論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Optimizing Task Design for Screen-based Cognitive Training
著者(和文)	Redlinger Eric
Author(English)	Eric Redlinger
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第12280号, 授与年月日:2022年6月30日, 学位の種別:課程博士, 審査員:小池 康晴,金子 寛彦,小尾 高史,永井 岳大,葭田 貴子
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第12280号, Conferred date:2022/6/30, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Tokyo Institute of Technology
School of Engineering
Department of Mechanical Engineering
Graduate Major in Human Centered Science and Biomedical Engineering

Optimizing Task Design for Screen-based Cognitive Training

Eric Redlinger

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy (Ph.D.)

Word Count: 38,104

Eric Redlinger: *Optimizing Task Design for Screen-based Cognitive Training*, Doctor of Philosophy (Ph.D), © May 2022

Supervisor:
Professor Koike Yasuharu

Location:
Tokyo, Japan

ABSTRACT

The search for an effective means of maintaining cognitive health through conscious intervention has been ongoing for over a century. In the past three decades alone, advances in computing power and display technology have enabled the emergence of a multi-billion-dollar industry that markets and sells screen-based training products. At the same time, the academic community has contributed numerous research studies examining potential approaches and protocols. Frequently investigated topics include training modality (e.g., working memory vs. divided attention) and implementation strategies (how much training, how often, etc.). Although most people now encounter cognitive training in the form of game-like smartphone apps, relatively few studies have examined the graphical design of the training task or the environmental and social circumstances under which training occurred.

With the current status-quo as a starting point, the aim of this thesis was therefore to specifically examine to what degree screen size, training environment and the game-like visual features that accompany today's popular training products might be impacting training outcomes. A secondary goal was to establish whether head-mounted displays (HMDs) might be a particularly suitable vehicle for cognitive training, given their inherent potential for reducing environmental distractions and focusing user attention during training sessions. Through a series of experiments using both HMDs and traditional displays, subjects performed cognitive tasks with different screen sizes and with varying levels of gamification. Biopotential (EEG) and performance outcomes were recorded and compared to determine the impact of individual interventions.

My findings show that the use of an HMD is advantageous not just for conducting cognitive studies, but potentially as a platform for end-users as well. Compared with a standard flat screen, HMD-based cognitive task performance was accompanied by significantly higher theta power in the frontal area, possibly signifying increased concentration and engagement with the task. The precise control over placement and size of stimuli afforded by HMDs also helped me determine that a stimulus size corresponding to a visual angle of approximately 20° resulted in optimal visual memory performance. This implies that small smartphone screens might not be the most efficient way to train. In contrast, none of the game-like conditions tested resulted in any significant performance handicaps.

In summary, the presentation and size of cognitive training tasks is clearly a factor in achieving optimal performance. There is, however, no evidence in my data to imply that simple game-like elements applied to the task have a negative impact. In fact, since longer-term benefits appear to manifest only after a significant investment of time in training, the increased subject motivation and engagement obtained from gamified task design might be instrumental in achieving satisfactory results.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my family for raising me under the motto "you can do anything if you set your mind to it".

Secondly, the tremendous support I derive from my devoted group of local and international friends has been absolutely essential. You all have inspired, encouraged and criticized me in equal measure to get me to do my best work. Many of you have had to endure endless semi-lectures on perception research and cognitive assessment methods when all you really wanted was to share a good glass of draft beer and lament about pandemics and politics.

I owe a similarly great debt to professors Tohru Yagi, Takako Yoshida, and Yasuharu Koike, along with their staff, for mentorship and guidance. I came to the Tokyo Institute of Technology with an undergraduate degree in psychology, an MS in engineering and a lot of enthusiasm. It was my advisors and mentors at Tokyo Tech who helped me to turn a broad range of ideas and potential experimental directions into the concrete goals that, with this thesis, hopefully make a meaningful contribution to the fields of cognitive training and human-computer-interaction.

