CLOSED-LOOP NON-INVASIVE BRAIN STIMULATION AS A FATIGUE COUNTERMEASURE

A Dissertation

by

ROHITH KARTHIKEYAN

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Ranjana K. Mehta
Committee Members,	M. Cynthia Hipwell
	Kiju Lee
	Anthony D. McDonald
	Jason B. Moats
Head of Department,	Guillermo Aguilar

May 2022

Major Subject: Mechanical Engineering

ABSTRACT

Cognitive fatigue is a pressing issue across multiple job domains, especially within safety critical systems, where fatigue-driven lapses in operator performance and behavior have resulted in costly errors. While the importance of rest and fatigue management is known and widely acknowledged, sustainable solutions to address the core problem remain far and few between.

A major bottleneck remains a lack of clarity on the nature of the phenomenon, where a task invariant operational definition is impeded by individual differences, confounds related to workload, motivation or saturation, and the classic incongruities between performance or perception-centric views of the construct. Furthermore, mitigation strategies have rarely strayed beyond consumables (e.g. caffeine) or pharmacological aids that are used on an ad-hoc basis. Therefore, there is an imminent need to develop alternatives that could sustainably redress these conditions, and advance solutions that are explicitly informed by user cognitive states.

To that end, this dissertation investigates the operational significance, the mechanistic relevance, and fieldability barriers associated with the use of non-invasive brain stimulation (NIBS) as a fatigue countermeasure. Low intensity electrical stimulation of brain regions responsible for cognitive faculties such as attention, vigilance, and working memory has shown promise across a myriad task contexts, and lends confidence to its use as a fatigue countermeasure in related experiments. A fatiguing working memory exercise serves as the sandbox to investigate the parameters of interest for NIBS on fatigue effects. Stimulation was shown to enhance task performance, while the subjective perceptions of fatigue remain unaltered. Cerebral hemodynamics reveal structural and functional disparities of fatigue effects over the time course of the experiment, with cardiac activity, captured unobtrusively, shown to mirror the changes observed across task-relevant cortical networks. This observation lends confidence to the predictive potential of cardiac measures as neurocognitive indices of fatigue states, however, the search for the underlying ground truth that qualifies these states remains an open question. This dissertation explores the specific challenges and benefits that come from taking performance and perception-centric views of fatigue, while

proposing alternatives derived from cardiac heuristics to successfully forecast these states over a desired prediction horizon.

Together the evidence found in this dissertation provide the necessary foundation for the development of a closed-loop framework for mitigating fatigue states in related experiments. However, the domain translation and task-invariability demands of these observations demand more ecologically valid task contexts, and extensive validation studies in future assessments.

DEDICATION

To my parents *Nithya Kalyani* and *Karthikeyan*.

For everything and more.

ACKNOWLEDGMENTS

I owe an immense debt of gratitude to all those people who have pushed me to this finish line. Their continuing faith and support have inspired me to new heights in more ways than one.

Thank you Dr. Mehta for your counsel, mentorship, and most importantly, for giving me the opportunity to make this research my own. Your motivation and faith in my ability has been instrumental in getting me here.

Thank you to each member of my committee – Dr. McDonald, Dr. Moats, Dr. Hipwell, and Dr. Lee for your constructive criticism, scholarly input and enthusiasm in discussing my work, and helping me improve my science.

Thank you to my friends and peers at the NeuroErgonomics Laboratory – Yibo, Oshin, John, Sarah and Maher for the many passionate discussions and camaraderie that made this process all the more enjoyable.

Many thanks to my team on the NHANCE project – Reed, Joceleen, Yixin, Joshua, Connor, Slava, Ethan, Caleb, and Santi. I have learned so much from you all, and this would be far from possible without your patience and skill.

I also wish to extend my gratitude to Dr. Seok Chang Ryu, Dr. Hangue Park, Dr. Alan Freed, and Dr. Theodora Chaspari for their mentorship and guidance early in my doctoral program.

My thanks also go to my friends Akshay, Namita, Kenny, Zaryab, Scheherzad, Abdullah, Varun, and the *running crew* who helped me keep my spirits and wits through the many ebbs and flows of graduate life.

Finally, I wish to express my deepest gratitude to my family – my parents for nurturing my passions through the years and encouraging me to pursue my ambition; my wife and partner, *Manasa* for her unconditional support, patience and love, and our dog, *Googly* – words cannot convey how much your presence have meant to me through these years, thank you!

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Dr. Ranjana K. Mehta (advisor), Department of Industrial and Systems Engineering and affiliated faculty with the Department of Mechanical Engineering; Dr. Anthony D. McDonald, Department of Industrial and Systems Engineering; Dr. Jason B. Moats, Emergency Services Training Institute/ Texas A&M Engineering Extension Service; Dr. Cynthia M. Hipwell and Dr. Kiju Lee from the Department of Mechanical Engineering.

Data collected for this dissertation was facilitated by a cohort of undergraduate researchers at the NeuroErgonomics laboratory through the Aggie Challenge program, they include, in chronological order of their participation, Ethan Vargas, Santiago Garcia, Caleb Jones, Joshua Carrizales, Meredith R. Smoot, Iaoroslava Konopleva, and Connor Johnson. All other work on this dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by the College of Engineering's Teaching Fellows program at Texas A&M University in Fall 2021, and the fellowship for Preparing Future Faculty in Mechanical Engineering from the Department of Mechanical Engineering in Spring 2022.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Work responsibilities in today's world demand a high level of cognitive function, sometimes with minimal to no physical exertion, or opportunities to recover under persistent task demands [1, 2]. Work days are no longer bounded by hours of daylight or related constructs that defined schedules since the early 1900s [3]. In fact, these distinctions have been skewed further with the onset of "work-from-home" and the hybrid work culture that has taken precedence since the onset of the pandemic [4]. As humans continue to invest in this altered format and the demanding schedules that follow, cognitive fatigue is now a ubiquitous human condition resulting in high levels of burn out across job domains [5]. In fact, the term "languishing" has entered our collective rhetoric, and has become a very familiar experience to us all [6]. This predilection and the incremental burden of work demands can contribute to symptoms such as increased distractibility [7], lack of motivation [8], changes in how we process information [9], and also an inability to manage moods [9]. Indeed, the underlying fatigue that drives these changes is known to inhibit our ability to detect errors, or take remedial action and promotes error-prone and risky behaviors in all individuals [10]. It comes as no surprise then, that these deficits have shown to result in progressively worsening health situations, performance and in some instances, an inability to serve essential job functions, e.g., impairment of treatment decisions by doctors [11], operator lapses in air traffic control [12], and in hospice care [13]. In safety-critical job domains such as emergency response (ER), or front-line medical practice, these problems take a particularly insidious turn, driven by multiple contributing factors such as long work hours, harsh working conditions, shift schedules, sleep deprivation, heightened risk, workload, and understaffed operations [14].

Sections of this chapter are reproduced with permission from "Towards a Closed-loop Neurostimulation Platform for Augmenting Operator Vigilance." by Karthikeyan, R., & Mehta, R. K. published in the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 3976-3983). Copyright IEEE.

1.1 Contributions

This dissertation explores the opportunities that enable, deficiencies that demand, and the mechanistic barriers that impede the development of effective, use-inspired solutions to combat cognitive fatigue. Specifically we consider the relevance of non-invasive brain stimulation as a fatigue countermeasure. The fatigue problem plagues different domains at different scales, yet, a consistent task invariant understanding of the phenomenon, a representation of associated biomarkers, and (or) sustainable mitigation strategies remain far from a reality. The reasons for this are several fold, including but not limited to the inherent complexity and subjectivity that underlies the human experience of these states, the lack of consistent experiment paradigms to study their manifestations, the challenge of conducting effective longitudinal investigations on healthy populations, and the absence of promising alternatives that look beyond pharmacological solutions. Therefore, the goal of this dissertation is to cut through some of these challenges by **(1)** providing clarity on the nature of fatigue during sustained cognitive activity, specifically fatigue as induced by the time-on-task effect, **(2)** introducing the prospect of neuromodulation as a countermeasure to those states, and **(3)** exploring techniques to forecast "fatigability" before it is perceptually available or evidenced in outward human behavior.

1.2 Motivating Factors

In safety-critical socio-technical systems, sustained attention or vigilance, and working memory form core competencies necessary to fulfill several primary job performance requirements [15]. With increasing *time-on-task* demands, operator cognitive states can be compromised as *vigilance decrements* begin [16], and individuals struggle to remain goal oriented. There are varied definitions for this effect, and the associated diminution, but there is some precision in characterizing it as the ability to sustain attention for an extended duration of time while maintaining optimal task performance. Therefore, contingent on the nature of the task, the observed decrease in attention capacity over a given time period is ascribed to vigilance loss driven by the time-on-task effect or task-induced cognitive fatigue [17, 18]. The consequence of vigilance decrements within these systems is not trivial, with operator lapses or error identified as a contributor to several well-researched case studies such as, Three Mile island, and Chernobyl [19]. Some studies attribute their cause to sleep deprivation, monotony, and (or) boredom [20], while others explore underlying cognitive factors that prime these states [21]. The ability to predict the onset of these conditions, and to compliment user cognitive deficits through intervention remains an active research effort [22, 23, 24].

1.3 A Neural Basis for Cognitive Fatigue

Studies have identified theoretical models, and measures for the quantitative, and qualitative estimation of fatigue-driven performance deficiencies in human operators. These include the *arousal theory* which posits cortical arousal as a causal indicator for these behaviors. Successive investigations with functional magnetic resonance imaging (fMRI), and electroencephalography (EEG) suggest that these changes are correlated with cortical arousal; however their causal relationship remains suspect [25, 26]. In contrast, the *underload* theory suggests that performance deficiencies induced by the time-on-task effect is due to the observer's withdrawal of focused attention to a task due to task monotony [27].

The underlying neural basis for fatigue is an important consideration in this investigation, and it remains a complex issue. There are multiple constructs and neural systems elementary to the relatively non-specific alertness and activation states associated with these phenomena, implying that they are not uni-dimensional [28]. Some attempts have been made on using EEG workload indices as predictors for monitoring such states, with the measured *Task Load Index*, and *Engagement Index* from lower-frequency *alpha* as an objective diagnostic measure for time-on-task driven deficiencies [29]. However, this and related metrics need further investigation on the variability associated with altering task demands. fMRI studies report that attentional performance is correlated with *BOLD-signals* especially in the parietal and pre-frontal cortices [30, 31], nevertheless, these methods are not optimized for everyday/ ambulatory use-cases. Another study reports the responsiveness of cerebral hemodynamics to operator fatigue, and vigilance states using near-infrared spectroscropy (NIRS) [32] – this non-invasive ambulatory technique shows promise in providing correlates for vigilance decrement but is limited by its spatio-temporal resolution, and operating constraints [33]. On the subject of deficit prediction, there is also emphasis on the need for *wearables* that are unobtrusive to task performance requirements, while remaining sensitive to the physiological, and cognitive triggers that capture operator lapses. For the systems discussed thus far, this remains an impediment, however, recent efforts suggest that heart rate and its variability (HR/V) are sensitive metrics for predicting the cognitive demands of sustained attention tasks [34], in particular, spectral power ratios, time-domain, and some non-linear parameters derived from the inter-beat interval (IBI). These metrics can be obtained using electrocadiography (ECG) enabled wearables that are easy-to-use, inconspicuous, and self-contained.

1.4 Non-invasive Brain Stimulation as a Countermeasure

With the ubiquity of fatigue exposure and risk, there is the need to sustainably reverse or otherwise delay the condition using some active, adaptive mechanism. Research in this space dwells on the use of both conventional and unconventional solutions. Conventional paradigms include – meditation, yoga, caffeine, and nicotine among others; with each showing varying levels of success, and some inherently limited by their mechanism of action [24]. Unconventional solutions include pharmacological agents such as *D-Amphetamine*, and *Modafinil* among others, with each having unique pharmacodynamics, activations, and long-term addiction potentials [35]. Some of these modalities are used on a spontaneous, ad-hoc basis, with no clear statistical validation of their efficacy (e.g., coffee in the workplace). Moreover, there is a need for tools that can act prior to symptomatic expression, and also those that are mapped to the operator's "real-time" cognitive state. Neuromodulation techniques that operationalize non-invasive mechanisms show promise in enabling some form of preemptive action, and in eliciting the desired response. In particular, the use of transcranial direct current stimulation (tDCS) has been explored by multiple research groups [36, 37]. In these studies, a small current (1 to 2 mA) is applied to the user's scalp for a short duration of time (up to 40 $min.$). These small currents are stated to influence levels of cortical excitability, thereby altering user experience of fatigue symptoms such as diminished working memory, vigilance capacity, etc. Research by Nelson et al. showed promise in vigilance enhancement via tDCS when administered over the dorsolateral prefrontal cortex (dlPFC), while performing a simulated air traffic control task [23]. Similar findings have been reported with participants subject to a psycho-motor vigilance Test (PVT), and the Mackworth clock test [38].

For the scope of this investigation, we consider the foundational aspects of developing a closed-loop, adaptive mechanism for tDCS administration – e.g. to identify specific brain regions impacted by fatigue processes, the markers that could signal the onset of those impacts, and the influence that lower bounds in stimulation intensity could bring to these conditions. In the studies prior, including those related to fatigue and related constructs, tDCS dosages were based on arbitrary, sequential time-intervals with non-specific relationships to user cognitive states [23, 38]. Our ultimate vision is to aid the development of a system that is explicitly, and concurrently informed by physiological triggers/ thresholds that determine the onset of fatigue. Therefore in subsequent chapters we provide clarity on the specific fatigue processes of interest, the early evidence in support of tDCS as a countermeasure to those faculties, and the challenge in finding an objective ground truth that characterizes these states sufficiently.

1.5 Dissertation Structure

In an ideal world, the steps that underpin this dissertation would read **(1)** understand the nature of time-on-task fatigue, **(2)** identify relevant biomarkers that signal deficiencies driven by (1), and **(3)** explore potential countermeasures informed by (2) to address the aforementioned deficiencies. In reality, however, our experiments, in a true alignment with the origins of psychophysics, was first centered on understanding the impact of stimulation on brain functions during a fatiguing activity, before working backwards to understand the nature of fatigue on the task itself, and to finally answer the challenge of forecasting those states. Therefore, this dissertation is organized as follows. In Chapter 2 we introduce our observations on the impact of NIBS on task performance and perceptions during a fatiguing working memory exercise, we explain the specific mode of cognitive fatigue of interest in this research, the parameters of concern for neuromodulation, and the metrics that qualify the experience of fatigue in our participants. Findings discussed in this chapter provide insight on the benefits enabled by intervention while also highlighting the barriers related to their translation. In Chapter 3, we take a high resolution lens to the experience of cognitive fatigue by relying on the fidelity and spatio-tempoeral resolution of wearable neuroimaging, i.e. functional near-infrared spectroscopy to explain the manifestation of fatigue during this exercise. We contrast neural indices to the fieldability of wearable electrocardiography (ECG) data to investigate the prospect of unobtrusive indices that signal the subjective experience of fatigue. Findings in this chapter introduce the relevance of cardiac activity and derived metrics as a reliable neurocognitive index of fatigue. In Chapter 4 we investigate the prospect of forecasting fatigue states by relying only on data derived from cardiac activity. However, in exploring this problem, we found other challenges, specifically those related to the annotation of fatigue or fatigue labels, i.e. how do we qualify states of fatigue. We discuss the merits of taking a perception-centric view of fatigue over performance and vice-versa, while proposing heuristics derived from cardiac indices seen in clinical manifestations of fatigue as a potential alternative. Findings discussed in this chapter raise the need to expand how we aggregate annotation methods to signal the underlying ground truth. Chapter 5 presents the conclusions of our effort and the future opportunities that could advance an operational closed loop NIBS solution as a fatigue countermeasure.

2. ANODAL TDCS AUGMENTS AND PRESERVES WORKING MEMORY BEYOND TIME-ON-TASK DEFICITS

2.1 Abstract

Transcranial direct current stimulation (tDCS) of the left dorsolateral prefrontal cortex (dlPFC) has been shown to promote working memory (WM), however, its efficacy against time-on-task-related performance decline and associated cognitive fatigue remains uncertain. This study examined the impact of anodal tDCS of the left DLPFC on performance during a fatiguing visuospatial WM test. We adopted a repeated measures design, where 32 healthy adults (16 female), underwent anodal, control and sham tDCS on separate days. They completed an hour long two-back test, with stimulation intensity, onset, and duration set at 1 mA, at the twentieth minute for 10 minutes respectively. Task performance, subjective responses, and heart rate variability (HRV) were captured during the experiment. Anodal tDCS substantially improved WM relative to sham tDCS and control for both sexes. These benefits lasted beyond the stimulation interval, and were unique across performance measures. However, no perceptual changes in subjective effort or fatigue levels were noted between conditions, although participants reported greater discomfort during stimulation. While mood and sleepiness changed with *time-on-task*, reflecting fatigue, these were largely similar across conditions. HRV increased under anodal tDCS and control, and plateaued under sham tDCS. We found that short duration anodal tDCS at 1 mA was an effective countermeasure to *time-on-task* deficits during a visuospatial two-back task, with enhancement and preservation of WM capacity. However, these improvements were not available at a perceptual level. Therefore, wider investigations are necessary to determine "how" such solutions will be operationalized in the field, especially within human-centered systems.

Sections of this chapter are reproduced with permission from "Anodal tDCS Augments and Preserves Working Memory Beyond Time-on-Task Deficits" by Karthikeyan, R., Smoot, M. R., & Mehta, R. K. published in Scientific Reports, 11(1), 1-11 (2021). Copyright Springer Nature.

2.2 Introduction

Fatigue in the emergency response (ER) domain has been typified as an accepted hazard of the job, with extended work hours, cognitive and physical demands, and sleep deprivation [39]. However, its impact on worker health and safety demands wider attention on contributory neurophysiological mechanisms and countermeasures. Performance decrements due to fatigue endanger not only the personnel involved, but also the public they serve [40]. Firefighters and emergency medical technicians, personnel in occupations with high fatigue-risk, are required to multitask using information from concurrent visual and auditory sources, where their working memory (WM) is frequently a key determinant of task success [41]. This ability can be significantly undermined in the field, for example, wild-land firefighters report sleep deficits exceeding ≈ 30 hours each week, which severely degrades their situational awareness and decision-making [42]. Schmiechel et al. found that the mere perception of fatigue can impede executive function [43]. Therefore, multimodal solutions to augment or otherwise preserve WM, beyond the subjective experience of fatigue, would bring great value to the operations of all first responders.

Neuroimaging studies have shown that Brodmann areas 46 and 9, i.e. primarily the left dorsolateral prefrontal cortex (DLPFC) [44] are most associated with WM and related executive function, and therefore, this region has been a focal point for targeted neuromodulation [45]. Transcranial direct current stimulation (tDCS), a non-invasive brain stimulation technique, was shown to improve cortical excitability in those regions and in turn promote WM in healthy adults [44]. Studies mainly report a polarity-dependent effect of tDCS on WM, where anodal tDCS of the left DLPFC with intensity $\in [1,\ 2]$ mA enabled performance enhancements across multiple WM tests [46, 47, 48, 49]. However, recent meta-analyses identified inconsistencies in the effectiveness of tDCS [50, 51, 52] owing to small effect sizes, heterogeneous stimulation parameters (e.g., timing, intensity, or duration) and study designs (e.g. repeated vs. between subjects design) resulting in variability of the observed impacts on accuracy, response time and other WM performance measures. Some evidence also questions the efficacy of sham tDCS as a control measure due to its effect on behavioral and physiological responses [53]. These discrepancies demand wider investigations into the effectiveness of tDCS with equity in participants and their experiences.

Of particular interest to us is the intersection of WM, cognitive fatigue, and neurostimulation. In examining the efficacy of tDCS as a fatigue countermeasure, McIntire et al. [54] demonstrated that anodal tDCS for 30 minutes at 2 mA resulted in improvements on attentional accuracy during vigilance tasks beyond the effects of sleep deprivation. However, they found no direct effect of stimulation on WM during those experiments. Similarly, Borragan et al. report that tDCS at 1.5 mA for 25 minutes did not counteract the behavioral influence of cognitive fatigue on a sustained WM paradigm [55]. Therefore, while the benefits of stimulation toward vigilance or attention capacity during fatiguing tasks have been reported [54], evidence for the same effect on WM remains lacking. Hence, research that evaluates the effectiveness of dosage, timing, and duration in task contexts beyond those previously tested is needed, as these factors drive subsequent use and acceptance in time-sensitive field applications such as ER.

In our previous investigation, we explored the role of tDCS on sustained attention during a prolonged (60 minute) psychomotor vigilance test (PVT) [56], where we relied on the PVT to serve both as a *fatigue induction* paradigm and as a measure of attention capacity – informed by related efforts that consider the influence of *time-on-task* and its impact on sustained attention [57]. The *time-on-task* effect is analogous to fatiguing protocols in the literature [58, 59], where sustained cognitive demand at fixed or varying workload levels is shown to elicit increases in subjective and objective fatigue indices. We found that anodal tDCS at 1 mA for 10 minutes enhanced vigilance capacity on the PVT beyond this *time-on-task* effect, while response time and accuracy were both seen to decrease otherwise. These findings were encouraging as they addressed the immediacy demands that are characteristic of the ER domain by showing the potential benefits of online tDCS on attention networks notwithstanding task demands that were present and continuing leading into the stimulation interval. Furthermore, the efficacy of a relatively short stimulation interval (10 minutes) at low current intensity (1 mA) is particularly relevant toward fieldability requirements in emergency response, where work conditions demand unobtrusive and expeditious modes of intervention. In the current study, we build on this investigation to consider the role of stimulation on *time-on-task* related deficits during a WM exercise. Under sustained periods of cognitive workload, human WM is compromised by *time-on-task* related fatigue [60]. This fatigue can manifest as an increase in response time, a decrease in accuracy, an increase in self-reported fatigue scores or a decrease in motivation to continue task performance [60, 61]. Previous investigations have successfully validated the use of a *two-back* protocol as a *fatigue induction* paradigm, specifically the work by Shigihara and Tanaka et al [62, 63] employed a two-back test for 30 minutes to induce fatigue in its participant pool before relying on a trail-making task to discern the effects of fatigue on cognition. The choice of a two-back task also affords the right balance between task workload and fatigability unlike the one-back or three-back protocols that are either too easy or challenging to the point of rapid saturation, an observation from our pilot efforts. Therefore, we co-opt this tested two-back protocol in our present investigation, but consider a singular task for 60 minutes to both induce fatigue and to evaluate WM across stimulation conditions.

We also espouse a vision of building a closed-loop solution for non-invasive neuromodulation in ER applications [56]. Current efforts with the use of electroencephalography show promise toward prescient state recognition [64, 65], however they remain impractical for field-use. Through this work we investigate the utility of heart rate variability (HRV) as an indicator of WM states across stimulation conditions. Prior studies have successfully demonstrated performance prediction when relying on HRV [66], however the impact of tDCS and the robustness of these prediction paradigms remain unclear [67]. The neurovisceral integration model (NVIM) explores functional associations between cardiac vagal activity, and activation in the prefrontal cortex [68], and could expand how we utilize HRV as a neurocognitive indicator. Therefore, the present study is focused on understanding the influence of anodal tDCS on WM performance during the fatiguing visuospatial two-back task. We consider a repeated measures, balanced experiment design with three conditions – anodal tDCS, sham tDCS, and control towards this goal. We hypothesize that anodal tDCS over the left DLPFC will prove an effective countermeasure to fatigue-related declines in WM capacity. Presently, we operationalize fatigue as a trait driven by the *time-on-task* effect and we characterize it by relying on a combination subjective self-reports and WM performance. Further, we look toward HRV as a means to understand changes in cognitive task performance using the NVIM.

2.3 Methods

2.3.1 Experiment Design and Methodology

This study employed a repeated-measures, counterbalanced *Latin* square design [69], with participants returning on separate days to complete a working memory exercise under three distinct conditions – control, anodal or sham tDCS. Participants were cast into three sex-balanced cohorts and the order in which each cohort was exposed to the treatment condition was based on the Latin square. Each session began with informed consent followed by subjective questionnaires on their mood (Profile of Mood States; POMS [70]) and sleepiness (Karolinska Sleepiness Scale; KSS [71]) levels (see Fig. 2.1 (a)); before their first sessions participants also responded to a background questionnaire. The sessions lasted ≈ 60 minutes each, and were divided into 12 blocks during which time participants completed a WM task. During anodal and sham tDCS conditions a researcher administered interventions at the start of the fifth block, i.e. approximately 20 minutes from the start of the session while remaining outside the participant's field of view. Prior investigations on performance deficits due to the *time-on-task* effect consider $20-30$ minutes as the ideal window for fatigue onset [63, 57], and it is our goal to explore potential opportunities for intervention around this interval. Furthermore, participants were blinded to the stimulation condition (sham or anodal), but were aware of control sessions due to the absence of stimulation peripherals. Between each block, participants responded to a subjective questionnaire related to their levels of fatigue, effort and perceived discomfort, they were instructed to complete this as quickly as possible. The average block transition time was ≈ 15 s. In addition, to reduce any anticipatory bias, participants were not informed of the precise duration of each block or experiment session.

2.3.2 Transcranial direct current stimulation

A 1×1 tDCS device (Soterix Medical, NY, USA) was used in this study. The primary region of interest for WM was the left DLPFC, i.e. anode over F3 according to the 10-10 EEG system,

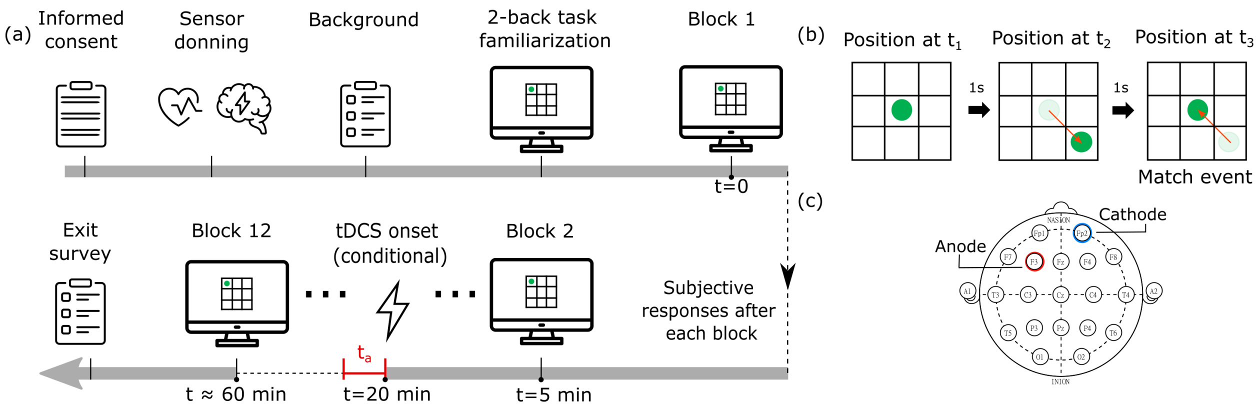

Figure 2.1: (a) Experiment protocol where t_a represents the 10 minute stimulation interval under the anodal condition, (b) schematic representation of a two-back match-event, and (c) tDCS electrode montage, cathode over the right supraorbital region (FP2), and anode over the left DLPFC (F3). Image created using *Inkscape 1.0.2-2*, https://inkscape.org/.

with the right supra-orbital (r-SO) region or FP2 used as the reference location (see Fig. 2.1(c)) [72]. Participants were instrumented with stimulation peripherals during anodal and sham tDCS conditions only. Before each session participants were acquainted to the sensation of the stimulus. The current intensity was set to 1 mA, and the current density was $0.028\ A/m2\ (area = 5{\times}7\ cm^2)$. Under anodal tDCS, the stimulation duration was 10 minutes at set point (1 mA); under sham tDCS, there was a ramp to the set point followed by a ramp to zero, lasting a total duration of $\approx\ 20$ s. The stimulation onset time was the same for both conditions – at the start of the fifth block, $\approx\ 20$ minutes from the start of each session. Stimulation onset time and condition were withheld from participants.

2.3.3 Metrics

2.3.3.1 Working Memory Performance

Three performance measures, namely accuracy, specificity, and sensitivity, were used to assess task performance on the WM test employed in this study. Table 2.1 presents a description of each measure. For analysis, the data was grouped into five phases I (baseline), II, III (stimulation), IV and V (terminal), such that phase I to IV consisted of two blocks each, and V was made of three blocks, with each block lasting five minutes. This phase resolution was adopted – (i) to

Table 2.1: Visuospatial two-back task performance metrics.

Performance measure	Description
Accuracy	$(TP + TN)/(P + N)$
Sensitivity	$(TP)/(P)$
Specificity	$(TN)/(N)$
Response delay (ms)	Time between stimulus and key-press

TP = **T**rue **P**ositives; P = all **P**ositive or *response selection* events;
TN = **T**rue **N**egatives; N = all **N**egative or *response inhibition* events.