Finally, none of this would have been possible without the assistance of numerous fellow students and lab members from Yagi, Yoshida and Koike labs, as well as visiting scholars from Australia and Germany. I'm especially looking at you: Graham, Yang, Claudia, Bernhard, Kim and Laura!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3D-MOT	Multiple Object Tracking
ACTIVE	Advanced Cognitive Training for Independent and Vital Elderly
BOLD	Blood Oxygen Level Dependent
CAVE	Cave Automatic Virtual Environment
CCT	Computerized Cognitive Training
CLT	Cognitive Load Theory
CT	Computerized Tomography
EEG	Electroencephalogram
EOG	Electrooculography
FFT	Fast Fourier Transform
FM	Frontal Midline
Gf	Fluid Intelligence
HMD	Head-Mounted Display
HRV	Heart Rate Variability
MIM	Motivational Intensity Model
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
SNR	Signal to Noise Ratio
SSVEP	Steady State Visual Evoked Potential
TP	Time Pressure
UFoV	Useful Field of View
VR	Virtual Reality
WM	Working Memory

1. INTRODUCTION –THE MEDIUM AND THE MESSAGE

I can still remember the moment when I became aware that the goal for the daily crossword puzzle that was delivered as part of our New York Times digital subscription was no longer to merely finish the puzzle, but to finish it within a given time-limit. The puzzle had always been the domain of my father and stepmother but as the necessity to conquer it increasingly became a non-negotiable part of our family's daily routine, I could no longer simply dismiss it as a generational pass-time. Its timely completion now occasionally impacted dinner preparation and other activities that I was far more interested in, so I sometimes found myself joining in to insure its speedy conclusion. On such occasions, I was further aware that subtle, encouraging sound effects accompanied our efforts and that our completion times were now being compared with others in an attempt to encourage us to improve our pace. In short, the humble, self-paced, solitary crossword puzzle experience had been transformed into a sometimes nerve-wracking, socially charged activity through the addition of time limits and various motivational elements. In the language of cognitive science, this process is known as "gamification."

Gamification of tasks to increase user motivation and engagement has wide-reaching implications for numerous industries. This is particularly apparent in the nascent field of screen-based cognitive training (aka brain training), where classic memory and logic tasks are repackaged into game-like applications (apps) for use on personal electronic devices. In fact, many individuals now encounter cognitive training for the first time via such apps. In exchange for just a few minutes of daily use, manufacturers promise a variety of benefits including improved memory, quicker recall and generally improved cognitive health [see section 2.7].

The appeal of such products is easy to grasp: just as my parents discovered as they entered their 60s and 70s, enjoying a high quality of life in one's later years is inextricably linked with maintaining good brain health. As modern medicine enables longer and longer life expectancies, the necessity for a truly effective cognitive training solution grows increasingly urgent. Yet, as an ever-larger portion of our day is spent engrossed in small screens, surrounded by countless competitors for our attention, the question arises: how can we be sure that the time we devote to keeping our brains active and healthy is being well spent? Moreover, is the current iteration of game-like training apps that are designed to be used in short, "on the go" sessions the best way to realize these cognitive health goals?

Just before his death in 1980, the eminent social philosopher Marshall McLuhan famously described his contemporary media landscape with the phrase "the medium is the message". McLuhan's intended meaning was that the means by which information is conveyed has the power to determine or alter its very contents. If this observation is applied to the current state of cognitive training, the logical conclusion must be that the trend towards gamified, portable apps runs the risk of relegating the very important matter of our brain health to one guided not by science but instead by entertainment and convenience. On the other hand, we must also acknowledge that in spite of our desire to do everything we can to insure our cognitive health, our busy lives may in the end leave us only the option of performing a quick, gamified version of a task on our personal devices, or none at all.

Given these complex realities, the goal of this thesis is not to find fault with, or attempt to completely remake, current cognitive training practices. Rather, I view the status quo instead as one that can almost certainly be improved. By systematically and thoroughly examining the recent history of screen-based cognitive training, filling in gaps along the way with original research, I hope to identify concrete ways in which we can increase the efficiency and effectiveness of our training. Ultimately, I feel it is only through a better understanding of the *message* (the science of cognitive training) that we will be able to optimally adapt the *medium* (the design and implementation of training tasks) to insure the best outcomes for our cognitive training efforts.

1.1. 40 Years of Screen-Based Cognitive Training Research

Decades of clinical research notwithstanding, the validity of cognitive training as a whole remains controversial. While the number of studies that show benefits of at least some types of screen-based cognitive training continues to grow, studies that report little or no benefit from cognitive training also exist in substantial numbers (Kable et al., 2017; Owen et al., 2010; Sala & Gobet, 2019; Souders et al., 2017; Stojanoski, Lyons, Pearce, & Owen, 2018). In 2017, the journal *Neurology Today* went so far as to publish an article under the provocative title "Is Successful Brain Training Fake News?" in which the authors cautioned against unrealistic expectations from cognitive training (Fitzgerald, 2017).

On the other side of the spectrum, numerous encouraging studies have demonstrated benefits ranging from better scores on standard cognitive assessment tests (Turunen et al., 2019; Wolinsky, Vander Weg, Howren, Jones, & Dotson, 2013) and improved performance in driving aptitude tests (Eramudugolla, Kiely, Chopra, & Anstey, 2017) to general functional gains in memory, attention and visual-spatial ability (Chi, Agama, & Prodanoff, 2017; Conklin et al., 2017). One particularly influential study, best known by its initials, ACTIVE (Advanced Cognitive Training for Independent and Vital Elderly), made news headlines in 2017 when it declared that a kind of adaptive, speed of processing task known as "Useful Field of View training" resulted in a significantly decreased risk of dementia up to ten years after the training intervention (Edwards et al., 2017).

What defines success in cognitive training? In the absence of results from longitudinal studies (ACTIVE being a notable exception) the most common metric for success is the so-called "transfer effect", or whether cognitive gains in a trained task lead to functional gains in related areas (see section 2.4 "Cognitive Training Tasks"). The presence or lack of short transfer (closely related tasks) or long transfer (slightly related or unrelated tasks) is a commonly accepted benchmark for determining whether the cognitive training intervention was successful or not. In general, studies claiming successful training interventions (e.g. Morrison & Chein, 2011) are able to show statistically significant transfer effects while skeptical studies (e.g. Owen et al., 2010; Stojanoski et al., 2018) are not. Additional commonly accepted assessment outcomes include standard psychological measures of cognition and psychosocial function (Hill et al., 2017).

With a general agreement on terminology and so many researchers examining the subject over an extended period of time, why is there still such a discrepancy in study outcomes? Many potential reasons have been suggested, including a lack of agreement on experimental methodology, outcome assessment methods, and the design and implementation of the cognitive training tasks themselves (Edwards, Fausto, Tetlow, Corona, & Valdés, 2018; García-Betances, Cabrera-Umpiérrez, & Arredondo, 2017). An example of the latter can be found in a recent study by Linares et. al., which found no near-transfer effect even between very similar working-memory tasks (Linares, Borella, Teresa Lechuga, Carretti, & Pelegrina, 2019). Closer inspection of their protocol, however, reveals that the training task in each case deviated from standard convention in that the task design was non-adaptive (i.e. task difficulty was not adjusted to match subjects' natural abilities or prior training gains). This one detail alone may have negatively impacted the study results as recent findings argue that the use of an adaptive task design is essential to the success of cognitive training (Edwards et al., 2018; Ross et al., 2019). In addition, environmental factors may have contributed to the lack of observed effect in this study as the assessment sessions were supervised by study staff whereas the training sessions were not.

Another factor potentially leading to discrepancies in study findings involves the choice of subjects. While many cognitive training studies target older individuals or those with mild cognitive impairment, there is evidence suggesting that the benefits from training are greatest if that training occurs *before* the onset of age-related cognitive decline. For example, Zając-Lamparska and her colleagues conducted a battery of training tasks with two separate groups of older individuals (60+), a mild impairment group and a no impairment group. They found significant gains only in the group of mentally fit individuals, concluding that cognitive training is a reasonable counter-measure for the decline experienced in normal cognitive aging but that waiting until signs of dementia are found is too late for beginning training (Zając-Lamparska et al., 2019).

In short, conflicting research findings may be the product of differing protocols, task modalities, training environments, or even just the criteria by which subjects are selected. The resulting lack of consensus hinders the creation of training guidelines and presents challenges for creators of cognitive training products and their users.