Table 2.2: Heart rate variability indices

HRV Measure	Description
RMSSD	**R**oot **M**ean **S**quare of the **S**um of **S**uccessive Inter-beat Interval (IBI) **D**ifferences
LF	Power density of the **L**ow **F**requency $(0.04 - 0.15\ Hz)$ component in the IBI spectrum

ensure that the two-block stimulation phase (i.e. blocks $5,\ 6$ in phase III) was preserved uniquely in our comparisons, while ensuring block integrity across all phases; and (ii) since block 12 was excluded from our analyses due to a self-reported anticipation-bias, where participants realized that the experiment was near its end (e.g. by counting the number of blocks) and returned to a state of alertness.

2.3.3.2 Heart Rate Variability

Cardiac electrical activity was obtained from a chest-based electrocardiogram (ECG) probe and amplifier interface at 128 Hz (Actiheart 4, CamNTech, Inc., UK). The electrodes were positioned at the base of the sternum and over the left pectoralis minor muscle. The raw ECG signal was filtered for motion artifacts [73], and corrected for ectopics with polynomial interpolation [74]. Subsequently, a peak detection algorithm was used to isolate the R peaks of the ECG signal [75]. The time between successive R-R peaks, i.e. the inter-beat-interval (IBI), was then derived from the processed peak signals. For subsequent analyses, we derived two representative statistics for every five minute window in the raw IBI data, one in the time domain (RMSSD) and another in the frequency domain (LF-power; see Table 2.2). The measures were $min - max$ normalized across all three sessions, for each individual, before statistical analyses.

2.3.3.3 Subjective Responses

Participants were subject to three one-point subjective questionnaires related to their levels of fatigue, expended effort and perceived discomfort during the brief transition interval (≈ 10 s) between each five minute experiment block. Participants rated each subjective attribute on a 10-point scale, where a score of 1 was "low or minimal", and a score of 10 was rated "extreme or unbearable." Besides the block transition questionnaires, participants were also subject to the POMS and KSS surveys before (PRE) and after (POST) each experiment session. We used an abridged version of the POMS survey with 39 questions across six emotive categories – tension-anxiety, depression-rejection, anger-hostility, vigor-activity, fatigue-inertia, and confusion-bewilderment[70]. Participants qualified each emotion on a 5-pt scale with 0 being "not at all", and 4 being "extremely." The score across all descriptors in a category was summed to generate a factor score; and this factor score was used to generate a composite mood disturbance score. The KSS reflected participant sleepiness levels ranging from 1 meaning "Extremely Alert" to 10 meaning "Extremely sleepy, can't keep awake".

2.3.3.4 Statistical Analysis

The primary goal of the statistical analysis in this investigation was to explore the influence of experiment condition on performance, physiological and subjective responses at different time points (phase), and across both sexes during the WM exercise. Independent analysis techniques were applied for each dependent measure due to constraints in the data, but a common time scale was used across performance, HRV, and inter-block subjective responses, where the data was grouped into five phases. For the POMS and KSS, statistical comparisons were drawn between the PRE- and POST-experiment reports across each condition. Bonferroni corrections were applied where relevant to set the significance level at an appropriate threshold for multiple comparisons.

For purposes of clarity, data in the figures are illustrated using the mean value and the standard error of the mean (S.E.M.); all graphs were created using Matplotlib 3.4.2 [76].

WM performance measures were not normally distributed, we relied on the non-parametric Friedman's test to assess the effects of condition and sex (male, female) in each phase. Values that were more than three standard deviations from the mean were labeled outliers and excluded prior to analysis. Kendall's W was used to determine the effect size. Significant outcomes on the Friedman's test were then evaluated with pairwise Wilcoxon signed-rank tests. The HRV metrics introduced in Table 2.2 agreed with normality and sphericity constraints of the repeated measures ANOVA. Therefore, we performed a three-way repeated measures ANOVA to test the main and interaction effects of condition, phase, and sex on each metric. All two way interactions were analyzed to assess simple main effects and simple pairwise comparisons through t-tests. In addition, we compared the *baseline-subtracted* median accuracy scores between cohorts using a Kruskal-Wallis one-way analysis of variance test, and found no significant difference between them in any phase or condition (H-statistic $\in [0.46, 5.15]$; all p-values> 0.053), suggesting a balanced distribution of WM capacities across cohorts. Subjective inter-block responses were not normally distributed, and therefore analyzed using a similar approach to that on performance data. Analyses on the PRE- and POST-experiment mood and sleepiness surveys relied on simple paired t-tests. Where relevant, normality was assessed using the Shapiro-Wilk's test, sphericity using Mauchly's test for sphericity, and the homogeneity of variance among participants using the Levene test. All statistical analyses were performed in R using the *rstatix* and *tidyverse* packages.

2.4 Results

2.4.1 Working Memory Performance

During baseline, i.e. phase I and through phase II, there was no effect of condition on any of the performance measures (all p-values > 0.2). During stimulation and beyond, i.e. phase III to V, a significant effect of condition was observed on all three metrics (all $p < 0.0033$; $k_w \in [0.188,\ 0.407]$). Post-hoc analyses revealed that during phases III, IV and V, all measures of

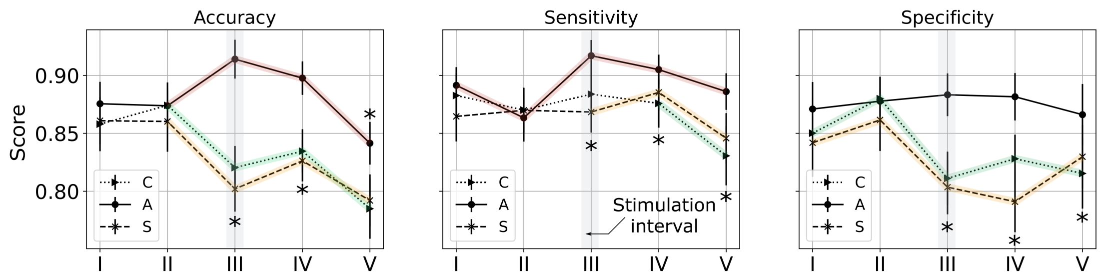

Figure 2.2: Trends in task performance for each phase as represented by the mean value and its standard error for each condition (control, anodal, and sham tDCS). The shaded segments represent *consecutive* time points where a statistically significant effect of time was found; the asterisk presents time points (phase) where a significant effect of condition was evidenced.

performance were significantly better under anodal tDCS relative to sham tDCS or control (all $p < 0.0022$). A main effect of time was observed across all conditions for the sensitivity and accuracy measures. For specificity, this was true only under control and sham tDCS (all $p < 0.022$; $k_w \in [0.056,\ 0.223]$). Overall, we note a decreasing trend in accuracy under sham tDCS and control. This decrease began from phase II, followed by a marginal improvement from phase III to IV, before declining further in phase V (all $p < 0.008$). Under anodal tDCS, we note an improvement in accuracy from phase II to III, and a gradual decrease across phases III and V (all $p < 0.00086$). Sensitivity under anodal tDCS exhibits a trend similar to accuracy across phase II and V (all $p < 0.027$). Under sham tDCS and control, sensitivity remained unaltered between phase I and III and phase I to IV respectively (all $p > 0.054$). However, in both conditions, we observed a decrease in the terminal phase (all $p < 0.048$). Specificity under sham tDCS and control was seen to improve in baseline and decrease from phase IIto III, followed by contrasting trends across phase III and V (all $p < 0.007$; see Fig. 2.2). No effect of sex was found on any condition or time variable (all $p > 0.71$).

2.4.2 Heart Rate Variability Measures

No effect of sex was found on any HRV measure considered (all $p > 0.89$). We found a significant main effect of time on the RMSSD values across all conditions (all $p < 0.0026$; $\eta_g^2 \in [0.16,\ 0.23]$). Post-hoc comparisons revealed an increase in RMSSD across phase I and II under

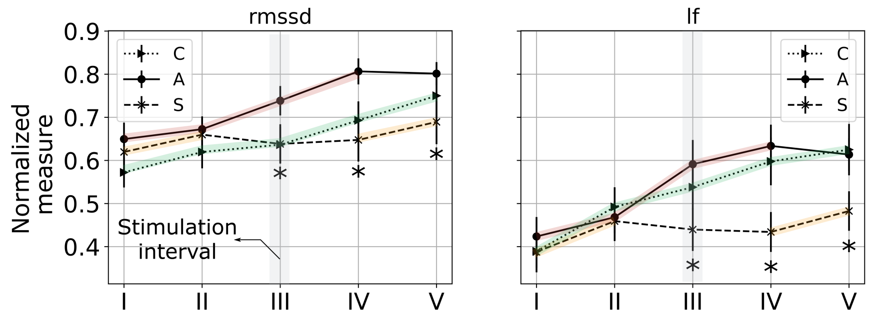

Figure 2.3: Mean and standard error of HRV features across the five phases: RMSSD (time-domain), LF-power (frequency-domain). The shaded segments represent *consecutive* time points where a statistically significant effect of time was found; the asterisk presents time points (phase) where a significant effect of condition was evidenced.

each condition (all $p < 0.007$). Under anodal tDCS, this increase in RMSSD was seen to persist until phase IV, followed by a plateau in phase V (all p < 0.001). Under sham tDCS, we observed that RMSSD remained relatively unchanged beyond phase II (all $p > 0.149$). Under control, we note an increasing trend in RMSSD, with statistically significant differences between phase III and IV, and phase IV and V (all $p < 0.019$). A main effect of condition was observed ($p < 0.0001$; $\eta_g^2 = 0.14$), where in phases III and IV, we found that RMSSD under the anodal condition was significantly higher than that under control or sham tDCS ($p < 0.019$). For the LF-power measure, a significant main effect of time was apparent in all three conditions (all $p < 0.0016$; $\eta_g^2 \in [0.08,\ 0.21]$). Post-hoc comparisons revealed an increase in LF power from phase I to II in all conditions (all $p < 0.041$). This increasing trend persisted under control and anodal tDCS until phase IV (all $p < 0.003$). Beyond phase IV, LF power under anodal tDCS was found to decrease in the terminal phase ($p = 0.002$), while remaining unchanged under the control condition ($p = 0.191$). Under sham tDCS LF-power was found to remain unchanged from phase II to IV (all $p > 0.076$), however in the terminal phase LF-power was elevated ($p = 0.041$); see Fig. 2.3.

2.4.3 Subjective Responses

No effect of condition was found across the subjective dimensions of effort, and fatigue (all $p > 0.122$). A marginal effect of condition was evident for the discomfort measure where participants

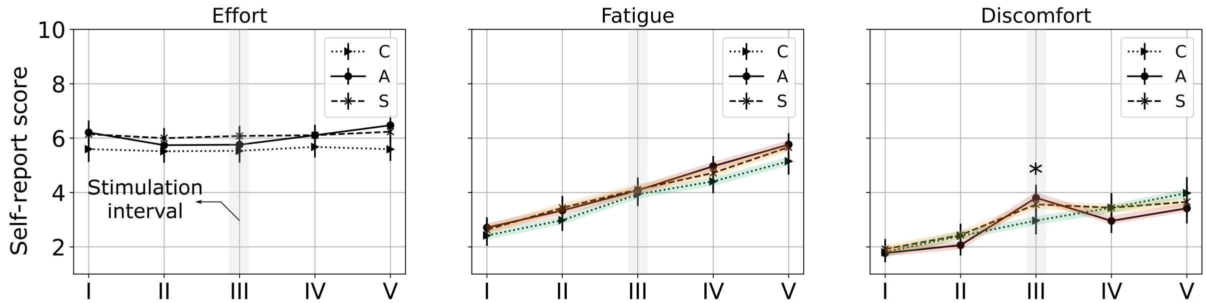

Figure 2.4: Perceived effort, fatigue, and discomfort scores as a function of time. The shaded segments represent *consecutive* time points where a statistically significant effect of time was found; the asterisk presents time points (phase) where a significant effect of condition was evidenced.

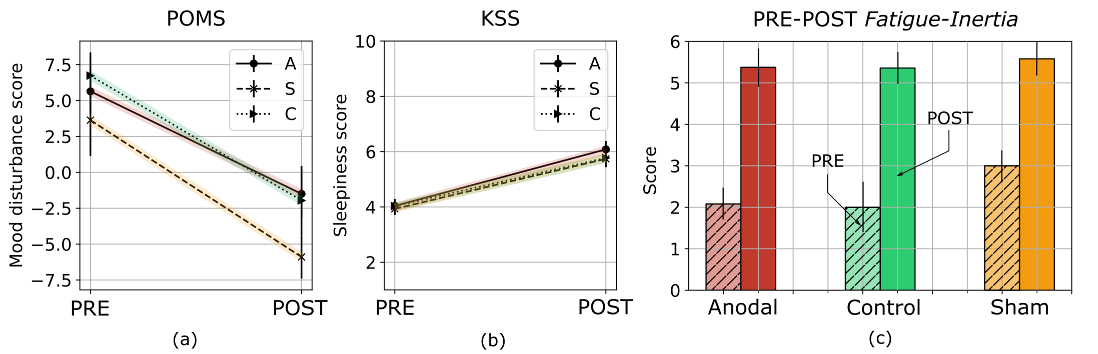

Figure 2.5: PRE- and POST- (a) mood disturbance and (b) sleepiness levels across all conditions as assessed by the Profile of Mood States survey and the Karolinska sleepiness scale. (c) PRE- and POST- *fatigue-inertia* scores from the POMS survey.

reported higher discomfort under sham, and anodal tDCS conditions relative to control during the stimulation interval, i.e. phase III ($p < 0.054$; see Fig. 2.4). A significant effect of time was observed across all three conditions for the discomfort and fatigue measures (all $p < 0.0001$; $k_w \in [0.27,\ 0.43]$), with participants reporting increased perceptions of fatigue and discomfort over time (all $p < 0.0022$). On the effort scores, a significant effect of time was reported only under the anodal tDCS condition (p= 0.027; $k_w = 0.08$). Post-hoc analyses revealed a marginal decrease of effort scores in phase III relative to phase I under this condition ($p = 0.054$). No effect of sex was found across the subjective responses (all $p > 0.561$). Furthermore, a significant difference was found in the PRE- and POST- mood disturbance and sleepiness scores across all conditions (all p

< 0.009), with a mood disturbance decrement of ≈ 8.4 points and sleepiness increment of ≈ 2.2 points after the completion of each session. No effect of sex or condition was evident for either measure (all $p > 0.316$; see Figs. 2.5 (a) and (b)).

2.4.4 Power Analysis

Thirty two participants completed this study (16 female), i.e. 96 sessions across all treatment variables. We employed a repeated measures design where participants were cast into three sex-balanced cohorts (10, 10, and 12) with the order in which they were exposed to each treatment condition (anodal, control or sham) counterbalanced through a Latin square. A power analysis was conducted to verify if our experiment had sufficient power to detect the effect sizes found significant in prior literature when employing a similar WM test, experiment design and stimulation protocol ($\eta_p^2 = 0.3$) [77]. Sample size estimation for a repeated measures ANOVA at a power level of 0.8, an error probability of 0.05 and partial eta-squared (η_p^2) of 0.3 using the G*Power software[78] suite revealed that a total sample size of 15 was required to test our hypothesis for the desired effect size. Given that we rely on non-parametric tests for significance in a subset of our variables, a 15% rule of thumb [79] increases the required sample size to 18 for the desired significance, 32 participants completed our experiments.

2.5 Discussion

The present study explores the role of anodal tDCS as a countermeasure to the *time-on-task* (TOT) effect during a fatiguing visuospatial WM exercise. The TOT effect, we hypothesized, would result in decreasing WM capacity and an increase in the perceptions of fatigue when individuals are engaged in tasks with sustained cognitive demand for extended periods of time. For example, in the work by Mockel et al. [58], on a three hour long *Simon* task [80], perceptions of fatigue, and performance levels were shown to progressively deteriorate within the first hour of task and further so with time. Therefore, in some ways, the TOT effect is analogous to the human fatigue response, although the timescales for typical fatigue experiments are longer. Previous efforts have established the role of stimulation as a fatigue countermeasure for attention and response inhibition when relying on tasks relevant to those constructs, e.g. Go/ No-Go tests, the Mackworth clock test, etc. [38, 54], indeed, investigations have also shown the efficacy of tDCS in improving WM in other conditions [81, 82]. However, evidence that supports the use of tDCS as a countermeasure for TOT effects on WM remains unclear, in fact, antecedent results show a lack of improvement in WM capacity under fatiguing conditions with intervention at different current intensities and duration [54, 55]. While the modes of fatigue onset remain disparate across these results, informed by prior work related to TOT-fatigue [63, 62], we employed a visuospatial two-back task to serve as both a *fatigue induction* mechanism and as a tool to evaluate WM. We compared performance, subjective and physiological responses, the influence of sex as a factor, and the effects of stimulation through a single-blind, repeated measures experiment, counterbalanced for learning across the three treatment conditions.

In both sexes, we found that anodal tDCS improved WM performance, beyond TOT-driven deficits that were otherwise evident across all performance measures (see Fig. 2.2). These improvements were found to last beyond the stimulation interval and enabled distinct changes that were unique to each performance measure for e.g., sensitivity improved under stimulation while specificity was maintained. However, no significant perceptual differences in self-reported effort or fatigue scores were noted between conditions. In general, participants indicated a marginal decrease in effort-scores under anodal tDCS and greater levels of discomfort during stimulation, as expected. Fatigue and discomfort reports were found to increase in a condition-agnostic manner over the time-course of the experiment, consistent with our expectations of the TOT effect, however, perceived effort remained similar throughout the hour long experiment, indicating participant motivation to stay on task. While mood and sleepiness changed with time, they were not significantly influenced by the condition variable.

The visuospatial two-back test, in its protracted format, demands both WM and sustained attention [83]. The central executive [84] is responsible for encoding and updating an internal WM buffer, recognizing match events, and priming task-relevant actions. The brain regions essential to this activity include the prefrontal, motor, and visual cortices [85]. These areas enable substantial behavioral adaptation, attention or response inhibition, and learning during WM exercises [86, 87]. However, with *time-on-task*, individuals invariably feel fatigued or otherwise depleted [57]. The onset of TOT-fatigue is marked by a decrease in neural activity across regions responsible for maintaining task-related behaviors, that are preempted by an increase in activation in regions peripheral to those networks [88]. These neurocognitive changes are supported by subjective fatigue scores, typically self-reported, or indirectly referenced from performance decrements [89]; in our protocol we rely on both modalities to operationalize TOT-fatigue. As reported, we found that subjective fatigue scores increased with time while PRE-POST surveys showed worsening mood after the completion of the experiment (see Figs. 2.5 (a) and (b)). Specifically, a substantial and statistically significant increase was found in the *fatigue-inertia* emotive category (see Fig. 2.5 (c)). According to the work by Schwartz et al. [90], the minimally important clinically-relevant difference (MICD) indicating fatigue on the POMS survey was 5.6 points, roughly 1.1 points for each category, and the MICD on a 1-pt fatigue questionnaire, similar to the one used in this investigation, was 2.4. In our observations, changes in the POMS mood disturbance scores, the POMS *fatigue-inertia* scores and the inter-block 1-pt fatigue responses exceeded these thresholds with a mean decrement of 8.45 ± 0.98 points, and increments of 3.077 ± 0.35 and 2.94 ± 0.14 points respectively, signaling the negative influence of sustained task performance on fatigue perception and reaffirming our TOT-fatigue hypothesis. Furthermore, this change was also reflected in performance behaviors as participants displayed a reduction in accuracy, sensitivity and specificity during both control and sham tDCS conditions beyond phase II, i.e. roughly twenty minutes into the experiment (see Fig. 2.2). Therefore, we reason that the observed changes in participant behaviors are driven by TOT-fatigue.

In light of the above evidence, the key finding in our work was affirmation that anodal tDCS at 1 mA for 10 minutes improved WM beyond TOT deficits. This improvement was seen both as the enhancement of WM capacity relative to baseline (accuracy, sensitivity), and as the preservation of performance levels over time (specificity; see Fig. 2.2). In addition, the effects were both concurrent and beyond transient, lasting for up to 20 minutes after the stimulation interval. Furthermore, we observed these changes in an experiment protocol, where intervention was provided after 20 minutes of sustained cognitive demand, which is noteworthy given mixed evidence thus far from studies that considered the online effects of tDCS on WM and contradictory observations of its effect on fatigue [52, 51]. Furthermore, earlier investigations with offline or online stimulation for longer durations and at current intensities greater than 1 mA reported no positive effect on WM under fatiguing or fatigued conditions [54, 55]. Additionally, we found no improvements in mood and sleepiness due to stimulation, contrary to some prior observations. Besides differences in the format of the WM task or stimulation parameters, we believe differences in the *fatigue induction* paradigm could be responsible for some of these divergent observations. For e.g., McIntire et al[54] explored fatigue driven by extended wakefulness, while our protocol demanded sustained attention for ≈ 70 minutes. Therefore, the nature of fatigue on the task is a factor to consider in future investigations.

Curiously, we note that although performance levels improved under anodal tDCS, changes in the level of perceived effort and fatigue remained similar across conditions (see Fig. 2.4). A recurrent hypothesis in this domain attributes cognitive enhancements enabled by tDCS to changes in neural efficiency [91], however subjective reports indicate that these "benefits" are mostly imperceptible, which demands consideration on how these intervention modes will be operationalized within human-centered systems. In addition, the self-reported fatigue scores appeared in contrast to those seen in the motor domain, where tDCS was shown to alleviate levels of perceived fatigue [92, 93]. The reasons for this disparity are perhaps related to the fundamentally distinct neurocognitive footprint of muscular and cognitive fatigue [94], regardless, larger studies with concurrent neuroimaging are necessary to unpack the neural mechanisms that underlie these distinctions, which remains the focus of our future efforts.

On performance improvements under anodal tDCS, we found that stimulation did not produce the same effect on sensitivity and specificity measures. During sham tDCS and control we discovered that sensitivity remained unchanged until the terminal phase, while specificity decreased beginning from phase III. Secondly, under anodal tDCS, sensitivity increased concomitant with stimulation, while specificity was maintained throughout the 60 minute task. It is likely that these distinctions are due to the fundamental differences in the behaviors these measures capture. Sensitivity identifies response selection, in this context, it characterizes event recognition and the commission of a prepotent motor response, while specificity identifies response inhibition and reflects the ability to withhold motor impulses [95]. Bender et al. demonstrated that these are fundamentally distinct cognitive operations governed by functionally and structurally independent brain regions [96], which may explain some of the observed differences. Studies also identify distinct cortical networks that underlie these processes, e.g. Rowe et al. showed the central role of frontoparietal networks, including the DLPFC, in response selection both when associated with WM and other tasks of willed action [97]; while others recognize the inferior frontal gyri (IFG), parieto-temporal junction, and supplementary motor areas as those essential toward inhibitory behaviors [98]. Filmer et al. [99], when using similar stimulation parameters, reported augmentation in response selection during anodal tDCS of the left lateral prefrontal cortex, while other efforts point to the role of the right IFG in enhancing task-relevant response inhibition under tDCS [100]. Therefore, we reason that the non-focal nature of 1×1 tDCS driven by the electrode size, current intensity and current density may be contributing to the differential impacts on response selection or inhibition behaviors observed in our investigation. Future investigations into the role of fatigue or *time-on-task* on these specific cognitive processes are necessary to shed light on the precise nature of their response and receptivity to stimulation.

HRV indexes the interaction between the central autonomic network and cardiac activity. The neurovisceral integration model posits that the main role of the prefrontal cortex during a WM task is toward attentional inhibition [68]. Empirical evidence points to a functional link between attentional inhibition mediated by the prefrontal cortex and vagal activity [101]. In our WM task, we expected that this would manifest as an increase in HRV during the early periods followed by a decrease due to TOT fatigue. However, HRV as indexed by RMSSD and LF-power increased under anodal tDCS until it plateaued in the terminal phase, while it appeared to increase throughout the experiment under control. In contrast, under sham tDCS, we found that the HRV indices increased through baseline before plateauing during the remainder of the experiment. Contextualized by performance data and subjective responses, we reason that HRV trends under sham or anodal tDCS agree with the expectations of the NVIM, however the control outcomes deviate from the model. It is possible that this deviation can be interpreted as the cumulative effect of fatigue and sustained WM demand on participants, with prior research indicating the linearly additive effects on HRV and similarly confounding observations [67]. Future investigations should consider HRV as an independent variable in investigating the influence of stimulation, which along with behavioral state dynamics and task state dynamics will serve as a crucial element in the development of fieldable closed-loop solutions for neuromodulation [102].

2.6 Limitations

We reported an anticipation bias in some participants during the terminal block, which required that we exclude it from our analyses, while this did not impact the trends observed in performance or subjective response, it remains a factor to consider in future study designs to avoid unanticipated influences on participant motivation while on task. Second, we do not adapt workload on the task to individual WM capacity, this appears as common practice in some earlier investigations on tDCS [103], however, we reasoned that a prolonged task format would elicit a consistent TOT-fatigue response in our participants which was determined to be the case. Third, our participant blinding strategy was potentially inadequate – during the experiments, participants were blinded to sham or anodal tDCS, based on existing best practices [104]. The consistent discomfort scores between the two conditions indicates that our blinding between stimulation sessions was likely satisfactory, however, the absence of stimulation peripherals during control may have skewed participant experience during that condition. Fourth, we do not capture motivation or engagement levels during active task performance. We made this decision to avoid disruptions to the participant's experience while on task, however, these measures could have improved how we understand the impact of tDCS on WM under fatigue [105]. Overall, our findings support the need for future investigations into the neural underpinnings that drive WM improvements, above and beyond the deficits induced by TOT-fatigue.

2.7 Conclusions

Cognitive fatigue can have serious consequences in safety-critical domains such as ER. Our research showed that, under controlled conditions, WM can be enhanced beyond the influence of TOT-fatigue with the help of anodal tDCS. Further investigation is necessary to provide clarity on the specific neural origins of these changes and their perceptual relevance to the human in the loop. Moreover, future explorations must consider how WM enhancements under abstract conditions in the laboratory translate to real-world ER WM demands. Additionally, although our work points to the positive influence of stimulation on task output, subjective feelings of effort, fatigue, and discomfort were not seen to benefit from neuromodulation, which demands consideration on "how" these technologies will be operationalized within human-centered, high-risk socio-technical systems.

3. A WINDOW INTO THE TIRED BRAIN: NEUROPHYSIOLOGICAL DYNAMICS OF VISUOSPATIAL WORKING MEMORY UNDER FATIGUE

3.1 Abstract

The neural and physiological drivers of fatigue are complex, coupled, and poorly understood. Investigations that combine the fidelity of neural indices and the field-readiness of physiological measures can drive translational research in recognizing and mitigating fatigue states in operational settings. We examine the spatio-temporal dynamics of neural activity and its physiological correlates in heart rate and its variability (HR/V) during a fatiguing visuospatial working memory task. Sixteen healthy adults, balanced by sex, completed a 60-minute fatiguing visuospatial working memory task. Changes in task performance, subjective measures of effort and fatigue, cerebral hemodynamics, and HR/V were analyzed. Peak brain activation, functional and effective connections within relevant brain networks were contrasted against spectral and temporal features of HR/V. We fouund that task performance elicited increased neural activation in regions responsible for maintaining working memory capacity. With the onset of time-on-task effects, resource utilization was seen to increase beyond task-relevant networks. Over time, functional connections in the prefrontal cortex were seen to weaken, with changes in the causal relationships between key regions known to drive working memory. HR/V indices were seen to closely follow activity in the prefrontal cortex. This investigation provided a window into the neural and physiological underpinnings of working memory under the time-on-task effect, with changes in functional and causal brain networks that unpack its influence on executive function. HR/V was largely shown to mirror changes in cortical networks responsible for working memory, therefore supporting the possibility of unobtrusive state recognition under ecologically valid conditions. Potential applications of this research include the development of a fieldable index for cognitive fatigue, and in bringing clarity to the nature of fatigue under the time-on-task effect.

Sections of this chapter are reproduced with permission from "A Window into the Tired Brain: Neurophysiological Dynamics of Visuospatial Working Memory under Fatigue" by Karthikeyan, R., Carrizales, J., Johnson, C., & Mehta, R.K. Currently in-review at Human Factors.

3.2 Introduction

The brain relies on a complex network of resources to facilitate working memory (WM) and associated executive functions [106]. The ability to sustain attention together with WM capacity remain central components of effective job performance in domains such as emergency response, frontline medical practice, and air-traffic control, where personnel are required to exhibit high levels of comprehension, reasoning, and vigilance for extended periods of time [88]. In these safety-critical systems, executive functions may be compromised by fatigue due to lapses in work conditions, the workload, hours on the job, or a combination of related factors. Fatigue due to time-on-task is known to induce additional cognitive burden, which can impair WM and limit our ability to manage task demands [58]. In particular, tasks that afford limited opportunities for individuals to implement compensatory strategies, for example, those with a high workload and a need for sustained attention, are known to be most vulnerable to the effects of fatigue [107]. In the laboratory, typical fatigue-WM experiments employ a battery of tests, such as the n-back test [61] the Sternberg task [108], the Simon task [58], etc., that provide a performance-oriented measure of the fatigue state experienced by an individual. Indeed, studies have also considered the use of these WM tests as the fatigue induction mechanism by manipulating the time-on-task variable [62]. In some cases, even shorter task durations with high workload have elicited an operationally significant fatigue response [109].Therefore, there exists a complex mapping between the workload, motivation, and time, among other factors, that appear to drive the human fatigue response [107].