1.2. Thesis Aim

This thesis seeks to establish whether we might be able to improve cognitive training outcomes by taking a closer look at the tasks used in common training regimens and the manner in which the tasks are implemented. Specifically, I aim to accomplish this by answering three questions:

> Does the introduction of game-like elements affect cognitive task performance and if so, which elements are primarily responsible for any observed effect?
>
> Does the size of the screen that we use impact the effectiveness of training? (is there an optimal size?)
>
> In light of recent technological advances in display technology, might HMD-based cognitive training be the best way to insure a precisely controlled, distraction free training environment?

By addressing these questions, this thesis aims to make a significant contribution to our understanding of screen-based cognitive training and help establish a "best practices" guide for creators and users of cognitive training products. As mentioned previously, the aim is not to find fault with current trends in cognitive training, but rather to provide evidenced guidelines to improve task design and avoid potentially counterproductive elements.

1.3. Thesis Scope

This thesis is necessarily interdisciplinary and requires that I examine literature from many different fields. In the interests of focus and brevity, there will be some fields that I naturally draw upon more frequently than others. The original research described in this thesis is also constrained by the 3-year time-period of a standard PhD, the laboratory budget, and by the available subject pool. As a result, it is important to acknowledge several limitations of this research from the onset.

1. I do not attempt to answer the question "does cognitive training work?" Providing an adequate answer to this question would require longitudinal data spanning a substantial period of time and access to a diverse subject pool. Furthermore, such data is in part already available (Edwards et al., 2017) and serves as an important point of reference for this thesis. By focusing on the psychological and psychometric aspects of cognitive task design, along with aspects related to the implementation of training, I hope to make an important contribution to the field that remains within the scope of what is possible.

2. Directly comparing variously sized screens across different display technologies introduces too many variables to be able to reach a satisfactory level of construct validity. Instead, I adopted an approach of first validating the use of an HMD for conducting psychometric experiments. I then conducted the bulk of my experiments on this platform. I found that a single visual environment with a common brightness, resolution and precise stimulus positioning capabilities provided the best opportunity for exploring the cognitive impact of various stimulus sizes and novel visual features such as 3-D depth.

3. I do not evaluate the motivational and engagement aspects of gamification. There is ample prior research addressing this and a working consensus has emerged that gamified design patterns do indeed increase subject motivation and satisfaction (Lumsden, Edwards, Lawrence, Coyle, & Munafò, 2016; Vermeir, White, Johnson, Crombez, & Van Ryckeghem, 2020). Rather, the oft-acknowledged heterogeneity and lack of precision in past experiments with regards to the performance impact of specific game-like elements provides me with ample opportunity to make a potentially valuable contribution.

4. There is a great diversity of training and assessment tasks employed in cognitive research. For the original research described in this thesis, I was faced with the challenge of finding or creating tasks that incorporated the most promising elements and then adapting those tasks to operate in a virtual reality environment. This involved careful consideration not just of task effectiveness but also protocol strategies to minimize individual subject differences and boost statistical power. The exact process of selecting which cognitive functions would be addressed for each experiment is described in detail in their respective *Methods* sections 3.3.1, 4.3.1 and 5.3.1. While it is possible that the study results might have been different had I used other cognitive tasks, I ultimately feel the choices I made allowed me to adequately meet my research objectives.

1.4. Thesis Overview

This thesis contains six chapters. Chapter 1 includes an overview of the current state of screen-based cognitive training and establishes the scope of this thesis.

Chapter 2 systematically lays out the research relevant to this thesis, including sections on cognitive function, training strategies, assessment methods and cognitive task design principles for screen-based training, including the recent use of virtual reality. It also includes an overview of the ongoing cognitive training controversy, as well as detailing the challenges inherent in examining a field that encapsulates, on the one hand, a legitimate academic research discipline pertaining to cognitive function and human-computer interaction and, on the other hand, a commercial industry that tends to promote only positive outcomes.

Chapters 3, 4 and 5 describe a trio of experiments that were conducted during 2020 and 2021 to address the research needs defined in section 1.2, "Thesis Aim". The chapters draw largely from the journal articles and conference presentations in which they were originally published. The experiments are as follows:

Experiment 1 (chapter 3) began as a simple feasibility study to determine whether scalp EEG could be reliably measured while simultaneously wearing an HMD. After initial validation, the study protocol was expanded to include EEG-based evaluations of a multi-modal visuo-spatial task performed in both VR and non-VR environments. The goal was to document any potential variation in the cognitive response between the two modalities while performing an otherwise identical task. A broad-spectrum EEG analysis revealed the presence of statistically significant differences in frontal theta power between the two experimental conditions, compared with a control. These results subsequently led to the formulation of a hypothesis regarding the relationship between stimulus size (or perceived stimulus size) and cognitive effort. This hypothesis became the basis for the subsequent experiment.