Different techniques have been proposed to estimate cognitive fatigue including objective indices, and behavioral measures or self-reports [110, 111]. In the field, there is a need to obtain these measures in an unobtrusive manner, while remaining sensitive to the overall cognitive state changes experienced by the human. Behavioral self-reports are known to be interruptive [112], place additional cognitive demands [113], and occur at timescales that lag the fatigue states of the individual [114], at which point intervention may no longer be feasible. Objective indices on the other hand can be prescient and, in some instances, unobtrusive, yet there remain challenges regarding their use in the field [115, 116], especially under the constraints of an emergency response setting [110]. Therefore, there exists an unfulfilled need for fieldable and proactive fatigue estimation tools in safety-critical field applications. A key requirement to meet this demand is to bridge the gap between the fidelity offered by objective neural indicators and the fieldability of physiological indicators and self-reports. The barriers to this goal are primarily mechanistic, given the complex neurophysiological dynamics of time-on-task fatigue, working memory, and human attention; and the lack of valid data sets to explore this problem at depth.

Neuroimaging studies have studied the activation of related brain regions during different types of WM tasks [117, 118], the relationship between activation and working memory load [119], and the role of network measures, such as connectivity and causality [120, 121], to develop insights around brain function and the neural underpinnings of WM. For example, workload-related activation differences in the prefrontal cortex and changes in effective connectivity [120], i.e. the influence that one neural system exerts over another [122], during an n-back WM test using functional near infrared spectroscopy (fNIRS). The influence of fatigue on WM and its neural correlates have been investigated using similar tools, where studies report the effect of time-on-task on executive function using complex network analyses on electroencephalogram (EEG)-based connectivity features and identified the presence of small-world characteristics that were representative of fatigue states [123]. Other studies have successfully utilized fNIRS based indices to predict workload demands and fatigue correlates when performing ecological valid WM tasks [124]. However, translational work that extend these observations to physiological, unobtrusive indicators such as heart rate and its variability (HR/V) remain far and few between. In one investigation, researchers found that mental fatigue led to an increase in HR/V with time-on-task, and counter to their initial hypothesis, found no relationship to motivation-related task engagement in either behavioral or physiological measures during a fatiguing protocol [125]. They allude to an exhaustion of neural resources or changes in cognitive control as possible factors that drive this process, however more evidence is needed to support this hypothesis.

The neurovisceral integration model (NVIM) [126] provides a framework to juxtapose vagal activity, prefrontal cortex (PFC) activation, and executive function, one we speculate will help elucidate the ambiguities of earlier findings. Vagal activity can serve as a useful analogue to neural data while relying on unobtrusive sensing instruments. Specifically, the work by Thayer et al. provides evidence to suggest that the primary role of the PFC during a WM task is toward sensory inhibition, where with increased PFC activity we expect an increase in parasympathetic tone, and therefore an increase in HRV [127]. However, study designs, and associated findings remain variable, with conflicting observations on the relationship between HRV indices and WM demand. For example, in a recent study, authors extended the NVIM to explore comparisons between neural activity and HRV during a response inhibition task [128], where they found that respiratory sinus arrythmia, a marker of vagal activity was negatively correlated with cerebral oxygenation at baseline – consistent with prior observations from the NVIM [129]. However, this relationship was seen to deviate from model expectations during active task demands. This reasserts the need for further exploration on task-specificity and environmental demands to assess the relevance of the NVIM framework. Addressing gaps in this space remains critical towards the development of robust state estimation methods free from the practical encumbrances of current neuroimaging tools.

To that end, the present study is centered on understanding WM capacity under the influence of time-on-task fatigue using neural and physiological indices. We approach this problem by employing a protracted version of a visuospatial two-back test which demands high WM under constant workload, and sustained attention. The primary aim was to examine the spatio-temporal dynamics of neural activity, and the temporal dynamics of physiological responses during this fatiguing visuospatial WM task. A secondary aim was to compare neurophysiological signal behaviors to expectations from the NVIM framework. Together, fNIRS and HRV based indices may enable advances toward a robust predictive framework for recognizing fatigue-related WM deficits in operational settings.

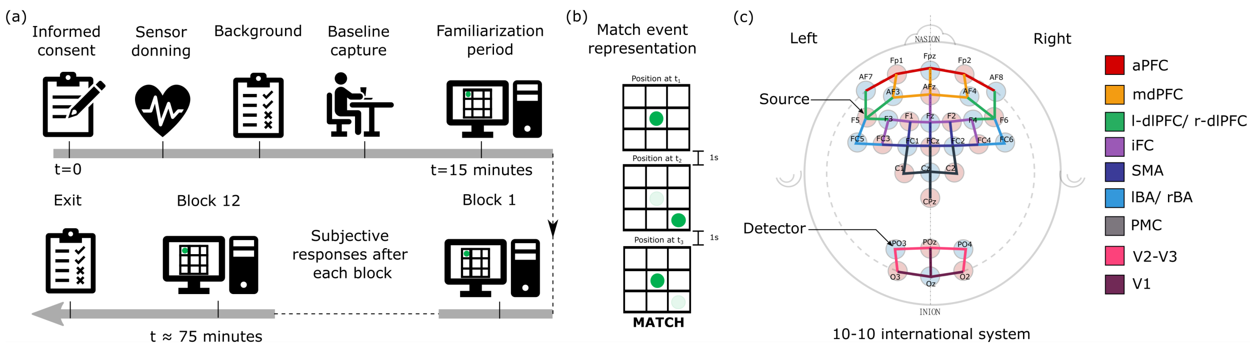

Figure 3.1: (a) Schematic representation of the experiment protocol. (b) Two-back match event where the user is expected to respond with a keypress. (c) Schematic representation of the probe-map used for neuroimaging via fNIRS. The probe-map consisted of eleven regions of interest derived from the 10-10 EEG system, where the red circles represent infrared (IR) sources, and the blue circles depict the IR detectors.

3.3 Methods

3.3.1 Participants

Sixteen participants were recruited (age: 25.12 ± 3.31 years; 8 female) from the local student population. For three participants, we had to interrupt the experiment when they needed a break, therefore only thirteen among them (seven female) produced unsegmented neural data compatible for subsequent analyses. All participants were self-reported to be right-hand dominant and provided informed consent before the start of the protocol. All procedures were approved by the university's Institutional Review Board and proceeded in accordance with the Ethics Code of the American Psychological Association.

3.3.2 Protocol

On informed consent, participants were equipped with relevant bio-instruments and responded to questionnaires on their background and demographics. Participants were then instructed to rest for five minutes with their eyes closed in a seated position to capture a baseline across all sensing instruments. They were then introduced to the visuospatial WM task, which included a training period, followed by the main experiment task. The task consisted of 12 blocks, with each block lasting a duration of five minutes. Between blocks, participants responded to single-element questionnaires on their perceived fatigue, effort, and discomfort on a scale of 1 to 10, with '1' being "low or minimal", and '10' being "extreme or very high". The specific phrase for each question was as follows: 1. Please rate the effort you expended in performing this task, 2. Please rate how fatigued you are from performing this task, and 3. Please rate your level of discomfort while performing this task. The time between any two blocks did not exceed 30s. The complete protocol is shown in Fig. 3.1 (a).

3.3.3 Visuospatial Working Memory Task

The experimental task employed in this study was a visuospatial two-back WM, consistent to the one reported in our earlier work (Karthikeyan et al., 2021). The task was presented on a static webpage using a desktop computer, where participants tracked a green circle (diameter = 20 mm) within a 3 x 3 grid (side = 130 mm), while seated comfortably in front of the screen (diagonal $\approx$ 600 mm) at a distance of $\approx$500 mm. The circle would appear in different sections of the grid; if the position of the circle matched the one from two steps prior, then participants would respond with a keypress. The inter-stimulus time was 1000ms, and the image persistence time was 900ms. The match probability was set to 0.6, where the interface provided a fixed, temporally randomized number of match events in each block (N = 94; see Fig. 3.1(b)). Before participants began the experiment, they could practice the two-back task under a training mode. The training interface provided feedback on response correctness and response time. During the experiment, this feedback was withheld from participants. The interface recorded every keypress or lapse event on the task with time stamps and a response correctness flag (hit, miss or false alarm). For subsequent discussions in this study, the performance measure used was the overall accuracy, which is defined as the ratio of hits + correct omissions to hits + correct omissions + misses + false alarms.

3.3.4 Bio-instruments

Participants were equipped with a head cap that housed sensor and detector probes of a continuous wave functional near-infrared spectroscopy (fNIRS) device (NIRSport2, NIRx Medical

Technologies LLC, USA). Cortical hemodynamics was obtained using the fNIRS device at 50Hz. Near infrared spectra were captured at two wavelengths (λ = 760 and 850nm). There was a total of 16 infrared (IR) sources and 16 IR detectors that characterized blood flow in the brain across 46 channels. These channels were originally focused on 11 regions: anterior prefrontal cortex (aPFC), dorsomedial PFC (mdPFC), right dorsolateral PFC (r-dlPFC), left dorsolateral PFC (l-dlPFC), intermediate frontal cortex, right Broca's area, left Broca's area, premotor cortex (PMC), supplementary motor area (SMA), secondary and tertiary visual cortex (V2-V3), and the primary visual cortex (V1; see complete probe-map in Fig. 3.1(c)). For the statistical investigations and results reported in this study, we focus on a subset of those regions, namely, l- and r- dlPFC, mdPFC, aPFC, SMA/PMC and the visual cortices (V; aggregating both V1 and V2-V3 regions as one). In addition to the fNIRS device, participants were instrumented with a chest-worn electrocardiography (ECG) device (Actiheart 4, CamNTech, Inc., UK) that was used to collect ECG data at 128Hz. Electrodes were placed at the base of the sternum and just beneath the left pectoralis minor muscle.

3.3.5 Signal Processing and Feature Extraction

3.3.5.1 fNIRS

Light intensity recorded from the fNIRS device was first converted to optical density. The optical density signal was low pass filtered at 3 Hz to attenuate high frequency noise. Motion artifacts were removed through peak detection and spline interpolation. The smoothed signals were band-pass filtered (0.016 $\sim$ 0.5 Hz) to reduce the effect of slow wave drifts and physiological noise in the data [130]. Lastly, the change in oxygenated, deoxygenated, and total hemoglobin concentration (ΔHbO/R/T) was derived using the modified Beer-Lambert principle using the HOMER2 toolbox [131] on MATLAB. For the scope of the analyses presented in this article we relied on the ΔHbO data which was used to derive region-wise peak activation (ΔHbOpeak), functional and effective connectivity measures. The raw time-series ΔHbO was sampled with a window of duration 15s which accommodates the underlying periodicity of the hemodynamic response. The peak values and functional connectivity (FC) measures were derived across each window, after grouping channels based on the relevant regions of interest (ROIs). For FC measures, we relied on Pearson's correlation coefficients that were transformed using Fisher's method [132]. Two ROIs were considered functionally connected only when the corresponding Fisher's z-score was $>= 0.4$ [133].

Time-domain effective connectivity (EC) analysis was performed to determine directed causal networks across the ROIs, namely l-dlPFC, r-dlPFC, mdPFC, PMC, SMA and V regions, using the Multivariate Granger Causality (MVGC) toolbox [134]. The MVGC, an autoregressive model, is based on the concept of Granger Causality, which posits that a time-series variable A drives another time-series variable B if the time-series history of A along with that of B improves the prediction of B better than its own time-history. For a complete guide to this method of analysis and the use of the MVGC toolbox, see [134]. The Grainger Causality magnitude was used as a measure of causal strength in our observations. Connections that were found significant were subject to Bonferroni corrections to account for multiple comparisons.

3.3.5.2 Heart Rate Variability

The raw ECG signal from the Actiheart was filtered for motion-related artifacts using a multi-resolution threshold [73], and ectopic beats were identified and removed by polynomial interpolation [74]. A peak detection algorithm was used to identify R peaks within the ECG signal [75]. The time between successive R peaks, i.e., normal-to-normal (NN) interval was then derived from the processed peak signals. We derived five representative statistics for every five-minute window for statistical analysis, three in the time domain (mean heart rate (HR), standard deviation of NN interval (SDNN), and root mean squared of successive differences (RMSSD)), and two in the frequency domain (low-frequency (0.04 - 0.15 Hz; LF) and high frequency (0.15-0.4 Hz; HF) power). These features were chosen given their empirical associations with executive function based on the NVIM [135]. All features were min-max normalized for each participant before statistical analysis [136].

3.3.6 Statistical Analyses

3.3.6.1 Data Partition

The fNIRS data, HR/V features, single-element subjective responses, and performance measure were partitioned into five phases – I to V; where the time-series variables were each characterized by their block mean. Each phase consisted of two experiment blocks from phase I to IV, while phase V was made of three blocks. Each block lasted a duration of five minutes, with $\approx 30s$ of transition time between them, where participants responded to the single-element subjective questionnaires. The last block (no. 12) was dropped from our analyses due to a self-reported anticipatory bias in some participants (N=6).

3.3.6.2 Analyses

The performance measure (accuracy) was not normally distributed, therefore we relied on the Friedman's test, a non-parametric equivalent to the one-way repeated measures analysis of variance (ANOVA), to assess the main effect of phase. Kendall's W (K_w) is reported as an estimate for effect size on the Friedman's test, with Wilcoxon signed rank tests for post-hoc analyses. On the fNIRS data, a one-way repeated measures ANOVA was applied to assess the main effect of phase on peak activation in each region, and between regions for functional and effectivity connectivity measures. Notably, we relied on aggregated activation data, i.e. mean across related regions, when analyzing changes in functional and effective connectivity. Therefore, while activation was captured across 8 regions, only 5 were used for connectivity analyses for clarity in our visuals and inference (see Figs. 3.2 and 3). On the ANOVA, we report the generalized eta-squared (η_g^2) as a measure of effect size [137]. All possible pairwise comparisons were made using paired *t*-tests to assess significance between levels of the within subjects' factor (phase). In subsequent discussions we primarily rely on comparisons between neighboring phases (i.e. between I-II, II-III, etc.) as highlighted in our plots. Heart rate-based measures were partitioned phase-wise, and subject to a one-way repeated measures ANOVA. Subjective responses were also non-normal, and therefore analyzed using non-parametric tests identical to those employed on the performance data. Bonferroni adjusted *p*-values

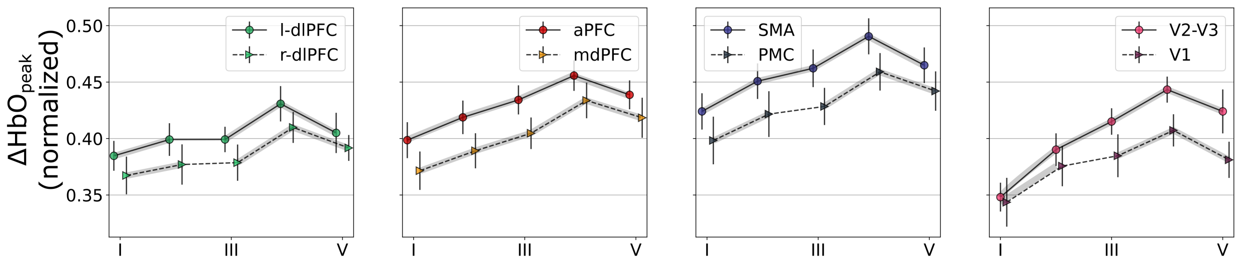

Figure 3.2: Peak activation across the five phases for each region of interest (ROI). All regions showed a significant main effect of time, shaded segments represent consecutive time points that were significantly different from each other. Error bars represent standard error. The plots are visualized with jitter on the x-axis for clarity of the error bars.

were used as a threshold to determine significance where relevant.

3.4 Results

3.4.1 Neural Activation

A significant main effect of time was found in peak activation across all brain regions (F(4, 48) $\in$ [6.82, 17.34], all $p < 0.04$, $\eta_g^2 \in$ [0.04 - 0.26]). 3.2 depicts the peak activation trend across each region. Post-hoc pairwise comparisons (10 in total for each region) revealed a consistent pattern of differences in the l-dlPFC, r-dlPFC, PMC and V2-V3 regions, where we observe an increase in peak activation going from phase I to phases II, III, and IV respectively ($t(12) \in$ [-4.47, -3.92], all $p < 0.001$); no significant differences were observed between phases II-III (all $p > 0.72$). Peak activation was found to increase further from phase III-IV ($t(12) \in$ [-3.62, -3.51], all $p < 0.0021$), and finally decrease from phase IV-V ($t(12) \in$ [2.31, 2.59], all $p < 0.02$). In the aPFC, mdPFC, SMA and V2-V3 regions, the increasing trend between peak values of phase I and phases II, III and IV persisted ($t(12) \in$ [-2.42, -1.98], all $p < 0.036$) with significant increments between each phase until phase IV, and a decrease in activation from phase IV-V, consistent with all other regions ($t(12) \in$ [-2.72, 2.15], all $p < 0.051$).

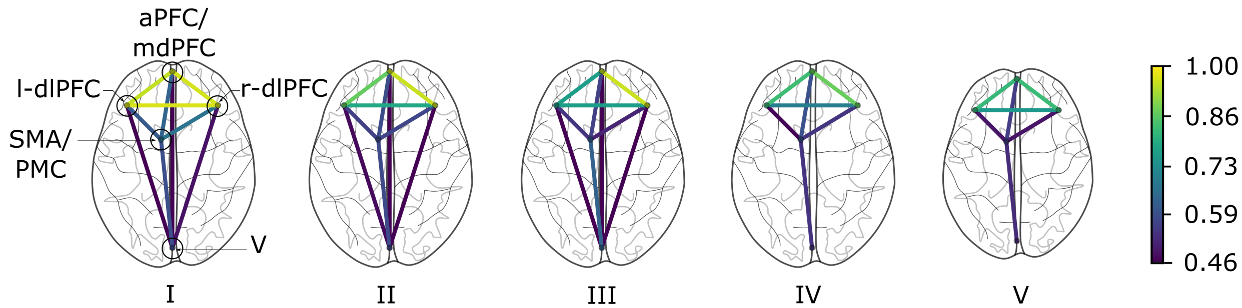

Figure 3.3: The graphic presents the mean z-score of functional connectivity strength by ROI across all participants at each phase, only those regions with significant connections are represented. Cranial positions shown here are approximate representations of the regions of interest.

Table 3.1: Mean z-scores of functional connectivity in each phase.

	FC_{mean} across each phase				
Connection	I	II	III	IV	V
l-r dlPFC	0.969 ± 0.088	$0.783^{I} \pm 0.075$	0.758 ± 0.085	$0.690^{III} \pm 0.088$	$0.645^{IV} \pm 0.090$
l-dlPFC - aPFC	0.951 ± 0.092	$0.871^{I} \pm 0.078$	$0.758II \pm 0.084$	$0.833^{III} \pm 0.092$	0.821 ± 0.103
l-dlPFC - SMA	0.609 ± 0.083	$0.563^{I} \pm 0.076$	0.543 ± 0.080	$0.463^{III} \pm 0.078$	0.491 ± 0.088
l-dlPFC - V	0.447 ± 0.104	0.474 ± 0.086	0.420 ± 0.089	0.340 ± 0.084	0.360 ± 0.098
r-dlPFC - aPFC	0.963 ± 0.088	0.955 ± 0.096	0.957 ± 0.101	$0.877^{III} \pm 0.098$	$0.827^{IV} \pm 0.107$
r-dlPFC -SMA	0.666 ± 0.103	$0.573^{I} \pm 0.080$	$0.514^{II} \pm 0.090$	0.543 ± 0.089	$0.488^{IV} \pm 0.090$
r-dlPFC - V	0.485 ± 0.100	0.455 ± 0.103	0.393 ± 0.102	0.337 ± 0.093	0.380 ± 0.099
aPFC - SMA	0.631 ± 0.106	$0.585^{I} \pm 0.086$	0.565 ± 0.099	0.594 ± 0.099	0.558 ± 0.105
aPFC - V	0.420 ± 0.106	0.460 ± 0.088	0.395 ± 0.086	0.346 ± 0.089	0.352 ± 0.100
SMA - V	0.589 ± 0.1214	0.582 ± 0.107	0.629 ± 0.108	$0.538^{III} \pm 0.089$	0.526 ± 0.102

*superscripts indicate when the mean z-score value of that phase was significantly different than the mean value in the preceding phase, as revealed through post-hoc comparisons, with all p < 0.003. The gray text cells represent insignificant functional connections (mean z-score < 0.4).

3.4.2 Functional Connectivity

Fig. 3.3 presents mean z-score of FC across all region-pairs at each time point. In general, we observe that (i) network-wide FC is positive, (ii) a global decrease in FC strength is apparent from phase I to V; and (iii) the number of significant connections were seen to decrease with time. Table 3.1 presents the functional connectivity strengths for each region and time-point. Here we aggregated activation data across the visual cortex i.e. V1, V2/V3 channels, the SMA/ PMC and aPFC/ mdPFC regions for convenience. A significant effect of time was observed in the connectivity strength across all inter-PFC connections, all PFC - visual cortex (V) connections, and between the SMA/PMC - V regions (F (4, 48) $\in$ [4.26, 22.34], all p < 0.04, $\eta_g^2 \in$ [0.14 - 0.36]). Connectivity in the PFC, across l-dlPFC, r-dlPFC and aPFC regions, had a mean magnitude of 0.961 in phase I; l-dlPFC and r-dlPFC associations were found to decrease from phase IV-V (t(12) = 3.61; p = 0.0017); while l-dlPFC – aPFC connections were shown to exhibit marginal recovery through phase III-V (t(12) $\in$ [-3.74, -3.52]; all p < 0.0021). Connectivity of the PFC with the visual cortices was weak yet significant, and relatively unchanged across phase I-II (all p > 0.716); a similar observation was found for connections between the PFC and SMA/PMC regions (all p > 0.143). In the terminal stages, i.e. phase IV to V, these connections were seen to weaken or were found insignificant.

3.4.3 Effective Connectivity

There was a main effect of time on the effective connectivity strengths across a subset of significant networks in this experiment (F (4, 48) $\in$ [3.42, 26.34], all p < 0.002, $\eta_g^2 \in$ [0.08, 0.19]); Fig. 3.4 presents all effective connections deemed significant in each phase using the causal strength metric. We found unique changes in causal dynamics over time, all significant observations are reported in Table 3.2, where significance is determined based on the p-values reported by the mvgc-pval function in the MVGC toolbox [134]. In phase I, we observed unidirectional effective connections originating from the l-dlPFC to the aPFC/ mdPFC and r-dlPFC regions (F_{mean} = 0.066; all p < 0.001; where F_{mean} is the mean causal strength). Significant unidirectional causality was also found between regions in the PFC, SMA/PMC and the visual cortices (F_{mean} = 0.056; all p < 0.001). In phase II we observed a decrease in F_{mean} compared to those levels seen in phase I (t(12) $\in$ [1.62, 2.31]; all p < 0.03), along with changes in causal directions, with bidirectional connectivity evidenced between the l-dlPFC-r-dlPFC and aPFC-r-dlPFC regions (F_{mean} = 0.051; all p < 0.001). In phase III connections between the l-dlPFC and visual cortex were not significant (all p > 0.71), and the direction of network causality in the PFC was reversed, with aPFC, r-dlPFC regions driving the l-dlPFC (F_{mean} = 0.052; all p < 0.001). This change persisted in phase IV,

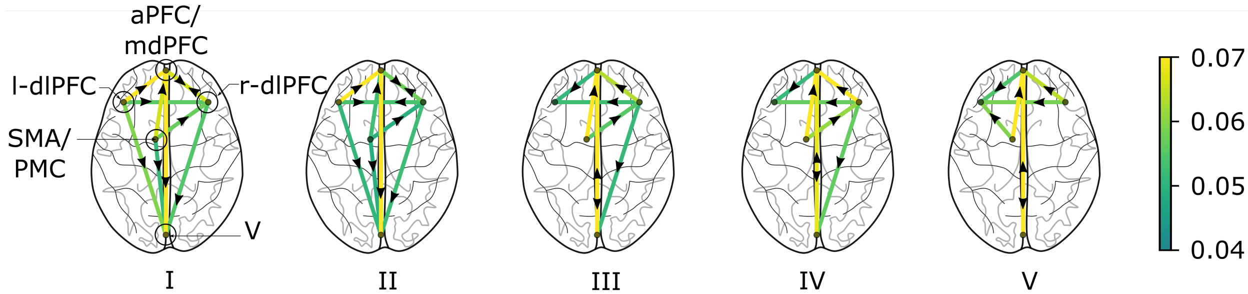

Figure 3.4: Effective connectivity during the visuospatial working memory task where arrows indicate the direction of causality. Only significant connections are shown. For bi-directional connections, the average strength is represented. Cranial positions shown here are approximate representations of the regions of interest.

where we also observed bidirectional causal connections between the aPFC - V regions (F_{mean} = 0.061; $p < 0.001$). In the transition from phase IV-V we found new causal pathways of significance where the SMA was observed to drive the l-dlPFC (F_{mean} = 0.066; $p < 0.001$), which until phase IV was shown to drive causal networks to the aPFC and r-dlPFC regions (F_{mean} = 0.049; all $p <$ 0.001).

Table 3.2: Significant effective connections across the phases of the experiment

	Connections		Effective Connectivity	
Phase	From	To	F_{mean}	σ
	aPFC	r-dlPFC*	0.064	0.0015
	aPFC	V	0.041	0.0024
	l-dlPFC	aPFC	0.069	0.0021
	l-dlPFC	r-dlPFC	0.051	0.0045
	l-dlPFC	V	0.050	0.0003
I	r-dlPFC	aPFC*	0.065	0.0005
	r-dlPFC	V	0.048	0.0017
	SMA	aPFC	0.065	0.0041
	SMA	r-dlPFC	0.064	0.0051

	SMA	V	0.046	0.0052
	aPFC	r-dlPFC*	0.053	0.0041
	aPFC	V	0.067	0.0069
	l-dlPFC	aPFC	0.069	0.0013
	l-dlPFC	r-dlPFC"	0.050	0.0005
	l-dlPFC	V	0.039	0.0027
II	r-dlPFC	l-dlPFC”	0.052	0.0048
	r-dlPFC	aPFC*	0.051	0.0061
	r-dlPFC	V	0.046	0.0017
	SMA	aPFC	0.046	0.0027
	SMA	rDLPFC	0.041	0.0023
	SMA	V	0.041	0.0052
	aPFC	l-dlPFC	0.046	0.0013
	aPFC	V*	0.067	0.0082
	r-dlPFC	l-dlPFC	0.049	0.0016
	r-dlPFC	aPFC	0.056	0.0019
III	r-dlPFC	V	0.046	0.0024
	SMA	aPFC	0.062	0.0006
	SMA	rDLPFC	0.052	0.0032
	V	aPFC*	0.063	0.0017
	aPFC	l-dlPFC	0.046	0.0022
	aPFC	V*	0.063	0.0061
	r-dlPFC	l-dlPFC	0.048	0.0053
	r-dlPFC	aPFC	0.068	0.0012
IV	r-dlPFC	V	0.046	0.0017
	SMA	aPFC	0.062	0.0042

	SMA	rDLPFC	0.052	0.0036
	V	aPFC*	0.066	0.0092
	aPFC	l-dlPFC	0.054	0.0071
	aPFC	V*	0.061	0.0036
	r-dlPFC	l-dlPFC	0.050	0.0022
V	r-dlPFC	aPFC	0.061	0.0043
	SMA	aPFC	0.069	0.0061
	SMA	l-dlPFC	0.055	0.0052
	V	aPFC*	0.064	0.0027

*represent bi-directional connections that were found significant.

3.4.4 Heart Rate-based Measures

A significant main effect of time was found across all four heart-rate based measures (F (4, 48) $\in$ [7.71, 14.46], all $p < 0.0001$, $\eta_g^2 \in [0.08, 0.13]$; Fig. 3.5). In mean HR we observed a significant increase from phase III-IV and IV-V as shown in Fig 3.5) ($t(12) \in [-3.61, -3.32]$; $p < 0.003$). For the LF measure, post-hoc comparisons revealed significant differences in the mean values between phase I, and all subsequent phases. Notably, we found that LF power density increases relative to phase I in all other phases ($t(12) \in [-3.15, -2.79]$; all $p < 0.01$). A similar increase was found going from phase II to III ($t(12) = -2.11$; $p = 0.028$); the measure was found to plateau across phase III-IV ($p = 0.132$), before decreasing across phase IV-V ($t(12) = 3.33$; $p = 0.003$). A significant increase in HF was evident between phase I and II ($t(12) = -2.38$; $p = 0.017$); thereafter HF was seen to remain unaltered across phase II - III ($p = 0.051$) before decreasing from phase III-IV ($t(12) = 2.46$; $p = 0.015$) and plateauing across phase IV-V ($p = 0.76$). In SDNN and RMSSD we observed significant increases across phase I-II ($t(12) \in [-3.211, -1.89]$; all $p < 0.041$), and II to III ($t(12) \in [-2.91, -2.23]$; all $p < 0.023$), while a plateau was observed between phase pairs III to IV (all $p > 0.054$), and a decrease in phase IV-V ($t(12) \in [3.55, 3.83]$; all $p < 0.002$).

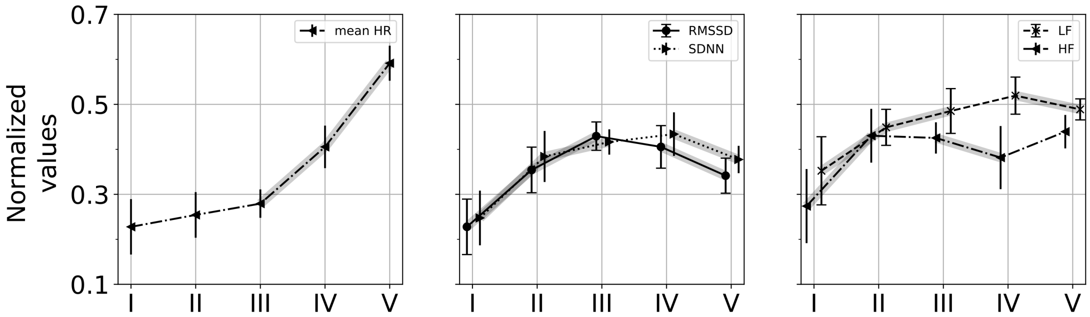

Figure 3.5: Trends in HR-based measures across the five phases. The plot represents the min-max normalized values for each feature. A significant effect of time was observed across all measures with a small to moderate effect size. Time domain features – Mean HR, RMSSD, and SDNN. Frequency domain features – spectral power densities in the LF (0.04 - 0.15Hz) and HF (0.15 - 0.40Hz) regimes. Shaded segments represent consecutive time points that were significantly different from each other, error bars represent standard error. The plots are visualized with jitter on the x-axis for clarity of the error bars.

3.4.5 Performance Accuracy

A main effect of time was found on the performance accuracy metric (χ^2_4 = 27.62, n = 13, $p <$ 0.0001, K_w = 0.21). Post-hoc analyses revealed a marginal increase in accuracy going from phase I to phase II ($t(12) = -1.91$; $p = 0.041$), a decrease in accuracy from phase III to phase IV ($t(12) =$ 1.76; $p = 0.052$), and a further decrease in accuracy levels from phase IV to phase V ($t(12) = 3.33$; $p = 0.003$); see Fig. 3.6 (a).

3.4.6 Subjective Responses

A main effect of time was found on all three subjective responses, i.e., perceived – effort, fatigue, and discomfort ($\chi^2_4 \in [47.4, 65.93]$, n = 13, all $p < 0.0001$, $K_w \in [0.08, 0.18]$). Post-hoc comparisons revealed an increase on all three self-reports early in the experiment (phase I-II; $t(12)$ $\in [-2.81, -2.43]$; all $p < 0.016$). This increasing trend persisted through phase III-V for discomfort and fatigue reports ($t(12) \in [-2.73, -2.33]$; all $p < 0.019$); however, effort scores did not vary significantly beyond phase II (all $p > 0.082$); see Fig. 3.6 (b).

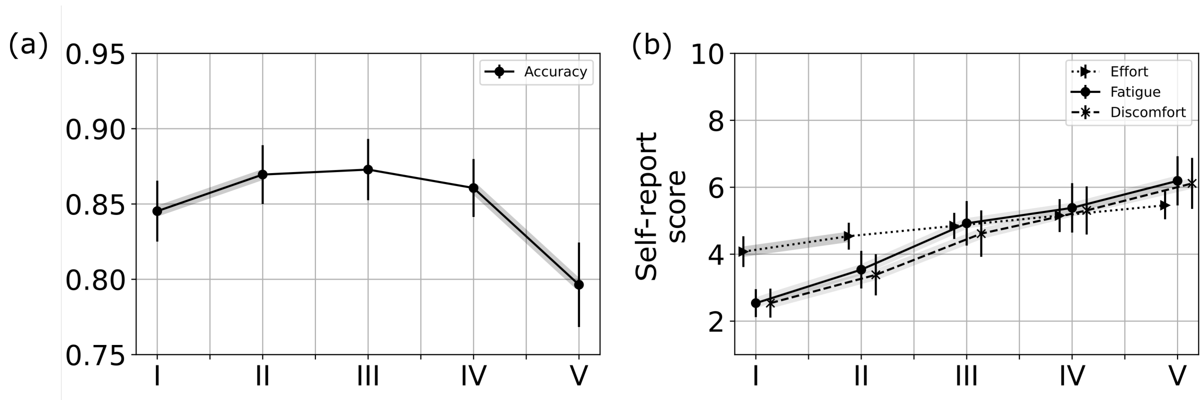

Figure 3.6: (a) Performance accuracy (%) during the time course of the experiment. (b) Subjective single-element responses on effort, fatigue and discomfort. Shaded segments represent consecutive time points that were significantly different from each other, error bars represent standard error. The plots are visualized with jitter on the x-axis for clarity of the error bars.

3.5 Discussion

In this study, we examined the spatio-temporal dynamics of brain activity, and changes in heart rate and its variability (HR/V) during a fatiguing visuospatial two-back WM test. In particular, we were interested in the time-on-task effect and related deficits during the WM exercise. We know that prolonged cognitive activity at fixed or varying workload levels is known to elicit an increase in both subjective and objective fatigue indices and a decrease in WM performance [59, 57]; and that this fatigue can manifest as a decrease in accuracy, an increase in self-reported fatigue scores or a decrease in the motivation to continue on task [60, 58]. Therefore, we hypothesized that these changes will be evidenced in the present study, driven by a combination of factors underlying the experience of the participant, including, learning during the initial stages of the experiment, a struggle to optimize available neural resources for staying on task, and a decline in the motivation to continue the exercise under the absence of additional rewards.

Our key findings were as follows, participants exhibited a largely stagnated trend in performance accuracy beyond the initial phase, before it worsened through the terminal stages of the experiment. We observed an increase in peak ΔHbO across key regions in the PFC, regions peripheral to the PFC, and the visual cortex until the penultimate experiment phase, where we note a global decrease in activation levels. These were accompanied by a decline in the number of significant functional associations across brain regions and an increase in perceived fatigue. Finally, the trends in peak activation were mirrored in HR/V, where mean HR remained relatively unchanged until phase II before increasing significantly until phase V, while temporal and spectral HRV features were seen to increase till the penultimate phase before diminishing (RMSSD, SDNN) or remaining unaltered (LF, HF - power) through the terminal phase. On the task we believe a few key processes were at play, early in the experiment (i.e. phase I-II) participants were likely "learning" the task until a stable performance threshold was reached beyond this period, participants expended effort to maintain task performance (III-IV), before they reached a state where they were unable to continue at that level or meet task demands altogether (IV-V)

Some of these observations are clearer when we consider the nature of the task. The visuospatial two-back test demands sustained attention and WM. Although workload on the N-back task was not adapted (N=2; constant), in its prolonged format , perceived workload was likely to vary [138], therefore eliciting distinct neural and physiological responses associated with the time-on-task effect [139], for example early in the experiment we believe that the perceived workload is high as participants learn to perform the task, beyond this period we anticipate the perceived workload to stagnate until time-on-task related effects render the task too demanding to continue. There are recognizable brain regions essential to the behavioral adaptation, attention and response inhibition, and learning characteristic of such WM exercises [84], including – the prefrontal, motor, and visual cortices [140]. However, with time-on-task, we believe that fatigue dominates the behavioral and neurophysiological responses that drive this process [120]. In our study, we found that participants' self-reports of fatigue increased with time, although their perceived effort remained unaltered beyond phase II. Concomitantly, we observed a significant decline in two-back performance accuracy during the latter half of the experiment. Fatigue is often characterized as the "the reluctance of further effort" [61], and given our observations with the self-reports and in performance outcomes, we reason that the changes seen during this experiment are largely fatigue related.

WM is primarily mediated by networks in the PFC, where domain-specific models postulate that the lateral PFC is functionally organized to process visuospatial information [141]. A key component of this network is the l-dlPFC region, which embodies specific computational mechanisms for monitoring and manipulating those representations [141]. The primary role of the frontal cortex during such WM tasks is towards attention inhibition, i.e., reducing the influence of distracting streams of information and to retain focus on the task at hand [142, 143]. Therefore, an increase in PFC activity, early in the experiment, is indicative of the effort employed by the participants in learning the WM task [144]. Beyond this period, activation in PFC regions was mostly stable, i.e., through phase II-III. One explanation for this is the static workload on the task, which is unlikely to elicit a fatigue response early on [145]. However, beyond phase III, we found that neural activity increased significantly into phase IV, and this was true across all regions, even those peripheral to the PFC. We reason that this indicates the onset of mental fatigue due to time-on-task effects, which prior investigations reveal is preempted by an increase in activation in regions peripheral to those essential for task-related behaviors [146]. Causse et al.in [146], argue that this is symptomatic of the additional resource demand placed by the need to sustain performance at some threshold, when performance accuracy itself appears to have saturated, further alluding to adaptations arising from fatigue in the form a distributed utilization of cerebral resources.

3.6 Conclusions and Limitations

Cognitive fatigue can have serious consequences in safety-critical domains such as emergency response. Our study captured fatigue-related neurophysiological dynamics during a 60-minute visuospatial WM task. We found that WM performance was significantly impacted by fatigue-related changes in neural activity. The changes in neural activity, and declines in functional and causal connections, were shown to be temporally coupled with heart rate and its variability. This observation reaffirms the prospect of operationalizing unobtrusive sensing paradigms for recognizing fatigue states in an applied setting. However, larger investigations under ecological valid conditions are necessary to ensure the generalizability and task-independence of our observations.

Brain activation across most of the regions monitored in this study showed an increasing trend, but beyond phase IV we found a global decrease in activation levels. This is likely due to a fatigue-driven lack of motivation to continue the task, especially in the absence of any additional rewards [61]. Our reasoning and inferences around the influence of fatigue and related task disengagement are further supported by our observations with the connectivity data, where we found elevated functional associations in the PFC early during the experiment, when performance was improving with participant learning on the task. Unlike our observations on regional activation, functional connections were seen to largely diminish in strength with time-on-task, with some connections between the prefrontal and visual cortices lost altogether in the later stages of the experiment. These observations are in alignment with prior evidence around the influence of time-on-task induced cognitive fatigue on functional connectivity, where bilateral connections in the PFC and peripheral regions were found to decrease over time [57].

The observations in effective connectivity reaffirm the centrality of the lateral PFC in enabling WM performance [141]. Early in the task, we found that the l-dlPFC was a prominent driver of connections in the PFC with unidirectional connections to the aPFC and r-dlPFC regions. These regions were also shown to exhibit effective connections with the visual cortices. Over time however, we observed **(i)** a role reversal in some of these pathways, with the r-dlPFC, aPFC and SMA region found to be driving l-dlPFC activity across phase III - V, **(ii)** some causal relationships between the prefrontal and visual cortex were lost altogether, and **(iii)** an overall decrease in network density as a function of time was observed. The reorganization of causal pathways is likely fatigue driven, with prior investigations reporting topology alterations, disintegrations, and directionality changes driven by fatigue due to time-on-task [147, 148, 149]. While the specific neural mechanisms that originate these changes remain unclear, we hypothesize that this could be an effect of the resource demand placed by sustained task performance, where with time-on-task additional cortical regions are recruited to preserve individual performance at some threshold. This hypothesis is supported by the observation that, beyond phase III, we found the r-dlPFC and APFC regions driving activity in the l-dlPFC region, a pathway that was not previously deemed significant.

Interestingly, the influence of fatigue in our observations was not limited to neural indices. We found distinctive parallels in the response of HR/V measures. The neurovisceral integration model (NVIM) posits that the key role of the prefrontal cortex during a WM task is toward attentional inhibition [68]. This inhibitory process is reflected in cardiac activity through the control of the vagus nerve. In our experiments we expected that this influence would manifest as an increase in heart-rate variability (HRV) during the early stages before fatigue-driven declines in the latter parts. The temporal characteristics of HRV, i.e. RMSSD and SDNN largely align with this expectation, i.e. as PFC activity increased with time-on-task, heart rate variability also increased. Mean HR was largely unchanged early in the experiment (phase I – III), however, beyond this interval, we found a significant increase in mean HR concomitant with a decrease in HRV measures. This fits our expectation of sympathetic dominance or parasympathetic withdrawal under cognitive saturation due to time-on-task [34, 68]. Furthermore, in the spectral domain, we found that the LF and HF-power densities mirrored these trends. Under controlled conditions, vagal activity is known to be associated with LF-power [150], and our observations corroborate this idea. The finding in this study that the dynamics of HR/V aligned with the NVIM and were seen to reflect changes in neural activity is one that is both interesting and in need of further investigation. In particular, there is need for clarity on the specific processes driving these relationships, for example, lower HR/V early on could be driven by learning on the task, before vagally-mediated effects or the influence of time-on-task fatigue. Nevertheless, we are optimistic that this could be a path forward in our search for robust and prescient state recognition in the field. Especially knowing that the recognition of WM deficits in emergency responders may benefit from technologies that capture such underlying neurophysiological dynamics unobtrusively.

Limitations of the study are as follows. First, in our experiments we found an anticipation bias in some participants (N= 6) during the terminal block, where they returned to an artificial state of alertness, this required that we remove the last block from our analyses to ensure that the true nature of time-on-task related fatigue is preserved. Second, we did not capture motivation or engagement levels during active task performance to minimize disruptions to the participant's experience. However, these measures could have improved how we understand the impact of motivation on fatigue and WM decline, which is shown to offset fatigue states in related experiment [151]. Third, besides the task demand, wearing the fNIRS instruments for a prolonged period of time is in itself uncomfortable, which may have skewed participant experiences and self-reports. Fourth, given the duration of the task, vascular nonlinearities may perturb the hemodynamic response signal and therefore our inferences [152]. While we account for issues such as sensor drift and avoid event-related analyses, the effects of vascular recovery and habituation need careful consideration. Finally, although our end goal is to translate laboratory results into fieldable fatigue detection solutions, our task and setup are not wholly congruent with real-world WM demands; this necessitates ecological valid experiment paradigms with stakeholders actively engaged in those responsibilities. Nevertheless, our findings support future investigations into the neural and physiological underpinnings that drive WM and related performance decline due to fatigue.

4. WHAT'S IN A LABEL? ANNOTATION DIFFERENCES IN FORECASTING COGNITIVE FATIGUE USING ECG DATA AND SEQ2SEQ ARCHITECTURES

4.1 Abstract

In this paper we consider strategies to annotate fatigue states during a prolonged visuospatial working memory exercise. Specifically, we look to address the need to forecast these states when relying on unobtrusive indices derived from cardiac electrical activity. We formulate this challenge as a multi-step-ahead, multivariate time-series forecasting problem, where we consider the effect of annotation variations for fatigue labels on model success and reliability. We found that a *sequence to sequence* (Seq2Seq) LSTM architecture was particularly faithful in identifying fatigue representations using heart-rate and its variability (HR/V). We also discuss the prospect of a heuristic-derived annotation index for fatigue that relies on changing patterns of HR/V as a signal for state changes in the human. We contrast perception-, performance- and heuristic-indices across other modeling variables such as the forecast horizon, the input vector horizon, and towards sample generalizability. Our findings indicate that, under controlled conditions, HR/V can serve as a useful neurocognitive index of fatigue, however, impact of activity, time scale, and other perturbations demand further investigation to evaluate the diagnosticity and robustness of the proposed indices.

4.2 Introduction

Cognitive fatigue is known to have significant consequences on human performance and behavior. In this paper we discuss strategies to discern fatigue states a-priori to their perceptual relevance to the human operator. This is an important problem because when subjective fatigue states become apparent it may be too late for individuals to find effective supplementation strategies and (or) to address the associated performance deficits [153]. This issue of fatigue is particularly relevant in risk-intensive and safety-critical task domains such as emergency response, or front-line medical practice, where slips or lapses can have significant consequences. There is an abundance of evidence to suggest that cognitive fatigue impairs critical executive faculties relevant to a wide-range of human executive functions such as situational awareness, working memory, attention capacity and vigilance [108]. Indeed, fatigue has been identified as among the key contributing factors in extraordinary disasters such as the Space-Shuttle Challenger, the Bhopal Gas Tragedy or more recently in the Deep Water Horizon oil spill [107, 154].

The importance of fatigue research is acknowledged widely by the human factors community, and current mitigation efforts largely relate to effective workload management and (or) sustainable shift practices in job domains such as emergency response, medical services or aviation [155]. While these measures are helpful they remain far from sufficient, the practical encumbrances of these guidelines demand consideration on how one could presciently and reliably forecast the onset of these states using unobtrusive/ wearable sensors. Cognitive fatigue, or fatigue as we refer to it for convenience in this work, reflects fundamental changes in neuronal function, a characteristic that is both influenced by the intrinsic motivation that an individual brings to the task at hand [156] and also the workload associated with performing the task itself. Therefore, the search for unambiguous biomarkers that reflect these traits is of interest to the larger community. Clinical researchers have explored this possibility in patients of Multiple Sclerosis who suffer from chronic fatigue [157, 158, 159], a body of work that directs some of our current research. However a larger question related to the experience of fatigue on the task demands further consideration.

There's some ambiguity on the nature of fatigue itself, especially knowing that a task invariant definition for the construct remains far from crystallized. For the purposes of this investigation, we primarily consider the effect of *time-on-task* fatigue, where individuals when engaged in a cognitive task for a long duration of time may experience a loss in productivity (efficiency) and worsening mood typical of an under-arousing and prolonged exercise [61]. Some researchers agree that cognitive fatigue in this context is synergistic with sleepiness states and can further exacerbate performance impairments [160]. A whole body of research operationalizes this definition especially in the context of driving and related drowsiness [161, 162]. In both contexts, the source of fatigue is derived from the task itself either from the associated workload, the duration, or a combination of the two resulting in some loss of productivity and negative outcomes. Studies on vigilance are especially relevant to reaffirm this task-induction hypothesis [20]. Incidentally, it could be argued that under the right circumstances and input conditions, even short duration of a sufficiently complex task may induces a fatigue experience consistent with that due to significant sleep deprivation, a possibility we discuss at length in Chapter 3. Especially a task where the workload and monotony afford little room for the individual to deploy compensatory strategies to recoup. In contrast, a task that is inherently interesting is seen as fatigue-resistant [163].

Given the myriad task-specific factors that influence these states, an operational definition for fatigue which is both sufficient and task invariant remains a challenging prospect [164]. Prior literature presents a myriad neuroscientific evidence [165], performance [166], sleep [167], circadian, [168], general health and task-derived factors that either add to or detract from a clear definition. Further, there is no consensus yet on the best techniques or tools to measure these states, therefore two common approaches are to take either a perception-oriented definition, i.e. subjective self-reports, or a performance-oriented definition, where decrements in performance capacity indexed by response correctness and time together serve as an index of fatigue [169]. The differences here largely originate from the domain where they are applied, example, in the clinical setting fatigue experience of patients predominate the need for care and subsequent prognosis [170], whereas in other settings such as emergency response performance demands take a precedent and therefore a tendency to operationalize fatigue in a performance-driven manner. However, in practice, performance measures are hard to define for most real-world tasks, and hence the ease of self-reports often appeal to convenience over correctness.

For the scope of the work presented in this paper, these distinctions (along performance and perception dimensions) are both interesting and important. In our experiments we found that while participants were fatigued by the task, intervention enabled a performance improvement while the subjective perceptions of fatigue continued to worsen with time [171]. Indicating that performance indicators and perception labels may not align and are susceptible to distortions due to annotator variability, performance saturation, etc. In the present study, we largely characterize fatigue as a consequence of the time-on-task effect, where with extended attention demand and sustained workload individuals are driven towards a regime of parasympathetic withdrawal, lack of motivation, and therefore performance deficiencies. We characterize these states as sub-optimal performance zones and discuss a technique on operationalizing heart variability as a predictor of those states. We expand on the relevance of heart rate variability using the neurovisceral integration model in Section 4.3. In this article we formulate this challenge as one similar to a multi-step ahead time-series forecasting problem (see [172] for a detailed review of techniques), with some evidence pointing to HR/V as a reliable index for fatigue in the clinical context, we investigate if it can serve a similar function under the *time-on-task* effect in a neurotypical population. Furthermore, we study the role of perception, performance, and underlying heuristics that derive from clinical research as indicators of fatigue. We also assess the effectiveness of a sequence to sequence deep learning model in the context of a noisy, low dimensional index such as heart rate variability.

4.2.1 Contributions

The key contributions of this work can be summarized as follows:

- We investigate the reliability of HR-based indices as a neurocognitive index for fatigue. We operationalize a sequence to sequence deep learning architecture to forecast fatigue states as a multistep-ahead, multivariate problem.

- We compare the differences that arise from annotation variability, specifically we consider the effect of performance-based labels, perception-based labels, a weighted-combination of both, and propose a heuristic-based alternative.

The rest of this chapter is organized as follows. In Section 4.3 we introduce background related to the relevance of HR/V as a neurocognitive index for fatigue and related attention regulation. In Section 4.4 we introduce the data, experiment protocol and modeling workflow. In Section 4.5 we present our results, followed by our discussion and conclusions in Sections 4.6 and 4.7 respectively.

4.3 Cardiac Activity as a Cognitive Index

In this section we expand on the rationale behind why cardiac activity has more to offer as a neurocognitive index, specifically relying on the neurovisceral integration model proposed by Thayer et al. [126] along with a brief overview of the current evidence from clinical and non-clinical research in this domain. Individual differences in autonomic balance are long known to be associated with a multitude of cardiovascular health and cognitive dysfunctions [68]. It is understood that vagally mediated heart rate and its variability (HR/V) is linked with neural structures implicated in executive function, and are known to support structures that contribute to several attention and executive functions [173]. Therefore there's sufficient historic evidence to suggest that cardiac activity, specifically, the beat-to-beat variability can serve as an index for efficient and effective cognitive task performance [127].

4.3.1 A Model for Neurovisceral Integration

This perspective is better understood through the neurovisceral integration model (NVIM), which outlines the functional and structural relationships between prefrontal cortex activation and cardiac activity. Specifically, the correlation between autonomic, attention and emotion regulation or dysregulation mediated by the vagus nerve [126]. According to the NVIM, the functions of pre-frontal and sub-cortical inhibitory circuits essential for self-regulation rely on the inhibitory inputs provided by the vagus nerve [174, 175]. Self regulation refers to our ability to control thoughts, emotions and attention behaviors therefore permitting individuals to remain goal-oriented across various environments. In [68], the authors describe several networks relevant to these constructs, and one such agglomeration is referred to as the Central Autonomic Network (CAN). The CAN is responsible for both top-down and bottom-up behavioral responses that are both adaptive and flexible to task demands. The spatial/ structural elements that constitute this network include the anterior cingulate, insular, orbitofrontal, and ventromedial prefrontal cortices, the central nucleus of the amygdala, the paraventricular and related nuclei of the hypothalamus, the periaquaductal gray matter, the parabrachial nucleus, the nucleus of the solitary tract (NTS), the nucleus ambiguus, the ventrolateral medulla, the ventromedial medulla, and the medullary tegmental field [150]; these structures are known to be bi-directional and non-rigid, meaning information can flow both ways either bottom-up or top down and additional resources can be recruited to meet specific task/ environmental demands.

The communication across these networks serve as critical elements of the feed-forward and feedback circuitry that determine attention responses such as response selection or inhibition, and therefore form a central component of the cognitive faculties of interest in this study. Under normal operating conditions the PFC is responsible for staying alert to changing safety cues in the environment by the inhibitory control mechanisms over the sympathoexcitatory subcortical circuits and the amygdala, an organ that is essential to our threat response and also to our susceptibility to distractions. Under threats/ distractions, the inhibitory control is diminished and sympathoexcitatory circuits dominate leading to a default mode threat response. Therefore reduced prefrontal regulation can lead to hyper vigilance type behaviors and perseverative cognition that are know to result in diminished resource capacities over time [175]. We hypothesize that under fatigue states such as the one induced by the *time-on-task* effect, critical signalling pathways across this network may be compromised either due to the imprecise attention scoping of prefrontal circuitry or due to fundamental changes leading to task disengagement, therefore there is a case to be made for the relevance of cardiac vagal activity as an index for task-induced cognitive fatigue.

4.3.2 Clinical and Non-clinical Evidence

Numerous studies with neuroimaging have found relationships between the inhibitory prefrontal-subcortical circuits and cardiac vagal tone, a relationship indexed by vagally mediated resting HRV [101]. Recent findings also suggest that this relationship is tied to the performance of those circuitry such that a higher resting HRV is associated with effective and adaptive responses to environmental demands [176, 177]. In fact, the chaos that underlies the beat-to-beat fluctuations is argued as an essential component of our capacity to adapt to fight or flee scenarios [178]. Therefore the continuous activation of the prefrontal cortex is essential to the optimal, flexible and sustained processing of input information that allows dynamic interaction and adaptation to changing demands of a task or its environment, e.g. as you read the text on this article, hold a conversation, navigate a busy street or respond to stimuli on a game window, and a myriad other examples in our day to day experience. This prefrontal network is also shown to be essential to inhibitory processes, such as ones that allow us to filter distractions and stay goal-oriented across those engagements [179]. Researchers have found a lot of robust relationships between resting state HRV and cognitive functions, such as higher resting HRV exhibit effective behavioral responses (e.g., faster response times and better accuracy) on executive cognitive tasks [180]. In contrast, lower resting HRV is associated with hypoactive prefrontal regulation; this results in hyperactive subcortical structures, which leads to maladaptive cognitive and emotional self-regulation. For example, people with lower resting HRV often fail to recognize safety cues or to habituate to novel, neutral stimuli [180, 178]. Recent evidence also point to these observations beyond resting state, where similar relationships are reported during active task engagement as a function of successive differences in these measures, and over time (see an overview here [135]).

Furthermore, clinical research supports the use of non-invasive HR/V measures as a reliable tool for predicting fatigue severity in chronic fatigue syndrome/ myalgic encephalomyelitis (CFS/ME). With specific observations identifying both time- and frequency-domain parameters that were associated with self-reported cognitive dysfunctions, anxiety, sleep quality, and depression in clinical out-patients [181]. Specifically, Escorihuela et al. report that CFS/ME patients had lower mean RR, SDNN, RMSSD and pNN50 than healthy controls, and lower LF, HF and HFnu but higher LF/HF. Specifically the trends associated with these measures were also important in their diagnosticity. Even though the first set of these HR/V parameters were obtained from the time domain analysis and the second set from the frequency domain analysis, they all indicate that CFS/ME patients showed decreased HR/V and trends, associated with autonomic dysfunction, sleep quality and anxiety/depression symptoms. Other studies on patients with multiple sclerosis also relate to similar ideas on how fatigability or fatigue risk is captured by changes in the temporal/ spectral characteristics of HR/V (e.g. [182, 183]). Therefore there is a need for more research to find transferable evidence in support of HR/V as a fatigue index in non-clinical settings.

In summary, the healthy heart is not a metronome [184], and the beat-to-beat fluctuations of the heart, i.e. heart rate variability, can serve as an information rich index, sensitive to the fatigue states of an individual. The cardiac autonomic network is long known to serve a central role in mediating self-regulation of attention, workload and performance of certain cognitive functions. Critically the CAN mediates cardiac response through the vagus nerve a pathway that is perturbed by fatigue/ workload related constraints. Therefore, operating on this hypothesis, we explore reliable fatigue state estimation using only electrocardiogram (ECG) data.

4.4 Methods

4.4.1 Data set

The data utilized in this study derives from experiments on cognitive fatigue, where 32 participants returned on separate days to complete a fatiguing working memory exercise under three different treatment conditions – control, sham or intervention. Furthermore, the order in which participants were subject to each condition was counterbalanced by partitioning them into three sex-balanced groups. The experimenters relied on a *Latin* square design for this repeated measures study. Participants began each session by responding to subjective questionnaires related to their mood and sleepiness levels on the day, following which they completed a 60-minute visuospatial two back working memory task on a desktop computer (see Fig. 2.1). Each experiment session lasted roughly 60 minutes, with a brief single element subjective questionnaire after every five minutes. The aggregated distribution of participant responses to those questions as a function of time was presented in Fig. 2.4 The average transition time was $\sim 15s$. Participants were not made explicitly aware of the experiment duration and (or) the number of subjective questionnaires involved to avoid biases in their experience. In this present article we only rely on data collected from the control condition of this experiment protocol. In addition to the 32 participants who completed the whole multi-day exercise, we also include data from 7 other individuals who completed the control conditions but not every other session. In the control condition, participants familiar with the activity underwent the task in the absence of peripheral instruments related to the broader context of the experiments reported in [171].