Experiment 2 (chapter 4) introduced a simpler visual memory task, inspired by the well-documented *n-back* task. The primary purpose of this cognitive task, which targets working memory, was to elicit sustained cognitive load in the frontal area that could be modulated by manipulating task difficulty. To eliminate the possibility of data contamination resulting from environmental and social distractions, it was decided that the entirety of the experiment would be conducted using an HMD. This setup had the additional advantage of making calculations related to stimulus brightness, size and position much more reliable, compared with a traditional LCD-monitor. Results confirmed the study hypothesis that stimulus size had a significant impact not just on EEG power, but also on task performance. Moreover, the data revealed that the optimal stimulus size for maximizing task performance corresponds to a roughly 20° visual angle, coinciding with the anatomical limits of the macular area in the human retina.

Experiment 3 (chapter 5) expanded on the previous study by fixing the visual angle at 20° for all experimental conditions and introduced a series of game-like visual elements to the core task as independent variables. The choice of visual elements included 3D depth, an element rarely examined in previous game-features impact studies due to the relative novelty of virtual reality headsets in cognitive training research. This represented one novel element of the study. Another novel element was the concurrent use of both task performance and EEG outcomes, enabling the possibility of direct correlations for each independent variable. Previous research was inconclusive regarding the performance impact of game-like features, with some studies documenting substantial effect sizes and others finding only a small effect or none at all. My study also ultimately found no significant differences for any of the conditions tested. An extensive discussion of the results at the end of chapter five points to possible explanations for these findings.

Finally, chapter 6 synthesizes the findings of all three experiments and offers an initial proposal for "best practices" screen-based cognitive training design, derived both from the results of my original research and an extensive prior literature review. I conclude by revisiting the limitations of the results presented in this thesis and offer my recommendations for possible future research directions.

2. BACKGROUND AND SYSTEMATIC REVIEW

Any analysis of cognitive training must necessarily start by defining which cognitive processes we are seeking to improve or maintain through training. Generally, these fall under the broader category of "executive functions," with a particular emphasis on working memory. A description of the key executive functions, along with related concepts such as *fluid intelligence*, comprises the first section of this chapter.

This will be followed by sections dedicated to the following topics:

- Methods for assessing cognitive function
- Traditional cognitive training tasks
- Commercial adaptations of traditional tasks
- Action games research and gamified tasks
- New possibilities for screen-based training, including virtual reality
- Design principles for constructing training tasks

Finally, I will discuss specific considerations for my research in light of the prevailing trends in training studies and the availability of novel hardware, including portable, wireless EEG amplifiers and virtual reality headsets.

2.1. Executive Functions, Working Memory and Fluid Intelligence

The body of mental processes that are responsible for the cognitive control of behavior are generally referred to as *executive functions*. They include planning, organization, decision making, problem solving, attention to detail, remembering, time management, spatial scanning, inhibition and self-control (Miyake et al., 2000). In addition, the core executive functions are often operationally summarized into three principle categories: working memory, inhibition, and mental flexibility (Lumsden, Edwards, et al., 2016).

Working memory (WM) refers to the brain's short term repository for items and is usually considered to be limited to 7-8 elements that are stored for no more than 10 seconds (Logie, Horne, & Pettit, 2014). WM has been linked to general cognitive capacity perhaps more than any other cognitive process and often serves as a proxy for cognitive ability in general.

A related concept, fluid intelligence (Gf), refers to the ability to reason and to solve new problems independently of previously acquired knowledge. While first proposed in 1963 by Raymond Cattell, the term has since gained broad acceptance and is frequently employed in working memory studies in particular (Susanne M. Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Unsworth, Fukuda, Awh, & Vogel, 2014).