4.4.2 Participants

In this investigation, we recruited a total of 54 participants, 32 among them completed all experiment conditions. We rely on those participants who participated and completed the control mode of the experiment, i.e. N = 39 (23 female) for the reported analyses. All experimental procedures were approved by Texas A&M University's Institutional Review Board (IRB2019-1591DCR), and proceeded in accordance with a strict infection control plan and the ethics code of the American Psychological Association. Participants provided written informed consent before the start of each experiment and were reimbursed for their time.

The median age of the participants was $27 \in [18, 34]$ years. All participants were neurotypical. 34 out of the 39 participants in this study reported good quality of sleep ($>$ 6 hours) in days preceding the experiment session, with a sleep quality rating median score of 7 on a 10 point scale. Here 0 indicated a state of "severe exhaustion" due to lack of sleep, and 10 indicated a state of being "extremely well rested". Roughly 70% (N = 27) of our participants reported a daily activity level of 4000-8000 steps, with 2 reporting activity levels in excess of 10000 steps and the remaining below 4000. Participants reported a median score of 8 on a 10 point scale, with regards to their level of motivation to participate in the study, where 0 indicated "absolutely no motivation" to proceed, and 10 being extremely motivated. In addition, during recruitment, participants were subject to a list of exclusion criteria [185] that included past neurological disorders, use of over the counter medication and caffeine habits.

4.4.3 Working Memory Task

Participants were subject to an hour long visuospatial two-back WM test, while seated in front of a personal computer and provided a keyboard to record their responses. On the task, they tracked the position of a circle within a 3×3 grid. If the position of the circle matched the one from two steps prior, they responded by pushing the space-bar (see Fig. 3.1). The inter-stimulus-interval was $1s$ with a persistence time of $900ms$. The two-back match probability was 0.6; the interface served a fixed number (N = 94) of randomly timed match events within each five minute block. Before starting their first session, participants were introduced to the task and allowed to practice for a minimum of five minutes under a training mode to ensure that they understood task instructions. In this mode, the interface provided textual feedback on their response time and correctness (RED = incorrect, GREEN = correct). During the actual experiments this feedback was withheld; participants began each session on self-indicating their willingness to proceed after the practice period. The interface recorded every key-press and stimulus match event presented to the participant along with a timestamp, response correctness tag (hit, miss, or false alarm), and response time (in ms).

4.4.4 Measures of Interest

4.4.4.1 Cardiac Activity

In these experiments we collected electrocardiogram (ECG) data using a chest worn, two lead amplifier and probe device (Actiheart 4, CamNTech, Inc., UK). Participants were first instrumented with the device to collect a baseline signal and to ensure an acceptable signal-to-noise ratio before beginning each experiment. The electrodes were positioned at the base of the sternum and over the left pectoralis minor muscle. The raw ECG signal was sampled at $128Hz$.

4.4.4.2 Subjective Responses

We query the participants' subjective state along three dimensions, i.e. perceived effort, fatigue and discomfort on an ordinal scale (1 - 10). Participants respond to these questions every five minutes during the 60 minute exercise, therefore each individual provides 12 samples of their subjective state in this protocol. In our present investigation, the fatigue responses are operationalized to serve as a perception-based index of fatigue.

4.4.4.3 Task Performance

On the task, we had different performance measures that qualified the participant's task behaviors. In the present context we rely on accuracy as the sole performance index in our experiments, which is defined as defined as the ratio of hits + correct omissions to hits + correct omissions +

misses + false alarms. In these experiments, a hit is when the user correctly responds to a target sequence, a miss is when the user fails to respond to a target sequence, a false alarm is when the user incorrectly responds to a non-target sequence, and correct omissions are events where the user withholds their response for non-target sequences. A target sequence is a two-back match event. The overall performance trend and distribution of participants included in this investigation is shown in Fig. 4.3.

4.4.5 Feature Extraction

Cardiac electrical activity was obtained from the ECG probe and amplifier interface at $128Hz$. The raw ECG signal was filtered for motion-related artifacts [73], and corrected for ectopics with polynomial interpolation [74]. Subsequently, a peak detection algorithm was used to isolate the *R* peaks from the ECG signal [75]. The time between successive *R-R* peaks, i.e. the inter-beat-interval ($\mathbf{X}_{IBI}$) or normal-to-normal (NN) interval was then derived from the processed peak signals for subsequent feature engineering. The corrected inter-beat-interval ($\mathbf{X}_{IBI}$) obtained from the continuous ECG signal was subject to participant-level feature-scaling (*min-max* normalization) and were then partitioned into time windows of $300s$ with a step size of 1. Temporal, spectral, and nonlinear features were then extracted from each window resulting in 50 features for heart rate and its variability (HR/V) using the Neurokit2 library [187]. The non-linear features were further categorized as measures of asymmetry [188], fragmentation [189], complexity and geometry [184] (see *Table 4.1* [186] © 2022 IEEE). With this technique, the composite HR/V feature matrix $\mathbf{X}_{HRV} \in \mathbb{R}^2$ had on average the following dimensions $[\sim 560 \times 50]$. *Subjective measures*, i.e. the fatigue response scores were preserved in the original scale, but min-max normalized across participants so that values were $\in [0, 1]$ with $\mathbf{X}_{sub} \in \mathbb{R}$ of shape $[12 \times 1]$. *Performance measures* were sampled on the interface as response omission and commission behaviors, we aggregated binary scores associated with each target sequence, i.e. do you have a correct response or not? over every $30s$ window to derive accuracy as defined in the preceding section, therefore for the roughly 60 minute exercise we derived an accuracy vector $\mathbf{X}_{perf} \in \mathbb{R}$ of shape $[120 \times 1]$ for each participant.

Table 4.1: Short-term heart rate variability (HR/V) features for each window ($60s$) of the filtered ECG signal. Reprinted with permission from [186] © 2022 IEEE

Temporal		
Measures of central tendency	RMSSD	RMS of successive differences (SD) of NN intervals
	SDNN	Standard deviation of NN
	SDSD	Standard deviation of SD
	MeanNN	Mean of NN
	pNN50	Proportion of intervals $> 50ms$
	MadNN	Median absolute value of SD
	CVNN	SDNN/MeanNN
	CVSD	RMSSD/MeanNN
	TINN	Base width of NN histogram
	HTI	HRV triangular index
	IQRNN	Inter-quartile range of NN
Spectral		
Power density	ULF	Ultra low (0 – 0.0033Hz)
	VLF	Very low (0.0033 - 0.04Hz)
	LF	Low (0.04 – 0.15Hz)
	HF	High (0.15 – 0.4 Hz)
	VHF	Very high (0.4 – 0.5 Hz)
Nonlinear		
Poincaré geometry	SD1,2	Transverse and longitudinal variability of poincaré plot
	CVI	Cardiac vagal index
	CSI	Cardiac sympathetic index
Asymmetry	SD1d, a	Short-term deceleration (d) and acceleration (a)
	C1d/C2d C1a/C2a	Contribution of HR *d*, *a* on short-term, and long-term HR/V
	SDNNd/a	Total variance from *d* or *a* changes
Fragmentation	PIP	Percentage of inflection points
	IALS	Inverse of average length of *a*, *d* segments
	PSS	Percentage of short segments
	PAS	Percentage of NN intervals in alternating segments
Complexity	ApEn	Approximate entropy
	SampEn	Sample entropy

* NN = Normal to Normal interval (ms); F = frequency (Hz).

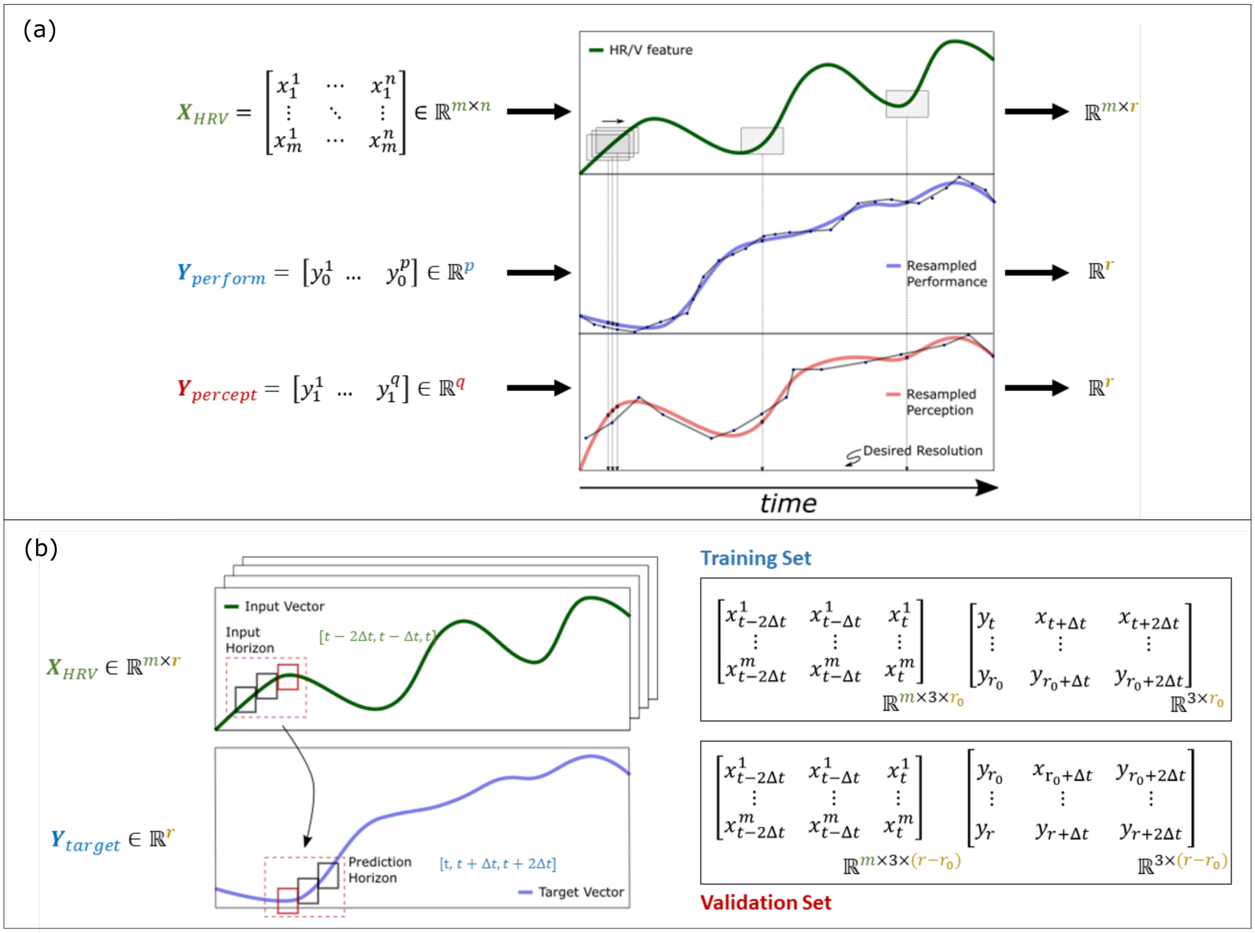

Figure 4.1: Schematic representation of the resampling technique and the input feature vector and targets of the mulitvariate, multi-step ahead forecasting problem. (a) Input feature vector and target vectors are resampled to a common resolution, resulting in r samples across all participants. (b) The input multidimensional feature vector and target vector are split into input and output sequences with a forecast and prediction horizon of size 3 with r_0 observations in training, and $r - r_0$ observations in the validation set.

4.4.6 Feature Representation

The feature representation problem is unique to this specific work given the need to contrast how different annotation techniques can lend to the underlying predictive power of the HR/V signal. We rely on a multidimensional input feature space i.e. $\mathbf{X}_{HRV} \in \mathbb{R}^2$ $[\sim 560 \times 50]$, and target vectors $\mathbf{X}_{perf}$ and $\mathbf{X}_{perf} \in \mathbb{R}$ of shape $[12 \times 1]$ and $[120 \times 1]$ respectively. We rely on a feature representation method as depicted in Fig. 4.1 (a), where every measure is re-sampled to a common resolution. During re-sampling, we aggregate over (down-sample) the higher resolution signal using a fixed sampling horizon and nearest neighbor mapping. When up sampling the lower resolution signal, i.e. subjective fatigue scores and accuracy, we consider the best fit second-order spline on the raw data and rely on a nearest neighbor map to the desired time stamp. These methods are pictorially represented in Fig. 4.1 (a). With the re-sampling techniques described here, every feature and target vector now share a consistent dimension i.e. $[\sim 500, n]$ where n = 50 for HR/V measures and 1 for the target vectors. All subsequent analyses rely on this data format as the base resolution. For training the models of interest in this study, the re-sampled observations where further split into input and output sequences of fixed lengths (see Fig. 4.1 (b)), while we investigate the effect of varying the prediction horizon and input vector dimensions on forecasting errors, the baseline problem is of *multi-step ahead* forecasting, for which we consider an input vector horizon consisting of three observations say $[t - 2\Delta t, t - \Delta t, t]$ and an equivalent forecast horizon of three observations $[t, t + \Delta t, t + 2\Delta t]$ where Δt, the step size, depends on the feature resolution and $\in [30,\ 300]\ s$. In the sections to follow we elaborate on our workflow and algorithms of interest.

4.4.7 Problem Formulation and Techniques

The mutlivariate time-series of interest here can be described as below. We have a two-dimensional vector $\mathbf{X}_{HRV} \in \mathbb{R}^2$, containing 50 features of HR/V (see Table 4.1), where each feature $x_i \in \mathbb{R}$ and $i \in [1,\ 50]$. Therefore, we can represent the input feature space as,

$$\mathbf{X} = \begin{bmatrix} x_{11} & \dots & x_{1t} \\ x_{21} & \dots & x_{2t} \\ & \vdots & \\ x_{n1} & \dots & x_{nt} \end{bmatrix}$$

A segment of this vector can be referred to as $\mathbf{X}_{t-p}^{t}$, where we have p elements or steps in the input feature space such that the first feature would be $[x_{1t-p}, ..., x_{1t}]$ and so on. Now for the multi-step ahead formulation of the forecasting problem, the target vector for this segment with a forecast horizon of h steps would be $\mathbf{Y}_t^{t+h} = [y_t, y_{t+1}, ..., y_{t+h}]$. The target vector here could be either perception-, performance- or heuristic-based depending on the type of data involved. We consider four multi-step ahead forecasting methods to contrast across these annotation techniques. Specifically, the statistical multiple output multivariate linear regression, the nonlinear support vector regression method, the ensemble gradient boosting regression trees, and finally a sequence to sequence deep learning architecture that uses Long Short Term Memory.

Multi-output Multiple Linear Regression

In multiple linear regression, we try to find the optimal combination of coefficients using multiple predictor variables, therefore here for the single-step case the problem can be defined as $y_t = \beta_0 + \beta_1 x_{1t-p} + ... + \beta_n x_{1t} + \epsilon$, where, β are the coefficient terms, and ϵ is the model error. The effective optimization problem now reduces to one of minimizing the least squares cost for each output $\mathbf{Y}_j$ where $j = [t, t+1, ..., t+h]$, i.e.

$$\underset{\mathbf{w}}{minimize} \ \frac{1}{P} \sum_{p=1}^{p} (\mathbf{X}_{t-p}^{t})^T \mathbf{w} - \mathbf{Y}_t^{t+h}$$

Support Vector Regression

Given training vectors $\mathbf{x}_i \in \mathbb{R}^p$, $i = 1, .., n$, and a target vector $\mathbf{y} \in \mathbb{R}^n$ ϵ-Support Vector Regression solves the following primal problem:

$$\begin{aligned} &\min_{w,b,\zeta,\zeta^*} \tfrac{1}{2} w^T w + C \textstyle\sum_{i=1}^{n} (\zeta_i + \zeta_i^*) \\ &\text{subject to } y_i - w^T \phi(x_i) - b \leq \varepsilon + \zeta_i^* \\ &\qquad w^T \phi(x_i) + b - y_i \leq \varepsilon + \zeta_i^* \\ &\qquad \zeta_i, \zeta_i^* \geq 0, i = 1, \ldots, n \end{aligned}$$

The expectation is that we penalize samples whose prediction is at least ϵ away from true target. Furthermore, the objective is penalized further by ζ_i^* or ζ_i^* depending on whether the prediction lies above or below the error bound. The mutlioutput and multivariate extension of the above optimization problem is detailed in [190].

Gradient Boosting Regression Tree

In gradient boosting regression, given a sample set with input variables say $\mathbf{x} = [x_{1t-p}, ..., x_{1t}]$ and corresponding output variable $\mathbf{y} = [y_t - p, ..., y_t]$, the objective is to find an optimal mapping $F(\mathbf{x})$ that minimizes the loss $L(\mathbf{y}, F(\mathbf{x}))$. GBRTs arrive at an optimal solution gradually through the weighting of some intermediate function $h(\mathbf{x})$ that minimizes this loss. Here, the function $h(\mathbf{x})$ is a basic regression tree derived by combining the input variables in the feature space and the negative gradient of the loss function in the previous model, therefore, the GBRT starts with some constant function $F_0(\mathbf{x})$ and builds the model in a greedy way. The model is formulated as below:

$$arg\ \underset{\gamma}{min} \sum_{k=t-p}^{t} L(y_k, \gamma)$$

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \gamma_m h_m(\mathbf{x})$$

Where $F_m(\mathbf{x})$ is the integral over prediction values of the basic regression trees. $h_m(\mathbf{x})$ is the m^{th} regression tree and γ_m is the coefficient of the m^{th} regression tree. This specific formulation of this model is for the univariate, single-step case. Extension towards the multivariate and multi-output formulations are explained in detail here [191].

Long Short Term Memory

The proliferation of deep learning architectures have considerably bolstered the time-series forecasting problem space. Specifically, Recurrent Neural Networks [192] enable neural networks to persist short-term dependencies in the input feature space, however they suffer from issues such as vanishing or exploding gradients that can skew the optimality constraints needed for the given application. Long Short-Term Memory networks overcome this problem of exploding or vanishing gradients by allowing the model to persist long-term dependencies in the input feature space [193]. An LSTM network is comprised of memory cells with self-loops (see Fig. 4.2). The self-loop allows it to preserve temporal information encoded on the cell's state. The flow of information through the network is handled by writing, erasing and reading from this memory state. These operations are handled by three constituent gates i.e the input, output and forget gates. The cell operation is formulated as below:

$$\mathbf{i}_t = \sigma\left(\mathbf{x}_t\mathbf{U}^i + h_{t-1}\mathbf{W}^i\right)$$
$$\mathbf{f}_t = \sigma\left(\mathbf{x}_t\mathbf{U}^f + \mathbf{h}_{t-1}\mathbf{W}^f\right)$$
$$\mathbf{o}_t = \sigma\left(\mathbf{x}_t\mathbf{U}^o + \mathbf{h}_{t-1}\mathbf{W}^o\right)$$
$$\tilde{\mathbf{C}}_t = \tanh\left(\mathbf{x}_t\mathbf{U}^g + \mathbf{h}_{t-1}\mathbf{W}^g\right)$$
$$\mathbf{C}_t = \sigma\left(\mathbf{f}_t * \mathbf{C}_{t-1} + \mathbf{i}_t * \tilde{\mathbf{C}}_t\right)$$
$$\mathbf{h}_t = \tanh\left(\mathbf{C}_t\right) * \mathbf{o}_t$$

where $\mathbf{i}_t$, $\mathbf{f}_t$, and $\mathbf{o}_t$ refer to the input, forget and output gates at time t, respectively. $\mathbf{x}_t$ and $\mathbf{h}_t$ are the number of input features and number of hidden units. $\mathbf{W}$ and $\mathbf{U}$ are weight matrices with a bias. σ is the sigmoid activation function. The architecture of a single LSTM cell is represented in Fig. 4.2 (a) (adapted from [194]). The input gate decides if the driving signal should modify the memory state or not, this decision is made with the help of a sigmoid activation function. Now, going back to the parent problem, where we have an input vector $\mathbf{X}_{HRV} \in \mathbb{R}^2$ and a target vector say $y \in \mathbb{R}$, the network relies on a back-propagation through time (BPTT) [195] algorithm to optimize the cost function defined as below:

$$L = \sum_{t=1}^{M}(y_{[t]} - \hat{y}_{[t]})^2$$

where, $\hat{y} \in \mathbb{R}$ is the forecasted target value at some future time step.

Sequence to Sequence LSTM

The Sequence to Sequence LSTM (LSTM*) architecture was developed to forecast multiple output sequences given an input sequence of varying lengths, primarily for its use case in NLP problems [196]. However, the architecture lends itself well to the related problem of multi-step ahead time-series forecasting. The architecture consists of two LSTM networks – an encoder and

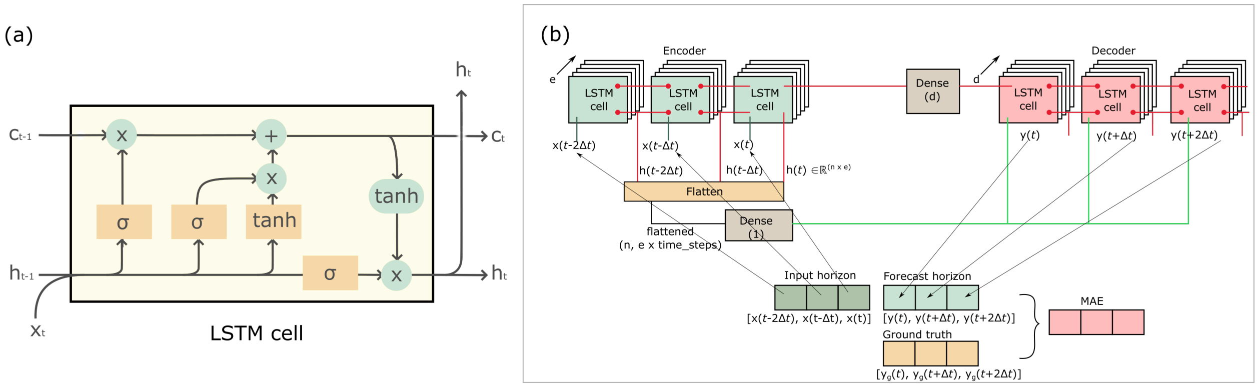

Figure 4.2: (a) LSTM cell adapted from LSTM Cell by *Guillaume Chevalier* licensed under Creative Commons Attribution-Share Alike 4.0 International license [194]; (b) *Seq2Seq* LSTM architecture adapted from [197]

a decoder (see Fig. 4.2 (b); adapted from [197]). The encoder converts input sequences into a fixed length vector, which then serve as the input state for the decoder. The decoder then generates an output sequence of length h, such that h is the desired forecast horizon. At it's core the problem now involves optimizing two cost functions similar to the optimization function for the LSTM, one for the encoder and the other for the decoder. Notably, the Seq2Seq architecture permits variability in the input feature length, while this is not a characteristic we utilize in the current effort it does permit the prospect of accommodating asynchronous sources. This can be advantageous considering the difference in sampling frequency across ECG, subjective and performance indices in this study. In our investigation, we also include a simple fully-connected neural network (FC-NN) to differentiate how the recurrence introduced by the LSTM* could improve forecast accuracy.

4.4.8 Annotating Fatigue States

A primary contribution in this work relates to the ground truth for cognitive fatigue. This is an important consideration given subjective variability in response to surveys, the differential sensitivity of single-element questionnaires [198] and the variability in response to performance demands during the working memory exercise [199]. Furthermore, evidence from our work reported in [171] shows that in task contexts such as the one used in this investigation, performance and perception measures do not always align, an observation that has been reported since the early days of fatigue research e.g. [200, 201]. We also found differential outcomes on performance behaviors contingent on the baseline performance capacity of the individuals studied which adds to the uncertainty on the expression of fatigue symptoms while engaged in this task. However, we know from our pre-post mood assessment questionnaires that all participants, regardless of baseline performance distinctions or reported perceptions during the task, experienced worsening mood along every relevant dimension (see Fig. 2.4). In fact the magnitude change of the total mood disturbance and fatigue-scores reported by the participants is consistent with a clinically significant change in fatigue perceptions as found in other studies [90]. Therefore, we are confident that the exercise is fatiguing, now the question is how do we capture those states of fatigue during the task as we hypothesize that this is not a binary state change and is in fact a gradient over the time-course of the participant's experience. A viewpoint that is supported by the evidence reported in prior literature e.g. [138, 59] and also the trends evident in the self-report scores (see Fig. 4.3).

Furthermore, our end goal is for such systems to enable a closed-loop countermeasure to those states of fatigue before their perceptual relevance to the human in the loop, with this vision, another constraint stems for our observation that interventions did produce an operationally significant change to performance outcomes but provided no change to perceptions [171]. This finding takes on an interesting form when we consider our takeaways from stakeholder interviewers, where several emergency responders shared the opinion that performance improvements were more important than perceptions in the field. Leading to the present question – what is the most helpful operational definition for cognitive fatigue states and can we reliably forecast these states when relying on a low-dimensional signal such as heart rate and its variability. To address these questions and to speak to the relevance of heart rate as a cognitive index we consider the annotation strategies as described in the following sections.

4.4.8.1 Perception-based Annotation

Here self-reported fatigue scores, derived from the ten-point scale responses serve as one among the ground truth labels for fatigue states and therefore a target variable of interest. Several works in the broad domain of affective computing have reliably shown the physiological estimation

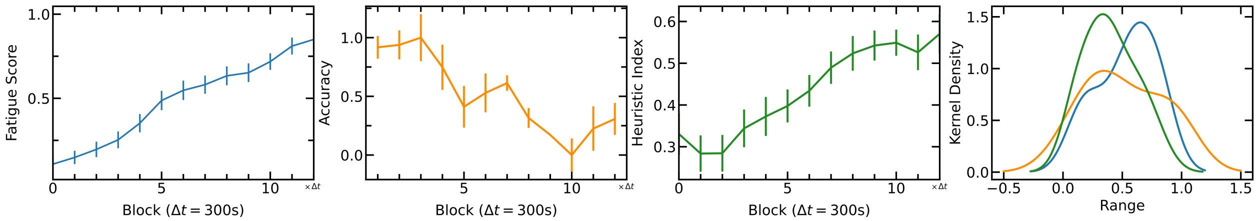

Figure 4.3: Annotation feature trend and distributions for perception-based, performance-based and heuristic-derived techniques aggregated across all participants. Twelve data points are reported where each observations corresponds to a time window (Δt) of $300s$. At each point we present the mean and its standard error (SE). The Kernel Density Estimate plot depicts the cumulative probability density function for each annotator.

of perceived emotions across multiple dimensions including stress [186], mental workload [202], valence among others using a similar technique. Therefore the construct of relying on subjective self-reports in qualifying states of affect is a tried and tested one, yet it does not guarantee whether the self-reports are well-aligned with physiological representations of the phenomena that result in them or the a-priori expectations of an experimental investigation. Another challenge relates to the sparse temporal resolution afforded by these indices, e.g. 12 data points in a 60 minute exercise, where it's not feasible to disrupt the experiment paradigm too often to collect this information.

In Fig. 4.3 we observe a worsening trend in fatigue response as a function of time, an observation that is discussed at length in preceding chapters. Notably, we found statistically significant changes in self-reported fatigue scores after every five minute interval and also that the post-experiment mood-disturbance score on the POMS survey indicated a significant worsening of fatigue perceptions over time. Therefore, with this evidence, we treat the perception of fatigue as a high-confidence indicator of subjective state, however evidence to suggest that heart-rate variability can index these changes across individuals is yet to be established, especially given the intrusive nature by which we sample these data which could permit compensatory mechanisms and recovery during the brief transition periods.

4.4.8.2 Performance-based Annotation

Here, we consider performance accuracy over the defined time window as a proxy indicator of fatigue states. Several studies have argued in favor and against the use of performance as an index for cognitive fatigue given the mixed responses elicited across study designs. A running concern relates to factors like individual variability, the ceiling effect in some cognitive tasks, where participants after some training period are able to maintain performance outcomes with ease, and conversely the lack of proper training exercises or performance stability on the task as a confound. These were equal concerns during the study development for us too, notably as shown in Fig. 4.3, overall we found that performance accuracy decreased as a function of time, however, we did find some learning behaviors early on in the exercise and some anticipatory confounds during the terminal stages of the experiment which could distort the reliability of performance as a ground truth indicator of fatigue. Regardless, given the degree of explainability performance outcomes afford and their operational relevance to faculties in the real world, performance-based labels merit consideration as a fatigue index.

4.4.8.3 Combining Perception and Performance-based Indices

In addition to exploring the independent effects of performance and perception based indices. We propose a weighted linear combination of the normalized perception and performance scores as a measure of fatigue on the task. We also investigate the effect of altering the weighting coefficients used. This naive approach has been tested in prior studies that look to operationalize machine learning models for affect detection [203]. Where a weighted aggregation of expert scores that annotate facial affect are used to improve the performance of a classifier. This method, although simplistic, was shown to alter the target function sufficiently to allow improved learnable representations in those experiments. We extend this consideration towards the present fatigue problem, while also discussing the relative information gain by biasing the index towards either perception or performance. The combined fatigue index **Y** can be represented as:

$$\boldsymbol{Y}_{[n\times 1]} = \mathbf{Y}_{\mathbf{tar}[\mathrm{n}\times 2]} \times \begin{bmatrix} 1-\gamma \\ \gamma \end{bmatrix}$$

Where $\gamma \in [0,1]$ is the bias factor that decides if the weight is skewed towards performance or perception. $\mathbf{Y}_{tar} \in \mathbb{R}^2$ is the performance and perception target vector, and $\mathbf{Y}$ is the aggregated/ combined fatigue score.

4.4.8.4 Cardiac Heuristic Fatigue Index

Here we rely on existing expectations from clinical research on cognitive fatigue in identifying fatigue states using heart variability. Specifically, we consider the following: a reduced heart rate variability trend, relative to baseline and preceding observations, increases the likelihood of fatigue, as does an increase in LF power density over time, a known index of parasympathetic dominance, as informed by the NVIM. Therefore, we define a heuristic-derived fatigue index using the expression below that combines RMSSD ($\mathbf{X}_{rms}$), pNN50 ($\mathbf{X}_{pnn50}$), and LF-power ($\mathbf{X}_{LF}$) indices such that the target vector $\mathbf{Y}$ is now formulated as:

$$\boldsymbol{Y}_{[n\times 1]} = f((1 - f(\Delta\mathbf{X}_{rms})) + (1 - f(\Delta\mathbf{X}_{pnn50})) + f(\Delta\mathbf{X}_{LF}))$$

where, $\Delta\mathbf{X} = \mathbf{X}_t - \mathbf{X}_{t-1}$, i.e. successive differences, and the $f(x)$ is $min - max$ normalization such that:

$$f(x) = \frac{x - min(x)}{max(x) - min(x)}$$

Interestingly, the sample trend of the heuristic index largely aligns with the observed trend in the self-reported fatigue scores. With significant differences in means between the terminal (50-60 minutes) and early (10-20 minutes) stages of the experiment ($p < 0.0001$; see Fig. 4.3).

4.4.9 Model Training, Evaluation, and Planned Analyses

In this paper, we evaluate the proposed model and annotation strategies to discus the opportunities and relevance of HR/V as a neurocognitive index for fatigue. In our comparison we rely on Mean Absolute Error (MAE) as the metric for model performance. MAE can be defined as:

$$MAE = \frac{\sum_{i=1}^{n} |e_i|}{n}$$

where, $|e_i| = |y_{ip} - y_{it}|$ is the absolute error between the predicted value y_p and target y_t. We conduct experiments to assess model performance across the following considerations – **(1)** choice of architecture, i.e. performance gains and deficiencies across statistical methods, traditional machine learning approaches, a single layer fully-connected neural network, and deep learning based methods; **(6)** the training paradigm, i.e. the differences in building individualized models against generalized models where sequence data is aggregated from across all participants during training; **(3)** the choice of annotation stategy or fatigue index, we rely on the same independent feature space, i.e. HR/V features but assess the relevance of performance, perception, and heuristic-based annotation strategies across every model of interest, we persist with the best performing model for subsequent analysis; **(4)** we investigate the boundaries of the forecast horizon, i.e. the number of steps into the future that can be reliably forecast by the trained model, and **(5)** the history of data or the input horizon and its impact on model performance; **(6)** we also study the effect of input feature resolution or the window-size used for aggregation across the HR/V feature space; and finally, **(7)** we comment on annotator noise or distortion by exploring a method of combining labeling strategies on forecasting outcomes. While this was not intended to be an exhaustive exercise, these specific considerations allow us to unpack and discuss potential use cases and variables of interest for our future efforts.

4.5 Experimental Outcomes

In this section we describe our observations across every planned experiment. We provide additional detail to contextualize how certain comparisons were drawn including information related to model training, tuning, and performance evaluation. Regardless of the modeling strategy or relevant variables, we used 10-fold, stratified, cross validation on the data, with 75% of participants in training, and the remaining as a "hold out" test set in each fold to avoid the bias of highly correlated samples coming from the same group of participants [204]. We optimized for MAE by tuning a subset of hyper-parameters for each model using a grid-search process. We report the outcomes of this process under the bench-marking experiments, the optimization parameters for each model are documented in Table 4.2. Unless otherwise indicated models were trained on individualized data, such that a unique model is developed for each individual in our sample. All HR/V features were derived at the limiting window size of $300s$ which is widely acknowledged to be the correct temporal resolution to derive the features of interest in this study [184]. There were 50 HR/V features in all including mean HR and its standard deviation across each window (see Table 4.1), we retain all 50 features in training across methods except for the heuristic annotation sub-case where we explicitly avoid $RMSSD$, $pNN50$, and LF power from our feature set given that these measures are relevant to the development of that annotation index. Furthermore, the step-size used for sampling the features of interest was $30s$, although we study this as a variable in one among our experiments. The MLR, SVR and GBRT algorithms were trained using the scikit-learn libraries API, while the FC-NN, and LSTM architecture were built using *PyTorch* with the *CUDA* API. All models were trained on compute resources using a Google Colab Pro account.

4.5.1 Model Comparison and Bench-marking

In our model comparisons we observed that the *Seq2Seq* LSTM (LSTM*) architecture outperformed other methods, regardless of the annotation strategy considered (see Fig. 4.4). The MAE for the LSTM* model was 0.071$\pm$0.005 across every annotation strategy of interest and ranged $\in [0.053, 0.102]$, with forecasting error found to be comparable across both perception indices and heuristic methods, while fairing relatively poorly when relying on performance as an index for fatigue (see Table 4.3). In contrast, we found that the classical MLR, the non-linear SVR and GBRT had relatively comparable MAEs when relying on performance and heuristic-based indices, although each method faired poorly relative to the LSTM*. Interestingly, the GBRT algorithm per-

Table 4.2: Optimization parameters of interest for the models under consideration

Model	**Parameters**
Regularized MLR	Solver = 'auto'; α = 0.8
SVR	rbf; γ = scale; C = 0.5; ϵ = 0.1
GBRT	n_estimators = 300, criterion = gini, max_depth = 50, min_samples_split = 2, min_samples_leaf = 5, max_features = 'auto', bootstrap = False
FC-NN	input_dim = 50, num_layers = 1, activation = 'relu', lr = 0.02, loss = L1, epochs = 500 , batch_size = 50, optimizer = adam
LSTM*	input_dim = 50, encoder_hidden_layers = 2, decoder_hidden_layers = 2, activation = 'relu', lr = 0.001, loss = L1, epochs = 500 , batch_size = 50, optimizer = adam

formed comparably when relying on perception indices with MAE $\in [0.071, 0.080]$. The disparity between performance-based annotation relative to both heuristic and perception-driven methods was apparent regardless of the modeling technique used, with perception-based indices resulting in roughly $10 - 15\%$ poorer model performance regardless of the technique used. For subsequent discussions and comparisons of interest in this paper we relied on the LSTM* architecture as the baseline model given the reliable performance seen across each annotation technique. The explicit performance trends for this architecture across each annotation technique is shown in Fig. 4.7.

Training Paradigm and Evaluation

Here we contrast the approach to model training and evaluation between individualized or generalized techniques. In the previous section on bench-marking, we report observations from a training strategy where a unique model is developed for each individual, by splitting the feature space into input and output sequence pairs for the desired three step ahead forecasting window. Here the feature vector is split into test and train segments while preserving the integrity of the chronological sequence. In this section, we compare the efficacy of the aforementioned approach

Table 4.3: Model comparison and benchmark evaluation across three annotation methods.The table presents the MAE $\pm$ SE for each method for a three step forecast horizon where $\Delta t = 30s$.

	MLR			SVR		
	t	$t+\Delta t$	$t+2\Delta t$	t	$t+\Delta t$	$t+2\Delta t$
Perception	0.107±0.009	0.109±0.009	0.109±0.009	0.097±0.007	0.098±0.007	0.101±0.007
Performance	0.149±0.009	0.142±0.008	0.139±0.009	0.153±0.008	0.144±0.007	0.141±0.008
Heuristic	0.111±0.004	0.120±0.004	0.124±0.004	0.118±0.004	0.125±0.004	0.130±0.005
	GBRT			**LSTM***		
	t	$t+\Delta t$	$t+2\Delta t$	t	$t+\Delta t$	$t+2\Delta t$
Perception	0.071±0.008	0.078±0.009	0.080±0.009	0.065±0.004	0.059±0.004	0.061±0.004
Performance	0.160±0.010	0.159±0.009	0.143±0.008	0.081±0.004	0.094±0.004	0.102±0.005
Heuristic	0.132±0.005	0.132±0.005	0.142±0.006	**0.053±0.005**	**0.056±0.005**	**0.064±0.006**

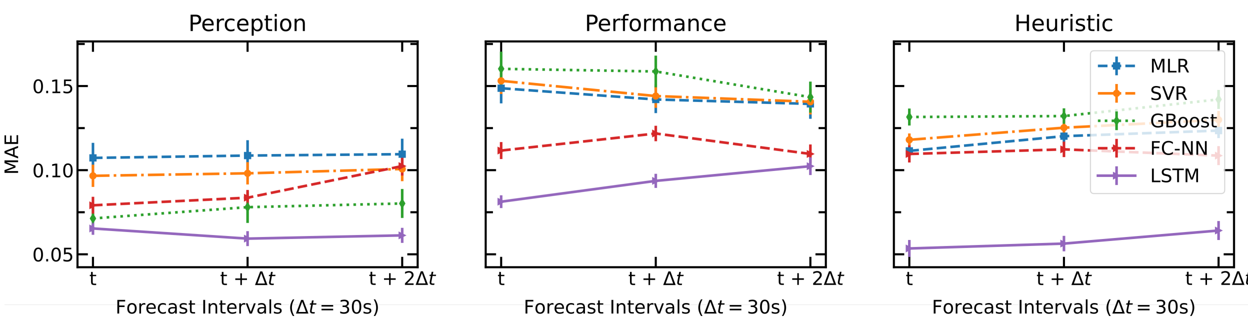

Figure 4.4: Model comparison and benchmark evaluation across three annotation methods. We rely on the Mean Absolute Error (MAE) to contrast the regularized Multiple Linear Regression (MLR), Support Vector Regression (SVR), Gradient Boosting Regression (GBR) models and the Seq2Seq formulation of the LSTM network. We consider a forecast horizon of three with $\Delta t = 30s$

to a paradigm where the input output sequences of all participants is aggregated into one large training set to develop what we call a generalized model, i.e. the network weights are optimized across multi-participant data in training before being evaluated on a hold-out set consisting of previously unseen participant data. The fundamental problem remains the same, i.e. to forecast fatigue states across each annotation technique. Our observations from this effort is summarized in Fig. 4.6 and Table 4.5. We note that the generalized approach fairs poorly across every annotation method relative to the personalized model development outcomes reported in Table 4.4 for the same

Figure 4.5: MAE for LSTM* under a generalized training paradigm for each annotation technique.

	Effect of Generalization in Training		
	t	$t + \Delta t$	$t + 2\Delta t$
Perception	0.295 ±0.168	0.297±0.163	0.299±0.172
Performance	0.378±0.162	0.340±0.225	0.320±0.137
Heuristics	0.099±0.065	0.134±0.086	0.156±0.097

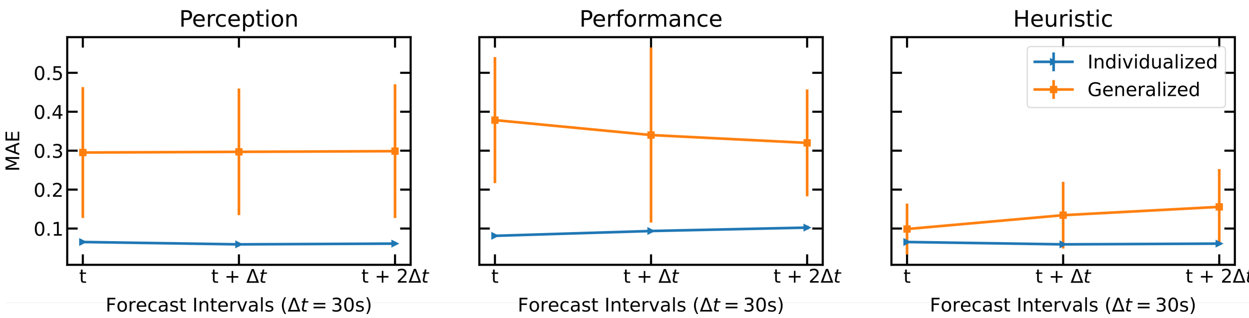

Figure 4.6: Comparing the outcomes of training paradigm: individualization vs. generalization, where data is aggregated from across all participants on the 3-steps ahead forecasting problem for each annotation technique.

architecture. The error margin is significant with roughly an order of magnitude difference for both perception and performance-based annotation methods under the generalized training approach.

Performance-Perception as a Combined Index

Here we introduce the observations from creating a combined index for fatigue using both performance and perception-based indices as introduced in Section 4.4.8. We found that the weighted combination of perception and performance enabled marginal improvements in forecast error, and fared better than when relying on perception/ heuristic indices alone. Our observations on this are summarized in Fig. 4.4 and Table 4.4. Here we also explore the effect of biasing the fatigue annotation index towards either the performance index or perception-index to learn where (if any) the outcome improvements were coming from. We observed that as the bias factor was increased towards perception, greater improvements were made in the MAE, but these benefits were seen to largely plateau or diminish beyond a bias (γ) value of 0.6 (see Fig. 4.7).

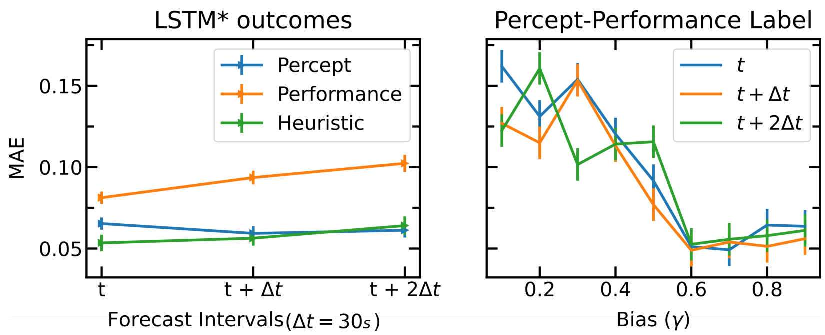

Figure 4.7: Performance outcomes of the LSTM* architecture across each annotation strategy with a 3-step ahead forecast horizon where $\Delta t = 30s$ (left); LSTM* architecture performance as a function of bias (γ) on the combined annotation strategy and for the same forecast horizon.

Table 4.4: LSTM* outcomes on the naive weighted linear combination of perception and performance indices as fatigue labels. With the bias (γ) toward perception-index (see Fig 4.7); SE $\in \pm[0.001, 0.007]$.

	Percept-Performance Annotation								
Bias (γ)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
t	0.162	0.131	0.154	0.120	0.092	0.051	0.061	0.068	0.087
$t + \Delta t$	0.127	0.115	0.153	0.113	0.077	0.049	0.054	0.051	0.056
$t + 2\Delta t$	0.122	0.161	0.102	0.114	0.116	0.053	0.056	0.058	0.061

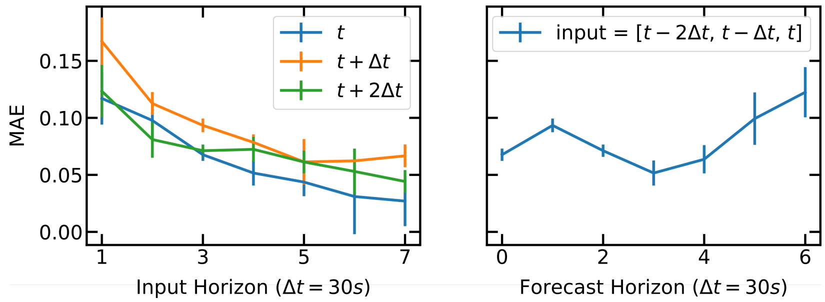

Figure 4.8: Lead and lag horizon effects on the performance outcomes of the LSTM* architecture when using the heuristic annotation technique. Varying the input vector horizon in steps of $t + \mathbf{n}\Delta t$ for the 3-step ahead forecasting problem (left); Observing the effect of increasing the forecast window in steps of $t + \mathbf{n}\Delta t$ for a fixed input horizon of three (right).

4.5.2 Lead and Lag Horizon Tuning

In Fig. 4.8 and Table 4.5 we introduce our observations from varying the input and output sequence lengths, or the input vector horizon and the forecast horizon respectively. We found that input vector horizon had a significant impact on forecast outcomes such that as the input sequence was increased in steps of $\Delta_t = 300s$ the forecast error was found to decrease as seen in the related plots. An observation which was found to be true for the entire three-step forecast horizon. The information gain and corresponding improvement in MAE was found comparable across all time-steps. Notably, here we only rely on the heuristic-derived annotation strategy for both experiments. On the lag horizon tuning, we found a limiting effect in terms of the reliable forecast window enabled by a fixed input sequence length i.e. $n = 3 : [t - 2\Delta t, t - \Delta t, t]$. We found that as the forecast horizon increased beyond $3\Delta t$ MAE worsened with a change of roughly 60% from the best outcome window. However, values were still comparatively lower than those seen in other annotation contexts.

4.5.3 Heart Rate Variability Feature Aggregation Resolution

In Fig. 4.9 and Table 4.6 we discuss the influence of feature resolution on modeling outcomes. Specifically, we explore the step-size t_{st} of the sliding window sampling method applied to the input feature space and its impact on the LSTM* architecture's ability to forecast fatigue indices

Table 4.5: Lead horizon/ Input Vector Horizon impact on the forecast horizon of the LSTM* architecture when using the heuristic annotation technique. SE $\in \pm[0.007, 0.033]$.

	Input Vector Horizon = $t + \mathbf{n}\Delta t$						
n	0	1	2	3	4	5	6
t	0.117	0.098	0.068	0.052	0.044	0.031	0.027
$t + \Delta t$	0.167	0.113	0.093	0.078	0.061	0.062	0.067
$t + 2\Delta t$	0.123	0.081	0.071	0.072	0.061	0.053	0.044

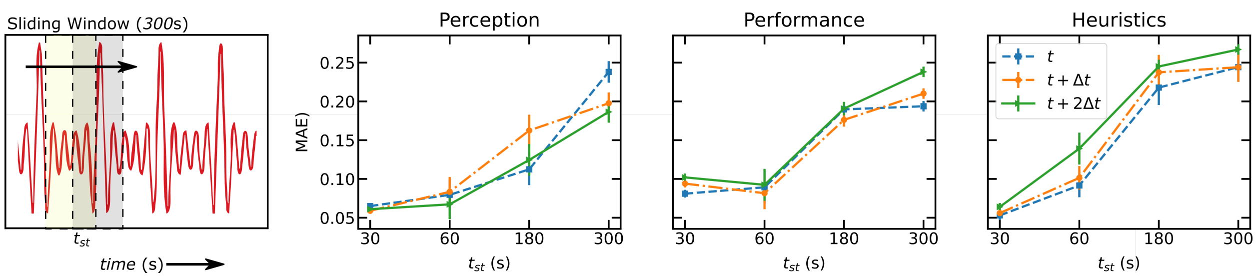

Figure 4.9: Effect of HR/V feature resolution on LSTM* performance when using each annotation technique. Varying the input sampling step size (t_{st}) for the 3-steps ahead forecasting problem.

at the resolution of interest. The sampling technique is illustrated in the Fig. 4.9. Notably, we consider four different overlap margins of roughly $90\%, 80\%, 30\%$, and 0% with t_{st} values of $30, 60, 180$, and 300 respectively. We found that the step-sizes had a significant effect on MAE scores across all annotation methods considered, with a consistent trend of worsening MAE as a function of increase in aggregation. We found an average increase of $> 70\%$ in MAE scores when contrasting the most coarse ($t_{st} = 30s$) representation with that of the finest resolution ($t_{st} = 300$), where $\mathbf{X}_{HRV}$ had the following shapes $[\sim 560, 50]$, and $[\sim 70, 50]$ respectively.

4.6 Discussion

In this paper we explore the relevance of heart rate and its variability (HR/V) as reliable neurocognitive indices of fatigue. We study mental fatigue as experienced during an hour long working memory exercise, induced by the *time-on-task* effect. This problem is wrought with challenges not only related to the type of sensing used but also one that relates to the search for ground truth – how do we qualify these states of fatigue, and would those measures share any parallels with vagally-

Table 4.6: Effect of HR/V feature resolution on LSTM* performance when using each annotation technique. Varying the input sampling step size (t_{st}) for the 3-steps ahead forecasting problem. SE $\in \pm[0.006, 0.022]$

	Feature Vector Resolution			
Step size (t_{st}) (s)	30	60	180	300
t	0.053	0.092	0.218	0.244
$t + \Delta t$	0.056	0.101	0.237	0.244
$t + 2\Delta t$	0.064	0.139	0.245	0.267

mediated cardiac activity? A pathway that is known to govern the attention and response behaviors of individuals engaged in similar exercises [68]. In exploring these considerations we also address the question - "Whats in a label?", more specifically, how do different annotation strategies influence our ability to forecast fatigue states when relying on HR/V based features as predictors. We consider performance-based and perception-based fatigue indices, and introduce novel alternatives that derive from existing fatigue literature on clinical populations, and observed practices in affective computing. We formulate the underlying problem as one of *multivariate*, *multi-step-ahead* time-series forecasting and rely on input sequences of multi-dimensional HR/V features to forecast output sequences of a relevant fatigue score. To this end, we utilize data collected as a part of a larger investigation on fatigue (see Chapters 2 and 3), where we have unique subjective responses, performance measures, and ECG data that support these explorations. In building models for this problem we also assessed the comparative effectiveness of annotation methods, modeling architecture, lead and lag horizons for forecasting and training, the impact of label combinations, the choice of training paradigm and evaluation technique, and the effect of input feature resolution. Our goal is for these observations to advance more focused scrutiny on the transferability and relevance of HR/V indices in related problems of cognitive dysfunction in the non-clinical setting.

We found that forecast outcomes with perception-labels fared better than those with performance-labels regardless of the model choice, forecast horizon, approach or feature resolution. We drew these comparisons primarily due to our observation that performance and perception-scores exhibited some incongruity across experiment conditions in our previous work (see Chapters 2 and 3), where while performance tendencies improved, perceptions of effort or fatigue states did not, suggesting that these behaviors are driven by disparate fatigue-impacted underlying faculties [120]. Therefore, the finding that perception-based labels do not lend well to the concept of HR/V as predictors is not wholly surprising.

Another point to note here are the limitations of the experiment protocol that could bias performance behaviors to regimes outside of the typical fatigue experience. For example, we know that the task format used in this investigation elicited distinct learning, and anticipatory behaviors in a subset of our participant pool. This was largely due to the differences in training demands and the lack of appropriate control on experiment termination procedures. Therefore, these are factors that confound the relevance of performance states as they relate to the fatigue dynamics of HR/V. In addition, workload on the task was constant and not adapted to individual capacity, a practice that is common in related studies of working memory [205], which could result in a performance ceiling effect or under-load related task disengagement [20]. However, this does not take away from the robustness of the fatigue-induction process given the subjective self-reports, and PRE-POST experiment survey responses (see Fig. 2.5).

On the other hand, our observation that perception-scores fared better as fatigue indices speaks to their use in the real world, where performance sampling is often inhibited. Factors such as the nature of the task, dynamic roles and responsibilities hinder the development of objective definitions and assessment methods for performance [206] on the job. Therefore, subjective-responses provide a more ecologically permissible format for querying the human fatigue states. However, perception-labels are served by *interruptive* sampling methods, place emphasis on the need for recall, have relatively poor temporal resolution [112, 113, 114], and are therefore also limited in their potential for real-world use. This need for reliable and unobtrusive alternatives to signal fatigue states, those that move past the barriers of current techniques, motivated us to consider cardiac heuristics. These indices rely on existing literature on fatigue in clinical populations, such as those with Multiple Sclerosis (MS), a condition known to elicit chronic symptoms of fatigue.

The basic principle that guides the engineering of this fatigue index is that lower HR/V is known to be associated with diminished cognitive capacities especially in tasks of executive control or willed action [135, 207, 208], similarly elevated levels of LF-power [180], an index for parasympathetic tone [184]. Interestingly, we found that the fatigue-index derived based on this heuristic was well correlated with fatigue-perceptions derived from our self-reports (see Fig. 4.3). Furthermore, we observed the best model MAE scores when relying on the heuristic-derived fatigue index regardless of the technique, approach or other variables of interest in our experiments. These findings signal the possibility of fatigue representations that are fundamentally encoded within cardiac activity, despite the relatively short duration of exposure to the fatiguing exercise. However, it is still premature to place confidence in these observations especially because, while we exclude the features directly related to the fatigue index in the input feature vector i.e.$[\mathbf{X}_{RMSSD}, \mathbf{X}_{pNN50}, \mathbf{X}_{LF}]$ during training, there remain highly correlated variables and latent dimensions that could easily approximate the target vector for the forecast horizons of interest.

We also observed that the *Seq2Seq* LSTM (LSTM*) architecture outperformed every other technique considered in our bench-marking exercise, regardless of the annotation method. This is not surprising given the degree of parameterization enabled by the architecture, and also with the operating hypothesis that fatigue in this exercise is driven by *time-on-task*, i.e. it is the cumulative outcome of activity and not an instantaneous event. Therefore, the temporal component and it's persistence is a central aspect of the predictive potential for related biomarkers. In other words, the ability to preserve long-term relationships in the underling signals and finding correlations across feature vectors is essential to tracking fatigue perceptions or related heuristics in these experiments. A function that the LSTM* is particularly well-suited to perform. In addition, we observed that the model performed poorly when aggregated across particpants, i.e. generalization resulted in poor model error tracking. We hypothesize that this is likely due to the degree of individual factors that differentiate HR/V and the periodicity of the fatigue response. Although population level heuristics do serve us well in this context (e.g. the heuristic-derived fatigue label), clinical observations also point to a similar degree of individual variability [181, 209].

Another fundamental question of interest here relates to learning a ground truth given multiple annotation strategies. As a first step, we consider the naive approach to combine perception and performance-indices which yielded a minor improvement in outcomes, especially with a bias towards perception-indices (see Fig. 4.7). It's hard to pin-point where this benefit is derived from, however we do expect that the function shape is likely altered in a way that is advantageous or relevant to the underlying representations in the HR/V time series, similar observations have been found in aggregation techniques for facial affect (e.g. [203]). However, a perspective than can really push these observations in the right direction is one that looks to treat each annotation technique as a noisy distortion of the true ground truth, e.g. the work by Raykar et al. proposes an Expectation Maximization based model to learn across noisy annotators in their work aptly titled, "Learning from Crowds" [210]. This idea is currently investigated in other contexts of affective computing such as facial emotion recognition, where multiple annotators may disagree on the degree of "affect" seen in a visual [211]. In those contexts, algorithms have found success to learn an optimal annotation technique or ground truth from across annotators (individuals), a related perspective here could be to treat each annotation method as a distortion of the underlying ground truth, a strategy that could improve the outlook and discussion across the annotation methods of interest in this study. We also observed that the input vector horizon and forecast horizon both impacted model performance in different ways. As the input time window increased, model performance improved significantly, we hypothesize that this is likely due to **(1)** *overfitting* on the data, and **(2)** the effect this has on the number of training and testing samples. Given that the experiment was only 60 minutes long, changing the input sample window significantly limits the number of independent observations that remain available for testing. On varying the forecast/ prediction horizon, we found that model performance worsened as a function of Δt, i.e. the forecasts were unreliable further ahead in time, which is unsurprising considering how short-term dynamics and correlations across the underlying HR/V can change with each Δt. Finally, we found that the input feature resolution as governed by the step size (t_{st}) parameter negatively impacted MAE scores, especially at lower resolutions, which was expected given the impact it has on temporal resolution of the training data, and how sensitive machine learning models are to these changes.

While these observations signal some optimism towards the techniques we could operationalize to forecast fatigue states under different constraints, there remain some key limitations which demand our attention; they are discussed in the section to follow.