Part of the enthusiasm surrounding WM training is due to the fact that it has been shown to improve or maintain cognitive abilities not just in older individuals, who are at greatest risk of declining function, but also in younger subjects. For instance, Artuso et al. reported that significant gains were obtained by 9-10-year-old children in both text updating and reading comprehension tasks in their 2019 study (Artuso, Carretti, & Palladino, 2019). Notably, they found a *far transfer* effect for both types of training. Borella et. al. found a similar result for students as young as 8 years old, with WM training reliably boosting reading comprehension skills up to 2 months later in their examination of 48 elementary school students (Borella, Carretti, Riboldi, & De Beni, 2010).

Hypotheses as to why working memory training seems to more reliably elicit long term gains in fluid intelligence and related executive functions tend to be divided into two schools of thought. The first of these proposes that training really just teaches us improved memory strategies, rather than improving our memory capacity per se. Grouping techniques such as *chunking* can be learned (or spontaneously acquired) and subsequently applied in unrelated memory tasks. In support of this theory, a 2019 paper by Meiran and Dreisbach entitled "Mechanisms of working memory training" notes that subjects use less and less actual WM capacity over time as they develop strategies (Meiran, Dreisbach, & von Bastian, 2019). Artuso and his team also acknowledged this possibility in their reading comprehension study, noting that while far transfer gains were significant, *near transfer* effects to other WM training were much more limited and they speculated about whether the reported gains might, in fact, be due to the acquisition of learning strategies (Artuso et al., 2019).

A second possibility is that correlations between WM and Gf may instead be due to some entirely separate higher-order function. Meiran and Dreisbach, for example, noted that the link between WM training and gains in Gf through correlation alone cannot be considered conclusive. They offer the example of grip strength: right hand grip strength is highly correlated with left hand grip strength, but benefits derived from physically training only the left hand will not transfer to right hand strength (Meiran et al., 2019).

Jaeggi et al., authors of an influential 2008 study that first introduced the popular *dual n-back* task, countered in a 2011 paper that attention and working memory may instead rely on similar neural networks. They further confirmed that WM training using n-back was particularly effective in achieving long term transfer gains, as shown by Gf tests conducted three months after initial training was complete (Susanne M. Jaeggi, Buschkuehl, Jonides, & Shah, 2011). Nevertheless, Studer-Luethi and Meier found that training tasks targeting other executive functions such as inhibition were just as effective as n-back, which is to say not very effective. They reported no far transfer for either n-back nor a separate auditory inhibition task in their 2020 study (Studer-Luethi & Meier, 2020).

As we shall see repeatedly when comparing studies that report differing results, however, small changes in protocol or nomenclature (for instance, how to define near/far transfer) may have an outsized effect on the study outcome. In short, a lack of shared methodology often hampers direct comparisons between study findings.

2.2. Measuring Cognitive Function with EEG

Encephalograms, or EEGs, are a widely used technology to assess cognitive activity. In its simplest form, it involves the placement of electrodes onto the surface of the human scalp. These electrodes, in combination with a powerful signal amplifier, are able to capture and record synaptic activity at the microvolt level. The accumulated raw data are then commonly separated into their component frequency ranges using a technique such as *fast Fourier transform* (FFT) and finally data-mined for specific patterns of activity, or *features* (Wolpaw & Wolpaw, 2012). While not as precise as other methods of measuring brain function, such as MRI or CAT, EEGs have the advantage of portability, processing speed and temporal precision. An in-depth exploration of the history and use of EEGs goes beyond the scope of this thesis, but as it plays an important role in the original research presented in this thesis, a closer look at prior studies linking EEG data with specific cognitive processes is warranted.

Most individuals reading this thesis have probably encountered the rows of wavy lines associated with EEG data visualization (figure 1).