4.6.1 Limitations

First, as discussed in the previous section, our observations are limited by the experiment protocol. A subset of participants reported an anticipation bias towards the end of the experiment that resulted in an artificial state of alertness. This resulted in changes to the performance tendencies and therefore is likely a confound to the relevance of those indices as a measure of fatigue, or distortions of the underlying ground truth (i.e. fatigue). On a related note, we know that learning played a role early in the experiment and therefore perceived workload and associated fatigue are enhanced until stable performance capacity or a "plateau" is reached, although it could be argued that these are factors relevant to the task-induced experience of fatigue. Our current efforts look to work past these deficiencies with more control on participant training protocols and their task experience. Second, the reliability of the heuristic index needs more investigation, especially knowing that there could be latent variables or other higher-order correlations that enable the architecture to track these indices over time, however, it is worth noting that this method is consistent with techniques commonly operationalized in time-series forecasting problems today, where the target and predictor variables are typically identical or related measures e.g. stock prices [212], power consumption [213], weather indices [214], etc. Third, HR/V is a low dimensional and degenerate index that is susceptible to several perturbations such as due to breathing or motor activity, the specific influence of these variables on masking fatigue representations need focused investigations and protocols. Fourth, while we establish the relevance of HR/V as an index for tracking perceptions and related indices of fatigue, we do so with the convenience of offline processing tools. However, we need to investigate the overheads associated with incorporating the feature extraction workflow and prediction pipeline in "real time", especially to accommodate the need to retrain learned representations over time contingent on the experiments or activities the user is engaged in.

4.7 Conclusions

In this paper, we discussed methods to forecast fatigue states using a multi-step ahead multivariate formulation of the problem, and features derived from heart-rate and its variability as the predictor variables. We found distinctions in model outcomes that were influenced by the choice of annotation index, i.e. ground truth, for the time-series fatigue label, where perception-indices and cardiac heuristics fared better than performance as indexed by features of HR/V. We observed that the proposed *Seq2Seq* architecture is particularly efficient in minimizing forecast error, specifically, the *Mean Absolute Error* across the desired prediction horizon. These observations lend confidence to the prospect of using a low-dimensional, and unobtrusive wearable sensor such as ECG devices to reliably forecast perceptions, performance measures, and underling cardiac heuristic indicators of fatigue under controlled experiment conditions. Focused research on translating these findings to the real world can be transformative to those job domains where fatigue exposure and risk remain insidious. However, there are specific barriers that demand our immediate research attention such as, **(1)** understanding the role of perturbations such as breathing/ activity on the robustness of these indicators; **(2)** studying the effect of circadian or sleep-dependent variations on the dynamics of HR/V; and **(3)** exploring the task-invariance of these observations by extending to related cognitive fatigue contexts like driving simulations, vigilance experiments among others could help advance robust and fieldable fatigue indices.

5. CONCLUSIONS

This dissertation explored several foundational questions related to the subjective experience of fatigue, associated biomarkers, the ability to counter those states using non-invasive brain stimulation, and the need to consider multiple ground truths when attempting to forecast fatigue. In Chapter 2 we shared evidence that anodal tDCS of brain regions responsible for working memory faculties promoted response commission behaviors, while preserving response inhibition behaviors on a visuospatial working memory task. However, the benefits found in performance outcomes were not seen in perception indices, with participants reporting a condition-agnostic worsening of mood, sleepiness and fatigue perceptions over the time course of the experiment. This raises important questions, especially those related to the ethics around how such mechanisms can translate to real world use cases or the decision making that can operationalize them within job domains. It is easy to see how one may choose a performance- or perception-centric view contingent on the objective function that underlies their optimization goal, i.e. promote worker safety or the overall fiscal gains associated with their job responsibilities, adding to the ethical undertones that demand more scrutiny.

The neural and physiological drivers of fatigue are complex, coupled, and poorly understood. Investigations that combine the fidelity of neural indices and the field-readiness of physiological measures can drive translational research in recognizing and mitigating fatigue states in operational settings. In Chapter 3 we examined the spatio-temporal dynamics of neural activity and its physiological correlates in heart rate and its variability (HR/V) during this fatiguing visuospatial working memory task. We observed that task performance elicited increased neural activation in regions responsible for maintaining working memory capacity. With the onset of time-on-task effects, however, resource utilization was seen to increase beyond task-relevant networks. Over time, functional connections in the prefrontal cortex were seen to weaken, with changes in the causal relationships between key regions known to drive working memory. HR/V indices were seen to closely follow activity in the prefrontal cortex. This investigation provided a window into the neural and physiological underpinnings of working memory under the *time-on-task* effect, with changes in functional and causal brain networks that unpack its influence on executive function. HR/V was largely shown to mirror changes in cortical networks responsible for working memory, therefore supporting the possibility of unobtrusive state recognition under ecologically valid conditions. However, this prospect of forecasting fatigue states or behaviors should also consider how we label fatigue on the task.

Chapter 4 explores the specific issues that originate from the differences in labeling strategies on the ability to forecast those indices. We treat the problem as a multivariate multi-step ahead forecasting problem where target vectors can be one of performance, perception or cardiac indices. The operating hypothesis here is that, perceptions queried through subjective questionnaires, performance indicators derived from the task, and the fatigue-driven cardiac heuristics are likely distortions of an underlying ground truth that relates to the objective fatigue experience of the individual. Models were found to perform equitably when trying to forecast both perception-based indices and cardiac-heuristics. However, performance was worse when attempting to forecast performance indices regardless of the technique used. While the observations indicate positively for the prospect of forecasting fatigue-related indices using cardiac measures, the disparity between performance and perception indices on model tracking error raises more questions on the optimal strategy for use under more ecologically valid conditions. These observations together provide a template of resources that could evolve a closed-loop solution for NIBS in the real world, however, some immediate research directions demand our attention towards this goal.

5.1 Future Work

5.1.1 Ecological Validity in Task Formats

In Chapter 2, we reported the benefits of stimulation during a well-controlled experiment paradigm that queried user visuospatial working memory. However, our ultimate goal is to translate these observations toward relevant uses in domains such as emergency response. The domain adaptation of these findings is essential for their real world use and purported benefits. Examples of job domains that relate to the visuospatial working memory demands operationalized in these experiments is the role of an air traffic controller or individuals responsible for emergency operations centers (EOCs), where there is need for auditory, verbal and spatial working memory in encoding incoming streams of information and relaying it to team members in some actionable form. Therefore, a key effort for us is to architect an experiment paradigm that allows us to move from the convenience of task formats derived from experimental psychophysics towards task formats synonymous to real-world use cases. However, there is a need to ensure that this transition is incremental, as the workload and training demands associated with a fully organic test-case are likely separate research efforts of their own.

5.1.2 Stimulation Parameter Space

The evidence reported in this dissertation serves as the foundation for personalized, non-invasive, closed-loop neuromodulation. However, unknowns related to the timing, location, intensity, repeated-use, and task-agnosticity of stimulation effects remain. There is a need to systematically explore this parameter space towards developing operationally significant tools for both clinical and non-clinical applications. In doing so, there is a need to – (1) identify causal networks for cognitive vulnerabilities in the brain (where and why); (2) harness downstream physiological and behavioral markers for predicting future states (when); (3) modulate targeted brain functions (what) to support or augment performance using non-invasive, i.e., non-implantable or ingestible, approaches; and (4) bridge gaps in stakeholder (who) understanding and perceptions of these technologies, towards wider-adoption and use.

5.1.3 Finding the Ground Truth

While we consider the influence of annotation strategies on our ability to forecast fatigue, we are still unable to comment on whether any of these techniques reflect a task invariant formulation of the construct. Moreover, we do not apply any optimization to collate all proposed methods to find an optimal ground truth. This concept of annotator noise, is not novel, and is one that is widely discussed in the context of learning from crowds [210], where annotator noise in labeling levels of affect can be regarded as some distortion of the underlying ground truth. In these contexts, researchers have considered different algorithms to find an optimal ground truth, example, using the *Expectation Maximization* algorithm to find the *Maximum Likelihood Estimate* that best combines each annotation for an *optimal* ground truth. In the present problem, we could apply a similar technique where we consider the differences across annotation techniques instead of individual annotators to find an optimal labeling technique for the given data set. However, the ultimate problem is likely to invovle to stages, **(1)** to find an optimal ground truth given these noisy annotations, and **(2)** validating this technique on an unrelated but inspired data set where the fatigue induction paradigm is consistent but there are other protocol related changes that deviate from the original experiment. This effort remains our present focus, and will likely signal the relevance of cardiac activity as a neurocognitive index with more certainty than presently claimed.

5.1.4 Technology Acceptance in the Real World

Finally, an important consideration for the translation of these ideas into the real world relates to stakeholder understanding, perceptions and acceptance of the proposed technologies. Our target demographic for the incremental evolution of this closed-loop framework are emergency responders, who operate in a domain that is significantly impacted by fatigue-related challenges. While this is a well-suited application domain, the technology acceptance barriers are nuanced, and our knowledge gaps demand more research. Specifically, we need to consider the prevalence, social acceptance, potential for misuse, and long-term safety of the end-users in developing tools or frameworks to assist cognitive faculties regardless of the task domain [215]. This process should include customer discovery efforts that are centered on interviews to both detail the mechanisms of the proposed intervention techniques, and to understand the fatigue risks and definitions that matter the most to these groups. This process is all the more important considering the dichotomy between performance and perception for both intervention and forecasting, where, we were able to improve performance outcomes with intervention (see Chapter 2), while our sensing mechanisms were better primed to forecast perception states (see Chapter 4). Suggesting the need to carefully determine use-inspired templates to advance these mechanisms for the real-world.

REFERENCES

[1] F. J. Carod-Artal and C. Vázquez-Cabrera, "Burnout syndrome in an international setting," in *Burnout for experts*, pp. 15–35, Springer, 2013.

[2] O. A. Mullette-Gillman, R. L. Leong, and Y. A. Kurnianingsih, "Cognitive fatigue destabilizes economic decision making preferences and strategies," *PloS one*, vol. 10, no. 7, p. e0132022, 2015.

[3] J. Arendt, "Shift work: coping with the biological clock," *Occupational medicine*, vol. 60, no. 1, pp. 10–20, 2010.

[4] S. Hayes, J. L. Priestley, N. Ishmakhametov, and H. E. Ray, ""i'm not working from home, i'm living at work": Perceived stress and work-related burnout before and during covid-19," 2020.

[5] E. Demerouti, A. B. Bakker, F. Nachreiner, and W. B. Schaufeli, "The job demands-resources model of burnout.," *Journal of Applied psychology*, vol. 86, no. 3, p. 499, 2001.

[6] A. Grant, "There'sa name for the blah you're feeling: It's called languishing," *The New York Times*, 2021.

[7] M. B. Herlambang, N. A. Taatgen, and F. Cnossen, "The role of motivation as a factor in mental fatigue," *Human factors*, vol. 61, no. 7, pp. 1171–1185, 2019.

[8] M. A. Boksem, T. F. Meijman, and M. M. Lorist, "Effects of mental fatigue on attention: an erp study," *Cognitive brain research*, vol. 25, no. 1, pp. 107–116, 2005.

[9] T. F. Meijman, "Mental fatigue and the efficiency of information processing in relation to work times," *International Journal of Industrial Ergonomics*, vol. 20, no. 1, pp. 31–38, 1997.

[10] G. R. J. Hockey, A. John Maule, P. J. Clough, and L. Bdzola, "Effects of negative mood states on risk in everyday decision making," *Cognition & Emotion*, vol. 14, no. 6, pp. 823–855, 2000.

[11] C. P. West, A. D. Tan, T. M. Habermann, J. A. Sloan, and T. D. Shanafelt, "Association of resident fatigue and distress with perceived medical errors," *Jama*, vol. 302, no. 12, pp. 1294–1300, 2009.

[12] P. S. Della Rocco *et al.*, "The role of shift work and fatigue in air traffic control operational errors and incidents," tech. rep., United States. Department of Transportation. Federal Aviation Administration ..., 1999.

[13] K. Alkema, J. M. Linton, and R. Davies, "A study of the relationship between self-care, compassion satisfaction, compassion fatigue, and burnout among hospice professionals," *Journal of Social Work in End-of-Life & Palliative Care*, vol. 4, no. 2, pp. 101–119, 2008.

[14] G. E. Vincent, B. Aisbett, S. J. Hall, and S. A. Ferguson, "Fighting fire and fatigue: sleep quantity and quality during multi-day wildfire suppression," *Ergonomics*, vol. 59, no. 7, pp. 932–940, 2016.

[15] F. M. Donald, "The classification of vigilance tasks in the real world," *Ergonomics*, vol. 51, no. 11, pp. 1643–1655, 2008.

[16] R. A. Grier, J. S. Warm, W. N. Dember, G. Matthews, T. L. Galinsky, J. L. Szalma, and R. Parasuraman, "The vigilance decrement reflects limitations in effortful attention, not mindlessness," *Human factors*, vol. 45, no. 3, pp. 349–359, 2003.

[17] R. Parasuraman, J. S. Warm, and W. N. Dember, "Vigilance: Taxonomy and utility," in *Ergonomics and human factors*, pp. 11–32, Springer, 1987.

[18] A. R. Neigel, V. L. Claypoole, S. L. Smith, G. E. Waldfogle, N. W. Fraulini, G. M. Hancock, W. S. Helton, and J. L. Szalma, "Engaging the human operator: a review of the theoretical

support for the vigilance decrement and a discussion of practical applications," *Theoretical Issues in Ergonomics Science*, pp. 1–20, 2019.

[19] E. L. Wiener, "Application of vigilance research: Rare, medium, or well done?," *Human Factors*, vol. 29, no. 6, pp. 725–736, 1987.

[20] N. Pattyn, X. Neyt, D. Henderickx, and E. Soetens, "Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue?," *Physiology & behavior*, vol. 93, no. 1-2, pp. 369–378, 2008.

[21] P. Thiffault and J. Bergeron, "Fatigue and individual differences in monotonous simulated driving," *Personality and individual differences*, vol. 34, no. 1, pp. 159–176, 2003.

[22] G. Matthews, D. R. Davies, and P. J. Holley, "Cognitive predictors of vigilance," *Human factors*, vol. 35, no. 1, pp. 3–24, 1993.

[23] J. T. Nelson, R. A. McKinley, E. J. Golob, J. S. Warm, and R. Parasuraman, "Enhancing vigilance in operators with prefrontal cortex transcranial direct current stimulation (tdcs)," *Neuroimage*, vol. 85, pp. 909–917, 2014.

[24] F. Al-Shargie, U. Tariq, H. Mir, H. Alawar, F. Babiloni, and H. Al-Nashash, "Vigilance decrement and enhancement techniques: a review," *Brain sciences*, vol. 9, no. 8, p. 178, 2019.

[25] R. Parasuraman, J. S. Warm, and J. E. See, "Brain systems of vigilance.," 1998.

[26] M. Körber, A. Cingel, M. Zimmermann, and K. Bengler, "Vigilance decrement and passive fatigue caused by monotony in automated driving," *Procedia Manufacturing*, vol. 3, pp. 2403–2409, 2015.

[27] W. S. Helton and P. N. Russell, "Working memory load and the vigilance decrement," *Experimental brain research*, vol. 212, no. 3, pp. 429–437, 2011.

[28] B. S. Oken, M. C. Salinsky, and S. Elsas, "Vigilance, alertness, or sustained attention: physiological basis and measurement," *Clinical neurophysiology*, vol. 117, no. 9, pp. 1885–1901, 2006.

[29] A. T. Kamzanova, A. M. Kustubayeva, and G. Matthews, "Use of eeg workload indices for diagnostic monitoring of vigilance decrement," *Human factors*, vol. 56, no. 6, pp. 1136–1149, 2014.

[30] G. Mangun and S. Hillyard, "The spatial allocation of visual attention as indexed by event-related brain potentials," *Human factors*, vol. 29, no. 2, pp. 195–211, 1987.

[31] M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature reviews neuroscience*, vol. 3, no. 3, pp. 201–215, 2002.

[32] C. Bogler, J. Mehnert, J. Steinbrink, and J.-D. Haynes, "Decoding vigilance with nirs," *PloS one*, vol. 9, no. 7, 2014.

[33] Y. Zhu, C. Rodriguez-Paras, J. Rhee, and R. K. Mehta, "Methodological approaches and recommendations for functional near-infrared spectroscopy applications in hf/e research," *Human factors*, p. 0018720819845275, 2019.

[34] A. Luque-Casado, J. C. Perales, D. Cárdenas, and D. Sanabria, "Heart rate variability and cognitive processing: The autonomic response to task demands," *Biological psychology*, vol. 113, pp. 83–90, 2016.

[35] C. Pierard, P. Satabin, D. Lagarde, B. Barrere, C. Guezennec, J. Menu, and M. Pérès, "Effects of a vigilance-enhancing drug, modafinil, on rat brain metabolism: 2d cosy 1h-nmr study," *Brain research*, vol. 693, no. 1-2, pp. 251–256, 1995.

[36] W. Paulus, "Transcranial electrical stimulation (tes–tdcs; trns, tacs) methods," *Neuropsychological rehabilitation*, vol. 21, no. 5, pp. 602–617, 2011.

[37] M. Ironside, J. O'Shea, P. J. Cowen, and C. J. Harmer, "Frontal cortex stimulation reduces vigilance to threat: implications for the treatment of depression and anxiety," *Biological psychiatry*, vol. 79, no. 10, pp. 823–830, 2016.

[38] L. K. McIntire, R. A. McKinley, C. Goodyear, and J. Nelson, "A comparison of the effects of transcranial direct current stimulation and caffeine on vigilance and cognitive performance during extended wakefulness," *Brain stimulation*, vol. 7, no. 4, pp. 499–507, 2014.

[39] V. A. Naushad, J. J. Bierens, K. P. Nishan, C. P. Firjeeth, O. H. Mohammad, A. M. Maliyakkal, S. ChaliHadan, and M. D. Schreiber, "A systematic review of the impact of disaster on the mental health of medical responders," *Prehospital and disaster medicine*, vol. 34, no. 6, pp. 632–643, 2019.

[40] M. Yung, B. Du, J. Gruber, and A. Yazdani, "Developing a canadian fatigue risk management standard for first responders: Defining the scope," *Safety science*, vol. 134, p. 105044, 2021.

[41] J. I. Westbrook, M. Z. Raban, S. R. Walter, and H. Douglas, "Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study," *BMJ quality & safety*, vol. 27, no. 8, pp. 655–663, 2018.

[42] G. E. Vincent, B. Aisbett, A. Wolkow, S. M. Jay, N. D. Ridgers, and S. A. Ferguson, "Sleep in wildland firefighters: what do we know and why does it matter?," *International journal of wildland fire*, vol. 27, no. 2, pp. 73–84, 2018.

[43] B. J. Schmeichel, "Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control.," *Journal of Experimental Psychology: General*, vol. 136, no. 2, p. 241, 2007.

[44] T. D. Wager and E. E. Smith, "Neuroimaging studies of working memory," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 3, no. 4, pp. 255–274, 2003.

[45] K. T. Jones, D. J. Peterson, K. J. Blacker, and M. E. Berryhill, "Frontoparietal neurostimulation modulates working memory training benefits and oscillatory synchronization," *Brain research*, vol. 1667, pp. 28–40, 2017.

[46] F. Keshvari, H.-R. Pouretemad, and H. Ekhtiari, "The polarity-dependent effects of the bilateral brain stimulation on working memory," *Basic and clinical neuroscience*, vol. 4, no. 3, p. 224, 2013.

[47] K. E. Hoy, M. R. Emonson, S. L. Arnold, R. H. Thomson, Z. J. Daskalakis, and P. B. Fitzgerald, "Testing the limits: investigating the effect of tdcs dose on working memory enhancement in healthy controls," *Neuropsychologia*, vol. 51, no. 9, pp. 1777–1784, 2013.

[48] J. Gill, P. P. Shah-Basak, and R. Hamilton, "It's the thought that counts: examining the task-dependent effects of transcranial direct current stimulation on executive function," *Brain stimulation*, vol. 8, no. 2, pp. 253–259, 2015.

[49] G. Giglia, F. Brighina, S. Rizzo, A. Puma, S. Indovino, S. Maccora, R. Baschi, G. Cosentino, and B. Fierro, "Anodal transcranial direct current stimulation of the right dorsolateral prefrontal cortex enhances memory-guided responses in a visuospatial working memory task," *Functional neurology*, vol. 29, no. 3, p. 189, 2014.

[50] J. Dedoncker, A. R. Brunoni, C. Baeken, and M.-A. Vanderhasselt, "A systematic review and meta-analysis of the effects of transcranial direct current stimulation (tdcs) over the dorsolateral prefrontal cortex in healthy and neuropsychiatric samples: influence of stimulation parameters," *Brain stimulation*, vol. 9, no. 4, pp. 501–517, 2016.

[51] L. E. Mancuso, I. P. Ilieva, R. H. Hamilton, and M. J. Farah, "Does transcranial direct current stimulation improve healthy working memory?: a meta-analytic review," *Journal of Cognitive Neuroscience*, vol. 28, no. 8, pp. 1063–1089, 2016.

[52] A. T. Hill, P. B. Fitzgerald, and K. E. Hoy, "Effects of anodal transcranial direct current stimulation on working memory: a systematic review and meta-analysis of findings from

healthy and neuropsychiatric populations," *Brain stimulation*, vol. 9, no. 2, pp. 197–208, 2016.

[53] C. Fonteneau, M. Mondino, M. Arns, C. Baeken, M. Bikson, A. R. Brunoni, M. J. Burke, T. Neuvonen, F. Padberg, A. Pascual-Leone, *et al.*, "Sham tdcs: A hidden source of variability? reflections for further blinded, controlled trials," *Brain stimulation*, vol. 12, no. 3, pp. 668–673, 2019.

[54] L. K. McIntire, R. A. McKinley, J. M. Nelson, and C. Goodyear, "Transcranial direct current stimulation versus caffeine as a fatigue countermeasure," *Brain stimulation*, vol. 10, no. 6, pp. 1070–1078, 2017.

[55] G. Borragán, M. Gilson, C. Guerrero-Mosquera, E. Di Ricci, H. Slama, and P. Peigneux, "Transcranial direct current stimulation does not counteract cognitive fatigue, but induces sleepiness and an inter-hemispheric shift in brain oxygenation," *Frontiers in psychology*, vol. 9, p. 2351, 2018.

[56] R. Karthikeyan and R. K. Mehta, "Towards a closed-loop neurostimulation platform for augmenting operator vigilance," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3976–3983, IEEE, 2020.

[57] J. Lim, W.-c. Wu, J. Wang, J. A. Detre, D. F. Dinges, and H. Rao, "Imaging brain fatigue from sustained mental workload: an asl perfusion study of the time-on-task effect," *Neuroimage*, vol. 49, no. 4, pp. 3426–3435, 2010.

[58] T. Möckel, C. Beste, and E. Wascher, "The effects of time on task in response selection-an erp study of mental fatigue," *Scientific reports*, vol. 5, no. 1, pp. 1–9, 2015.

[59] P. L. Ackerman, R. Kanfer, S. W. Shapiro, S. Newton, and M. E. Beier, "Cognitive fatigue during testing: An examination of trait, time-on-task, and strategy influences," *Human performance*, vol. 23, no. 5, pp. 381–402, 2010.

[60] M. Krimsky, D. E. Forster, M. M. Llabre, and A. P. Jha, “The influence of time on task on mind wandering and visual working memory,” *Cognition*, vol. 169, pp. 84–90, 2017.

[61] J. F. Hopstaken, D. Van Der Linden, A. B. Bakker, and M. A. Kompier, “A multifaceted investigation of the link between mental fatigue and task disengagement,” *Psychophysiology*, vol. 52, no. 3, pp. 305–315, 2015.

[62] Y. Shigihara, M. Tanaka, A. Ishii, S. Tajima, E. Kanai, M. Funakura, and Y. Watanabe, “Two different types of mental fatigue produce different styles of task performance,” *Neurology, Psychiatry and Brain Research*, vol. 19, no. 1, pp. 5–11, 2013.

[63] M. Tanaka, K. Mizuno, S. Tajima, T. Sasabe, and Y. Watanabe, “Central nervous system fatigue alters autonomic nerve activity,” *Life sciences*, vol. 84, no. 7-8, pp. 235–239, 2009.

[64] J. K. Johannesen, J. Bi, R. Jiang, J. G. Kenney, and C.-M. A. Chen, “Machine learning identification of eeg features predicting working memory performance in schizophrenia and healthy adults,” *Neuropsychiatric electrophysiology*, vol. 2, no. 1, pp. 1–21, 2016.

[65] J. E. Reiser, E. Wascher, G. Rinkenauer, and S. Arnau, “Cognitive-motor interference in the wild: Assessing the effects of movement complexity on task switching using mobile eeg,” *European Journal of Neuroscience*, 2020.

[66] K. Tsunoda, A. Chiba, K. Yoshida, T. Watanabe, and O. Mizuno, “Predicting changes in cognitive performance using heart rate variability,” *IEICE TRANSACTIONS on Information and Systems*, vol. 100, no. 10, pp. 2411–2419, 2017.

[67] S. Nikolin, T. W. Boonstra, C. K. Loo, and D. Martin, “Combined effect of prefrontal transcranial direct current stimulation and a working memory task on heart rate variability,” *PloS one*, vol. 12, no. 8, p. e0181833, 2017.

[68] J. F. Thayer, A. L. Hansen, E. Saus-Rose, and B. H. Johnsen, “Heart rate variability, prefrontal neural function, and cognitive performance: the neurovisceral integration perspective

on self-regulation, adaptation, and health," *Annals of Behavioral Medicine*, vol. 37, no. 2, pp. 141–153, 2009.

[69] J. V. Bradley, "Complete counterbalancing of immediate sequential effects in a latin square design," *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 525–528, 1958.

[70] S. Shacham, "A shortened version of the profile of mood states.," *Journal of personality assessment*, 1983.

[71] K. Kaida, M. Takahashi, T. Åkerstedt, A. Nakata, Y. Otsuka, T. Haratani, and K. Fukasawa, "Validation of the karolinska sleepiness scale against performance and eeg variables," *Clinical neurophysiology*, vol. 117, no. 7, pp. 1574–1581, 2006.

[72] M. A. Nitsche, L. G. Cohen, E. M. Wassermann, A. Priori, N. Lang, A. Antal, W. Paulus, F. Hummel, P. S. Boggio, F. Fregni, *et al.*, "Transcranial direct current stimulation: state of the art 2008," *Brain stimulation*, vol. 1, no. 3, pp. 206–223, 2008.

[73] F. Strasser, M. Muma, and A. M. Zoubir, "Motion artifact removal in ecg signals using multi-resolution thresholding," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pp. 899–903, IEEE, 2012.

[74] V. Marked, "Correction of the heart rate variability signal for ectopics and missing beats," *Heart rate variability*, 1995.

[75] C. Li, C. Zheng, and C. Tai, "Detection of ecg characteristic points using wavelet transforms," *IEEE Transactions on biomedical Engineering*, vol. 42, no. 1, pp. 21–28, 1995.

[76] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[77] F. Teo, K. E. Hoy, Z. J. Daskalakis, and P. B. Fitzgerald, "Investigating the role of current strength in tdcs modulation of working memory performance in healthy controls," *Frontiers in psychiatry*, vol. 2, p. 45, 2011.

[78] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior research methods*, vol. 39, no. 2, pp. 175–191, 2007.

[79] E. L. Lehmann and H. J. D'Abrera, *Nonparametrics: statistical methods based on ranks.* Holden-day, 1975.

[80] J. R. Simon, "The effects of an irrelevant directional cue on human information processing," in *Advances in psychology*, vol. 65, pp. 31–86, Elsevier, 1990.

[81] F. Fregni, P. S. Boggio, M. Nitsche, F. Bermpohl, A. Antal, E. Feredoes, M. A. Marcolin, S. P. Rigonatti, M. T. Silva, W. Paulus, *et al.*, "Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory," *Experimental brain research*, vol. 166, no. 1, pp. 23–30, 2005.

[82] H. Arciniega, F. Gözenman, K. T. Jones, J. A. Stephens, and M. E. Berryhill, "Frontoparietal tdcs benefits visual working memory in older adults with low working memory capacity," *Frontiers in aging neuroscience*, vol. 10, p. 57, 2018.

[83] W. S. Helton and P. N. Russell, "Visuospatial and verbal working memory load: effects on visuospatial vigilance," *Experimental brain research*, vol. 224, no. 3, pp. 429–436, 2013.

[84] M. D'esposito, J. A. Detre, D. C. Alsop, R. K. Shin, S. Atlas, and M. Grossman, "The neural basis of the central executive system of working memory," *Nature*, vol. 378, no. 6554, pp. 279–281, 1995.

[85] M. d'Esposito, G. Aguirre, E. Zarahn, D. Ballard, R. Shin, and J. Lease, "Functional mri studies of spatial and nonspatial working memory," *Cognitive Brain Research*, vol. 7, no. 1, pp. 1–13, 1998.

[86] C. Kim, N. F. Johnson, and B. T. Gold, "Conflict adaptation in prefrontal cortex: now you see it, now you don't," *Cortex*, vol. 50, pp. 76–85, 2014.

[87] J. Tiego, R. Testa, M. A. Bellgrove, C. Pantelis, and S. Whittle, "A hierarchical model of inhibitory control," *Frontiers in psychology*, vol. 9, p. 1339, 2018.

[88] M. Causse, F. Dehais, and J. Pastor, "Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots," *The International Journal of Aviation Psychology*, vol. 21, no. 3, pp. 217–234, 2011.

[89] R. H. Paul, W. W. Beatty, R. Schneider, C. R. Blanco, and K. A. Hames, "Cognitive and physical fatigue in multiple sclerosis: relations between self-report and objective performance," *Applied Neuropsychology*, vol. 5, no. 3, pp. 143–148, 1998.

[90] A. L. Schwartz, P. M. Meek, L. M. Nail, J. Fargo, M. Lundquist, M. Donofrio, M. Grainger, T. Throckmorton, and M. Mateo, "Measurement of fatigue: determining minimally important clinical differences," *Journal of clinical epidemiology*, vol. 55, no. 3, pp. 239–244, 2002.

[91] R. McKendrick, B. Falcone, M. Scheldrup, and H. Ayaz, "Effects of transcranial direct current stimulation on baseline and slope of prefrontal cortex hemodynamics during a spatial working memory task," *Frontiers in human neuroscience*, vol. 14, p. 64, 2020.

[92] P. S. Williams, R. L. Hoffman, and B. C. Clark, "Preliminary evidence that anodal transcranial direct current stimulation enhances time to task failure of a sustained submaximal contraction," *PloS one*, vol. 8, no. 12, p. e81418, 2013.

[93] L. Angius, E. Santarnecchi, A. Pascual-Leone, and S. M. Marcora, "Transcranial direct current stimulation over the left dorsolateral prefrontal cortex improves inhibitory control and endurance performance in healthy individuals," *Neuroscience*, vol. 419, pp. 34–45, 2019.

[94] C. Christodoulou, "The assessment and measurement of fatigue," *Fatigue as a window to the brain*, pp. 19–35, 2005.

[95] V. M. Goghari and A. W. MacDonald III, "The neural basis of cognitive control: Response selection and inhibition," *Brain and cognition*, vol. 71, no. 2, pp. 72–83, 2009.

[96] A. D. Bender, H. L. Filmer, K. Garner, C. K. Naughtin, and P. E. Dux, "On the relationship between response selection and response inhibition: An individual differences approach," *Attention, Perception, & Psychophysics*, vol. 78, no. 8, pp. 2420–2432, 2016.

[97] J. B. Rowe, I. Toni, O. Josephs, R. S. Frackowiak, and R. E. Passingham, "The prefrontal cortex: response selection or maintenance within working memory?," *Science*, vol. 288, no. 5471, pp. 1656–1660, 2000.

[98] L.-W. Ko, Y.-C. Shih, R. K. Chikara, Y.-T. Chuang, and E. C. Chang, "Neural mechanisms of inhibitory response in a battlefield scenario: A simultaneous fmri-eeg study," *Frontiers in human neuroscience*, vol. 10, p. 185, 2016.

[99] H. L. Filmer, J. B. Mattingley, and P. E. Dux, "Improved multitasking following prefrontal tdcs," *Cortex*, vol. 49, no. 10, pp. 2845–2852, 2013.

[100] J. Leite, Ó. F. Gonçalves, P. Pereira, N. Khadka, M. Bikson, F. Fregni, and S. Carvalho, "The differential effects of unihemispheric and bihemispheric tdcs over the inferior frontal gyrus on proactive control," *Neuroscience research*, vol. 130, pp. 39–46, 2018.

[101] G. Park and J. F. Thayer, "From the heart to the mind: cardiac vagal tone modulates top-down and bottom-up visual perception and attention to emotional stimuli," *Frontiers in psychology*, vol. 5, p. 278, 2014.

[102] C. Zrenner, P. Belardinelli, F. Müller-Dahlhaus, and U. Ziemann, "Closed-loop neuroscience and non-invasive brain stimulation: a tale of two loops," *Frontiers in cellular neuroscience*, vol. 10, p. 92, 2016.

[103] S. P. Ruf, A. J. Fallgatter, and C. Plewnia, "Augmentation of working memory training by transcranial direct current stimulation (tdcs)," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.

[104] W. Dinn, F. Göral, S. Adigüzel, S. Karamürsel, F. Fregni, and A. Aycicegi-Dinn, "Effectiveness of tdcs blinding protocol in a sham-controlled study," *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, vol. 10, no. 2, p. 401, 2017.

[105] K. T. Jones, F. Gözenman, and M. E. Berryhill, "The strategy and motivational influences on the beneficial effect of neurostimulation: a tdcs and fnirs study," *Neuroimage*, vol. 105, pp. 238–247, 2015.

[106] A. M. Owen, K. M. McMillan, A. R. Laird, and E. Bullmore, "N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies," *Human brain mapping*, vol. 25, no. 1, pp. 46–59, 2005. Publisher: Wiley Online Library.

[107] G. Matthews and P. A. Hancock, *The handbook of operator fatigue*. CRC Press, 2017.

[108] J. Persson, K. M. Welsh, J. Jonides, and P. A. Reuter-Lorenz, "Cognitive fatigue of executive processes: Interaction between interference resolution tasks," *Neuropsychologia*, vol. 45, no. 7, pp. 1571–1579, 2007.

[109] J. G. Temple, J. S. Warm, W. N. Dember, K. S. Jones, C. M. LaGrange, and G. Matthews, "The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task," *Human factors*, vol. 42, no. 2, pp. 183–194, 2000. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[110] R. K. Mehta, J. Nuamah, S. C. Peres, and R. R. Murphy, “Field methods to quantify emergency responder fatigue: lessons learned from sUAS deployment at the 2018 Kilauea Volcano eruption,” *IISE transactions on occupational ergonomics and human factors*, pp. 1–9, 2020. Publisher: Taylor & Francis.

[111] R. K. Mehta and M. J. Agnew, “Subjective evaluation of physical and mental workload interactions across different muscle groups,” *Journal of occupational and environmental hygiene*, vol. 12, no. 1, pp. 62–68, 2015. Publisher: Taylor & Francis.

[112] R. D. O’Donnell, “Workload assessment methodology,” *Cognitive processes and performance*, vol. 2, pp. 1–49, 1986. Publisher: John Wiley & Sons.

[113] K. E. Garrison, *The Influence of Performance Incentives on Subjective Experiences of Mental Effort*. PhD Thesis, 2020.

[114] M. Lohani, B. R. Payne, and D. L. Strayer, “A review of psychophysiological measures to assess cognitive states in real-world driving,” *Frontiers in human neuroscience*, vol. 13, p. 57, 2019. Publisher: Frontiers.

[115] R. K. Mehta, S. C. Peres, P. Kannan, J. Rhee, A. E. Shortz, and M. S. Mannan, “Comparison of objective and subjective operator fatigue assessment methods in offshore shiftwork,” *Journal of Loss Prevention in the Process Industries*, vol. 48, pp. 376–381, 2017. Publisher: Elsevier.

[116] Y. Zhu, R. R. Jankay, L. C. Pieratt, and R. K. Mehta, “Wearable sensors and their metrics for measuring comprehensive occupational fatigue: a scoping review,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, pp. 1041–1045, SAGE Publications Sage CA: Los Angeles, CA, 2017. Issue: 1.

[117] D. E. Nee and M. D’Esposito, “The representational basis of working memory,” *Behavioral neuroscience of learning and memory*, pp. 213–230, 2016. Publisher: Springer.

[118] M. M. Owens, B. Duda, L. H. Sweet, and J. MacKillop, "Distinct functional and structural neural underpinnings of working memory," *NeuroImage*, vol. 174, pp. 463–471, 2018. Publisher: Elsevier.

[119] E. B. Klaassen, R. H. de Groot, E. A. Evers, J. Snel, E. C. Veerman, A. J. Ligtenberg, J. Jolles, and D. J. Veltman, "The effect of caffeine on working memory load-related brain activation in middle-aged males," *Neuropharmacology*, vol. 64, pp. 160–167, 2013. Publisher: Elsevier.

[120] P. Qi, H. Ru, L. Gao, X. Zhang, T. Zhou, Y. Tian, N. Thakor, A. Bezerianos, J. Li, and Y. Sun, "Neural mechanisms of mental fatigue revisited: New insights from the brain connectome," *Engineering*, vol. 5, no. 2, pp. 276–286, 2019. Publisher: Elsevier.

[121] R. Sala-Llonch, C. Pena-Gomez, E. M. Arenaza-Urquijo, D. Vidal-Piñeiro, N. Bargallo, C. Junque, and D. Bartres-Faz, "Brain connectivity during resting state and subsequent working memory task predicts behavioural performance," *Cortex*, vol. 48, no. 9, pp. 1187–1196, 2012. Publisher: Elsevier.

[122] K. J. Friston, "Functional and effective connectivity: a review," *Brain connectivity*, vol. 1, no. 1, pp. 13–36, 2011. Publisher: Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA.

[123] Y. Sun, J. Lim, K. Kwok, and A. Bezerianos, "Functional cortical connectivity analysis of mental fatigue unmasks hemispheric asymmetry and changes in small-world networks," *Brain and cognition*, vol. 85, pp. 220–230, 2014. Publisher: Elsevier.

[124] F. Dehais, A. Dupres, G. Di Flumeri, K. Verdiere, G. Borghini, F. Babiloni, and R. Roy, "Monitoring pilot's cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid fNIRS-EEG passive BCI," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 544–549, IEEE, 2018.

[125] M. Gergelyfi, B. Jacob, E. Olivier, and A. Zénon, “Dissociation between mental fatigue and motivational state during prolonged mental activity,” *Frontiers in behavioral neuroscience*, vol. 9, p. 176, 2015. Publisher: Frontiers.

[126] J. F. Thayer and R. D. Lane, “A model of neurovisceral integration in emotion regulation and dysregulation,” *Journal of affective disorders*, vol. 61, no. 3, pp. 201–216, 2000.

[127] J. F. Thayer and R. D. Lane, “Claude bernard and the heart–brain connection: Further elaboration of a model of neurovisceral integration,” *Neuroscience & Biobehavioral Reviews*, vol. 33, no. 2, pp. 81–88, 2009.

[128] E. E. Condy, B. H. Friedman, and A. Gandjbakhche, “Probing Neurovisceral Integration via Functional Near-Infrared Spectroscopy and Heart Rate Variability,” *Frontiers in Neuroscience*, vol. 14, 2020. Publisher: Frontiers Media SA.

[129] C. Chang, C. D. Metzger, G. H. Glover, J. H. Duyn, H.-J. Heinze, and M. Walter, “Association between heart rate variability and fluctuations in resting-state functional connectivity,” *Neuroimage*, vol. 68, pp. 93–104, 2013. Publisher: Elsevier.

[130] J. K. Nuamah, W. Mantooth, R. Karthikeyan, R. K. Mehta, and S. C. Ryu, “Neural efficiency of human-robotic feedback modalities under stress differs with gender,” *Frontiers in Human Neuroscience*, vol. 13, p. 287, 2019.

[131] T. J. Huppert, S. G. Diamond, M. A. Franceschini, and D. A. Boas, “Homer: a review of time-series analysis methods for near-infrared spectroscopy of the brain,” *Applied optics*, vol. 48, no. 10, pp. D280–D298, 2009.

[132] J. Rhee and R. K. Mehta, “Functional connectivity during handgrip motor fatigue in older adults is obesity and sex-specific,” *Frontiers in human neuroscience*, vol. 12, p. 455, 2018.

[133] M. Rubinov and O. Sporns, “Complex network measures of brain connectivity: uses and interpretations,” *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010. Publisher: Elsevier.

[134] L. Barnett and A. K. Seth, “The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference,” *Journal of neuroscience methods*, vol. 223, pp. 50–68, 2014. Publisher: Elsevier.

[135] G. Forte, F. Favieri, and M. Casagrande, “Heart rate variability and cognitive function: a systematic review,” *Frontiers in neuroscience*, vol. 13, p. 710, 2019.

[136] R. L. Burr, “Interpretation of normalized spectral heart rate variability indices in sleep research: a critical review,” *Sleep*, vol. 30, no. 7, pp. 913–919, 2007. Publisher: Oxford University Press.

[137] R. Bakeman, “Recommended effect size statistics for repeated measures designs,” *Behavior research methods*, vol. 37, no. 3, pp. 379–384, 2005. Publisher: Springer.

[138] M. R. Grech, A. Neal, G. Yeo, M. Humphreys, and S. Smith, “An examination of the relationship between workload and fatigue within and across consecutive days of work: Is the relationship static or dynamic?,” *Journal of occupational health psychology*, vol. 14, no. 3, p. 231, 2009.

[139] G. N. Dimitrakopoulos, I. Kakkos, Z. Dai, H. Wang, K. Sgarbas, N. Thakor, A. Bezerianos, and Y. Sun, “Functional connectivity analysis of mental fatigue reveals different network topological alterations between driving and vigilance tasks,” *IEEE transactions on neural systems and rehabilitation engineering*, vol. 26, no. 4, pp. 740–749, 2018. Publisher: IEEE.

[140] S. Ahn, T. Nguyen, H. Jang, J. G. Kim, and S. C. Jun, “Exploring neuro-physiological correlates of drivers’ mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data,” *Frontiers in human neuroscience*, vol. 10, p. 219, 2016. Publisher: Frontiers.

[141] A. K. Barbey, M. Koenigs, and J. Grafman, “Dorsolateral prefrontal contributions to human working memory,” *cortex*, vol. 49, no. 5, pp. 1195–1205, 2013.

[142] R. W. Engle, A. R. Conway, S. W. Tuholski, and R. J. Shisler, "A resource account of inhibition," *Psychological Science*, vol. 6, no. 2, pp. 122–125, 1995. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[143] M. J. Kane and R. W. Engle, "The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective," *Psychonomic bulletin & review*, vol. 9, no. 4, pp. 637–671, 2002. Publisher: Springer.

[144] V. A. Petruo, M. Mückschel, and C. Beste, "On the role of the prefrontal cortex in fatigue effects on cognitive flexibility-a system neurophysiological approach," *Scientific reports*, vol. 8, no. 1, pp. 1–13, 2018. Publisher: Nature Publishing Group.

[145] J. Fan and A. P. Smith, "The impact of workload and fatigue on performance," in *International symposium on human mental workload: Models and applications*, pp. 90–105, Springer, 2017.

[146] M. Causse, Z. Chua, V. Peysakhovich, N. Del Campo, and N. Matton, "Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS," *Scientific reports*, vol. 7, no. 1, pp. 1–15, 2017. Publisher: Nature Publishing Group.

[147] H. Wang, X. Liu, H. Hu, F. Wan, T. Li, L. Gao, A. Bezerianos, Y. Sun, and T.-P. Jung, "Dynamic reorganization of functional connectivity unmasks fatigue related performance declines in simulated driving," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 28, no. 8, pp. 1790–1799, 2020. Publisher: IEEE.

[148] S. Yu, A. Bezerianos, N. Thakor, and J. Li, "Functional brain network analysis reveals time-on-task related performance decline," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 271–274, IEEE, 2018.

[149] C. Zhao, M. Zhao, Y. Yang, J. Gao, N. Rao, and P. Lin, "The reorganization of human brain networks modulated by driving mental fatigue," *IEEE journal of biomedical and health informatics*, vol. 21, no. 3, pp. 743–755, 2016. Publisher: IEEE.

[150] J. F. Thayer, A. L. Hansen, J. J. Sollers III, and B. H. Johnsen, “Heart rate variability as an index of prefrontal neural function in military settings,” in *Biomonitoring for physiological and cognitive performance during military operations*, vol. 5797, pp. 71–77, International Society for Optics and Photonics, 2005.

[151] M. A. Boksem, T. F. Meijman, and M. M. Lorist, “Mental fatigue, motivation and action monitoring,” *Biological psychology*, vol. 72, no. 2, pp. 123–132, 2006. Publisher: Elsevier.

[152] T. J. Huppert, “Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy,” *Neurophotonics*, vol. 3, no. 1, p. 010401, 2016. Publisher: International Society for Optics and Photonics.

[153] M. Engle-Friedman, G. M. Mathew, A. Martinova, F. Armstrong, and V. Konstantinov, “The role of sleep deprivation and fatigue in the perception of task difficulty and use of heuristics,” *Sleep Science*, vol. 11, no. 2, p. 74, 2018.

[154] A. Sneddon, K. Mearns, and R. Flin, “Stress, fatigue, situation awareness and safety in offshore drilling crews,” *Safety Science*, vol. 56, pp. 80–88, 2013.

[155] P. D. Patterson, D. J. Buysse, M. D. Weaver, C. W. Callaway, and D. M. Yealy, “Recovery between work shifts among emergency medical services clinicians,” *Prehospital Emergency Care*, vol. 19, no. 3, pp. 365–375, 2015.

[156] M. B. Herlambang, F. Cnossen, and N. A. Taatgen, “The effects of intrinsic motivation on mental fatigue,” *PloS one*, vol. 16, no. 1, p. e0243754, 2021.

[157] L. Barrios, P. Oldrati, S. Santini, and A. Lutterotti, “Recognizing digital biomarkers for fatigue assessment in patients with multiple sclerosis,” in *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; PervasiveHealth*, EAI New York, NY, 2018.

[158] L. A. Walker, J. Berard, L. Berrigan, L. Rees, and M. Freedman, "Detecting cognitive fatigue in multiple sclerosis: method matters," *Journal of the neurological sciences*, vol. 316, no. 1-2, pp. 86–92, 2012.

[159] M. J. Bradshaw, S. Farrow, R. W. Motl, and T. Chitnis, "Wearable biosensors to monitor disability in multiple sclerosis," *Neurology: Clinical Practice*, vol. 7, no. 4, pp. 354–362, 2017.

[160] T. J. Balkin and N. J. Wesensten, "Differentiation of sleepiness and mental fatigue effects.," 2011.

[161] M. van der Hulst, T. Meijman, and T. Rothengatter, "Maintaining task set under fatigue: a study of time-on-task effects in simulated driving," *Transportation research part F: traffic psychology and behaviour*, vol. 4, no. 2, pp. 103–118, 2001.

[162] P. Tejero Gimeno, G. Pastor Cerezuela, and M. Choliz Montanes, "On the concept and measurement of driver drowsiness, fatigue and inattention: implications for countermeasures," *International journal of vehicle design*, vol. 42, no. 1-2, pp. 67–86, 2006.

[163] G. Gunzelmann, L. R. Moore, K. A. Gluck, H. Van Dongen, and D. F. Dinges, "Fatigue in sustained attention: Generalizing mechanisms for time awake to time on task.," 2011.

[164] A. D. Kohl, G. R. Wylie, H. Genova, F. G. Hillary, and J. Deluca, "The neural correlates of cognitive fatigue in traumatic brain injury using functional mri," *Brain injury*, vol. 23, no. 5, pp. 420–432, 2009.

[165] M. W. Chee, L. Y. Chuah, V. Venkatraman, W. Y. Chan, P. Philip, and D. F. Dinges, "Functional imaging of working memory following normal sleep and after 24 and 35 h of sleep deprivation: Correlations of fronto-parietal activation with performance," *Neuroimage*, vol. 31, no. 1, pp. 419–428, 2006.

[166] S. M. Doran, H. P. Van Dongen, and D. F. Dinges, "Sustained attention performance during sleep deprivation: evidence of state instability.," *Archives italiennes de biologie*, vol. 139, no. 3, pp. 253–267, 2001.

[167] M. P. Walker and R. Stickgold, "Sleep, memory, and plasticity," *Annu. Rev. Psychol.*, vol. 57, pp. 139–166, 2006.

[168] H. P. Van Dongen and D. F. Dinges, "Circadian rhythms in fatigue, alertness, and performance," *Principles and practice of sleep medicine*, vol. 20, pp. 391–399, 2000.

[169] S. Wessely, M. Sharpe, and M. Hotopf, *Chronic fatigue and its syndromes*. Oxford University Press, 1998.

[170] J. Haselkorn, C. Balsdon Richer, and D. Fry Welch, "Multiple sclerosis council for clinical practice guidelines. overview of spasticity management in multiple sclerosis. evidence-based management strategies for spasticity treatment in multiple sclerosis," *J Spinal Cord Med*, vol. 28, no. 2, pp. 167–199, 2005.

[171] R. Karthikeyan, M. R. Smoot, and R. K. Mehta, "Anodal tdcs augments and preserves working memory beyond time-on-task deficits," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

[172] S. B. Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa, "A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition," *Expert systems with applications*, vol. 39, no. 8, pp. 7067–7083, 2012.

[173] A.-M. Krypotos, S. Jahfari, V. A. van Ast, M. Kindt, and B. U. Forstmann, "Individual differences in heart rate variability predict the degree of slowing during response inhibition and initiation in the presence of emotional stimuli," *Frontiers in psychology*, vol. 2, p. 278, 2011.

[174] T. Force, "Standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[175] J. F. Thayer, F. Åhs, M. Fredrikson, J. J. Sollers III, and T. D. Wager, "A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 2, pp. 747–756, 2012.

[176] L. S. Colzato, B. J. Jongkees, M. de Wit, M. J. van der Molen, and L. Steenbergen, "Variable heart rate and a flexible mind: Higher resting-state heart rate variability predicts better task-switching," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 18, no. 4, pp. 730–738, 2018.

[177] C. Ottaviani, P. Zingaretti, A. M. Petta, G. Antonucci, J. F. Thayer, and G. F. Spitoni, "Resting heart rate variability predicts inhibitory control above and beyond impulsivity," *Journal of Psychophysiology*, 2018.

[178] B. H. Friedman, "An autonomic flexibility–neurovisceral integration model of anxiety and cardiac vagal tone," *Biological psychology*, vol. 74, no. 2, pp. 185–199, 2007.

[179] K. R. Ridderinkhof, W. P. Van Den Wildenberg, S. J. Segalowitz, and C. S. Carter, "Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning," *Brain and cognition*, vol. 56, no. 2, pp. 129–140, 2004.

[180] A. L. Hansen, B. H. Johnsen, and J. F. Thayer, "Vagal influence on working memory and attention," *International journal of psychophysiology*, vol. 48, no. 3, pp. 263–274, 2003.

[181] R. M. Escorihuela, L. Capdevila, J. R. Castro, M. C. Zaragozà, S. Maurel, J. Alegre, and J. Castro-Marrero, "Reduced heart rate variability predicts fatigue severity in individ-

uals with chronic fatigue syndrome/myalgic encephalomyelitis," *Journal of translational medicine*, vol. 18, no. 1, pp. 1–12, 2020.

[182] G. Videira, P. Castro, B. Vieira, J. P. Filipe, R. Santos, E. Azevedo, M. J. Sa, and P. Abreu, "Autonomic dysfunction in multiple sclerosis is better detected by heart rate variability and is not correlated with central autonomic network damage," *Journal of the neurological sciences*, vol. 367, pp. 133–137, 2016.

[183] F. Shirbani, E. Barin, Y.-C. Lee, K. Ng, J. D. E. Parratt, M. Butlin, and A. P. Avolio, "Characterisation of cardiac autonomic function in multiple sclerosis based on spontaneous changes of heart rate and blood pressure," *Multiple Sclerosis and Related Disorders*, vol. 22, pp. 120–127, 2018.

[184] F. Shaffer and J. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, p. 258, 2017.

[185] A. R. Brunoni, M. A. Nitsche, N. Bolognini, M. Bikson, T. Wagner, L. Merabet, D. J. Edwards, A. Valero-Cabre, A. Rotenberg, A. Pascual-Leone, *et al.*, "Clinical research with transcranial direct current stimulation (tdcs): challenges and future directions," *Brain stimulation*, vol. 5, no. 3, pp. 175–195, 2012.

[186] R. Karthikeyan, A. D. Mcdonald, and R. Mehta, "Stress detection during motor activity: Comparing neurophysiological indices in older adults," *IEEE Transactions on Affective Computing*, no. 01, pp. 1–1, 2022.

[187] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and A. S H Chen, "Neurokit2: A python toolbox for neurophysiological signal processing," 2020.

[188] J. Piskorski and P. Guzik, "Asymmetric properties of long-term and total heart rate variability," *Medical & biological engineering & computing*, vol. 49, no. 11, pp. 1289–1297, 2011.

[189] M. D. Costa, R. B. Davis, and A. L. Goldberger, “Heart rate fragmentation: a new approach to the analysis of cardiac interbeat interval dynamics,” *Frontiers in physiology*, vol. 8, p. 255, 2017.

[190] L. Zhang, W.-D. Zhou, P.-C. Chang, J.-W. Yang, and F.-Z. Li, “Iterated time series prediction with multiple support vector regression models,” *Neurocomputing*, vol. 99, pp. 411–422, 2013.

[191] X. Zhan, S. Zhang, W. Y. Szeto, and X. Chen, “Multi-step-ahead traffic speed forecasting using multi-output gradient boosting regression tree,” *Journal of Intelligent Transportation Systems*, vol. 24, no. 2, pp. 125–141, 2020.

[192] D. Mandic and J. Chambers, *Recurrent neural networks for prediction: learning algorithms, architectures and stability*. Wiley, 2001.

[193] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[194] G. Chevalier, “LSTM Cell.” `https://commons.wikimedia.org/wiki/File:LSTM_Cell.svg#filelinks`, 2018. [Online; accessed 20-November-2021].

[195] P. J. Werbos, “Backpropagation through time: what it does and how to do it,” *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[196] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” *Advances in neural information processing systems*, vol. 27, 2014.

[197] Ignatius, “Encoder-Decoder LSTM for Time Series Sequence to Sequence.” `https://tinyurl.com/encoder-decoder-se2seq`, 2018. [Online; accessed 20-November-2021].

[198] F. Wolfe, "Fatigue assessments in rheumatoid arthritis: comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients.," *The Journal of rheumatology*, vol. 31, no. 10, pp. 1896–1902, 2004.

[199] N. Unsworth and R. W. Engle, "The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory.," *Psychological review*, vol. 114, no. 1, p. 104, 2007.

[200] E. Thorndike, "Mental fatigue. i.," *Psychological Review*, vol. 7, no. 6, p. 547, 1900.

[201] D. Bryant, N. D. Chiaravalloti, and J. DeLuca, "Objective measurement of cognitive fatigue in multiple sclerosis.," *Rehabilitation psychology*, vol. 49, no. 2, p. 114, 2004.

[202] J. Heard, C. E. Harriott, and J. A. Adams, "A survey of workload assessment algorithms," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 5, pp. 434–451, 2018.

[203] M. Feffer, O. O. Rudovic, and R. W. Picard, "A mixture of personalized experts for human affect estimation," in *International conference on machine learning and data mining in pattern recognition*, pp. 316–330, Springer, 2018.

[204] G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo, Y. Schwartz, and B. Thirion, "Assessing and tuning brain decoders: cross-validation, caveats, and guidelines," *NeuroImage*, vol. 145, pp. 166–179, 2017.

[205] L. F. Barrett, M. M. Tugade, and R. W. Engle, "Individual differences in working memory capacity and dual-process theories of the mind.," *Psychological bulletin*, vol. 130, no. 4, p. 553, 2004.

[206] R. M. Epstein and E. M. Hundert, "Defining and assessing professional competence," *Jama*, vol. 287, no. 2, pp. 226–235, 2002.

[207] A. Zeki Al Hazzouri, M. N. Haan, Y. Deng, J. Neuhaus, and K. Yaffe, "Reduced heart rate variability is associated with worse cognitive performance in elderly mexican americans," *Hypertension*, vol. 63, no. 1, pp. 181–187, 2014.

[208] P. Nicolini, M. M. Ciulla, G. Malfatto, C. Abbate, D. Mari, P. D. Rossi, E. Pettenuzzo, F. Magrini, D. Consonni, and F. Lombardi, "Autonomic dysfunction in mild cognitive impairment: evidence from power spectral analysis of heart rate variability in a cross-sectional case-control study," *PloS one*, vol. 9, no. 5, p. e96656, 2014.

[209] K. Fuentes, M. A. Hunter, E. Strauss, and D. F. Hultsch, "Intraindividual variability in cognitive performance in persons with chronic fatigue syndrome," *The Clinical Neuropsychologist*, vol. 15, no. 2, pp. 210–227, 2001.

[210] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy, "Learning from crowds.," *Journal of machine learning research*, vol. 11, no. 4, 2010.

[211] R. Gupta, K. Audhkhasi, Z. Jacokes, A. Rozga, and S. Narayanan, "Modeling multiple time series annotations as noisy distortions of the ground truth: An expectation-maximization approach," *IEEE transactions on affective computing*, vol. 9, no. 1, pp. 76–89, 2016.

[212] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Applied soft computing*, vol. 90, p. 106181, 2020.

[213] D. L. Marino, K. Amarasinghe, and M. Manic, "Building energy load forecasting using deep neural networks," in *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*, pp. 7046–7051, IEEE, 2016.

[214] Z. Karevan and J. A. Suykens, "Transductive lstm for time-series prediction: An application to weather forecasting," *Neural Networks*, vol. 125, pp. 1–9, 2020.

[215] N. Voarino, V. Dubljević, and E. Racine, "tdcs for memory enhancement: analysis of the speculative aspects of ethical issues," *Frontiers in Human Neuroscience*, vol. 10, p. 678, 2017.