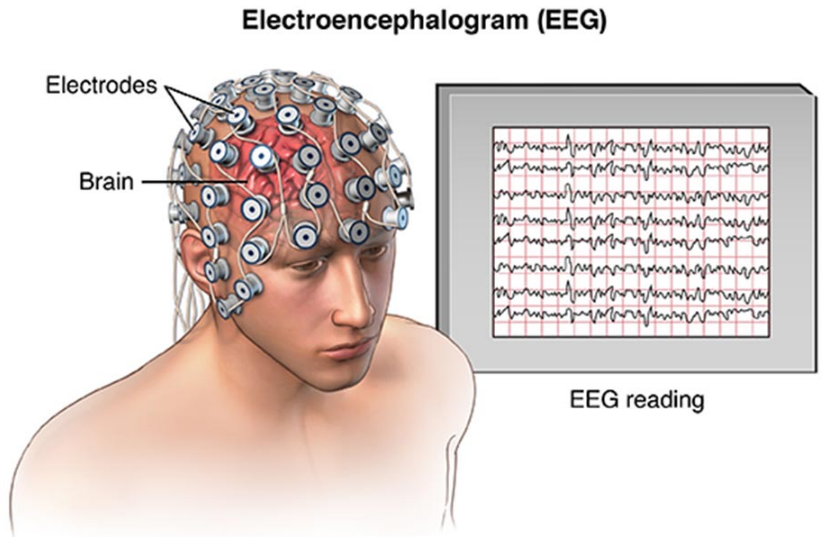


Figure 1: *EEG data is commonly visualized in the time-dimension with rows corresponding to the activity captured by individual electrodes. (Image credit: mayoclinic.org)*

Like other direct visualizations of brain activity such as the aforementioned MRI or CAT scans, these images impart an air of scientific irrefutability. Decades of mostly uncritical use of such imagery in the popular press has led many casual observers to place an often-unmerited trust in media reports that make reference to such technologies. A 2015 New York Times article, for example, famously informed its readers that brain scans *proved* that fans of Apple Computer, Inc.'s products loved their iPhones because, the article claimed, scans showed activations in the "pleasure center" area of the brain during use of the popular smartphones (Lindstrom, 2011).

The interpretation of data obtained from scanning technologies such as EEG is, in reality, much more complicated. Results often rely on correlations that are highly specific to the particular conditions of the observed task. In the aforementioned example, for instance, *disliking* one's iPhone could just as easily have generated activations in the exact same area of the brain (amigdala) which, contrary to popular belief, is associated with not just one but a host of different emotions and cognitions (Ressler, 2010).

While claims of one-to-one associations between specific brain locations and individual human emotions are best approached with a healthy dose of skepticism, biometric technologies such as EEG do excel at establishing correlations between broader *categories* of cognitive activity and their related areas of cortical activation. One such well-documented relationship involves the executive functions of attention and memory, and corresponding cognitive activity in the frontal and parietal cerebral regions. A long history of experiments extending back more than two and a half decades has documented clear patterns of synchronization between task-induced working memory load and frontal hemisphere activity in the theta frequency range (4-8Hz) (Başar, Başar-Eroglu, Karakaş, & Schürmann, 2001; Alan Gevins, Smith, McEvoy, & Yu, 1997; Holm, Lukander, Korpela, Sallinen, & Müller, 2009; Inanaga, 1998; Kahana, Seelig, & Madsen, 2001; Klimesch, 1996, 1999; Pope, Bogart, & Bartolome, 1995; Sauseng et al., 2005; Schacter, 1977; Wilson & Russell, 2003; Yamada, 1998). Our understanding of this relationship was further extended by subsequent research showing that those same cognitive demands also often result in a marked suppression of alpha wave activity (8-13Hz), particularly in the parietal area (Antonenko, 2007; Ewing, Fairclough, & Gilleade, 2016; Fairclough & Venables, 2006; Fournier, Wilson, & Swain, 1999; Alan Gevins et al., 1998; Ryu & Myung, 2005; Sterman & Mann, 1995).

Because I, too, am interested in tracking working memory load for my research, I decided to do a small feasibility study (n=4) to see if I could achieve similar results. While subjects alternated between resting and performing a simple one-back working memory task I recorded EEG signals in the frontal area. I found that my results largely matched the findings documented by prior researchers: engagement with the cognitive task led to an immediate spike in theta activity at electrode position Fz and a corresponding suppression of spectral power in the alpha range. Plotted spectral data for a typical subject can be found in Appendix A.1.

Another advantage of using EEGs to measure task engagement is that EEG spectral amplitude (the amount of measured EEG power) has also been shown to be extremely sensitive to minor changes in the degree and type of cognitive demand. For instance, Naumann et al. found that they were able to accurately distinguish between multiple difficulty levels during video game play by observing only the theta and alpha rhythms of their subjects (Naumann, Schultze-Kraft, Dähne, & Blankertz, 2017). Similarly, Slobounov and his colleagues found that merely adding a time pressure component to their cognitive task was sufficient to significantly impact the observed theta and alpha power (S. M. Slobounov, Fukada, Simon, Rearick, & Ray, 2000). Most recently, Puma and his associates confirmed a consistent relationship between cognitive load and theta EEG power in their experiment that manipulated load by varying the number of concurrent tasks (Puma, Matton, Paubel, Raufaste, & El-Yagoubi, 2018). In short, the use of EEGs to document a variety of cognitive function is a validated and well-understood process that is used in numerous studies every year.

2.3. Cognitive Assessment Surveys

Cognitive assessment surveys employ a wide range of methods and strategies. The ones that are the most pertinent to cognitive training cover two basic areas: general intelligence/cognitive impairment and overall cognitive capacity. The former category includes popular batteries such as the RPM/APM (Raven's Progressive Matrices/Raven's Advanced

Progressive Matrices), MMSE (Mini-mental State Examination) and the IADLs (Instrumental Activities of Daily Living scale). Generally speaking, these tests include a series of assessment questions from a variety of cognitive domains and are considered highly reliable with good concurrent validity (Bauco et al., 1998; S. Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963; Monroe & Carter, 2012; Raven, 2000; Schwarz, Diener, & Kahneman, 1999). Of these, the RPM is entirely non-verbal and focusses on reasoning capacity as each test item requires the subject to complete a visual series by choosing an answer that completes a proposed pattern. At the other end of the spectrum, the IADLs asks the subject specific questions about day-to-day capabilities, such as their ability to independently take their medications or do their own laundry (Lawton & Brody, 1969). Differences aside, all of these tests assume at least a minimal level of literacy on the part of the subject.

Cognitive capacity tests, on the other hand, are primarily perceptual tests and are not limited or biased by the literacy level of the subject. They are generally more narrowly focused in scope but nevertheless correlate highly with a variety of cognitive functions, including working memory capacity (Tiego, Testa, Bellgrove, Pantelis, & Whittle, 2018). Of these, the following two tasks deserve a more detailed introduction due to their widespread use.

1) Flanker:

   As a classic test of *attentional inhibition*, the flanker test assesses the subject's ability to quickly evaluate similar-looking stimuli placed immediately to the left or right of the primary stimulus. When the "flanking" stimuli are the same as the central stimulus, the condition is considered congruent, as opposed to an incongruent condition where the stimuli differ (see figure 2).

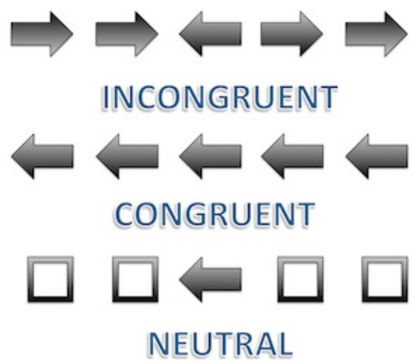


***Figure 2**: Flanker test requires user to compare central target stimulus with "flanking" stimuli.*

Generally, subjects take longer to identify the incongruent state, and the difference between the reaction time for congruent and incongruent responses is assessed as the *interference effect*.

The underlying concept being very simple, many variations on the flanker test exist. For example, some versions use a combination of text and images while some use images alone. The flanker test has been repeatedly validated and remains one of the most widely used cognitive tests due to its adaptability and ease of implementation (e.g., Green & Bavelier, 2003; Lumsden, Skinner, Woods, Lawrence, & Munafò, 2016; Tiego et al., 2018).

2) Stroop:

Response inhibition can be defined as voluntary control over goal-irrelevant stimuli, cognitions and responses (i.e., ability to ignore distractions). The Stroop test, first proposed by John Ridley Stroop in 1929, makes use of this effect to explore how cognitive processing is impacted by conflicting information. In the classic version of the test, the names of colors are written in either the same color ink as their names, or in an alternate color. Subjects are asked to read through the list of colors as quickly as possible but in doing so should ignore the word itself and rather simply announce the *color* that the word is printed in (figure 3).

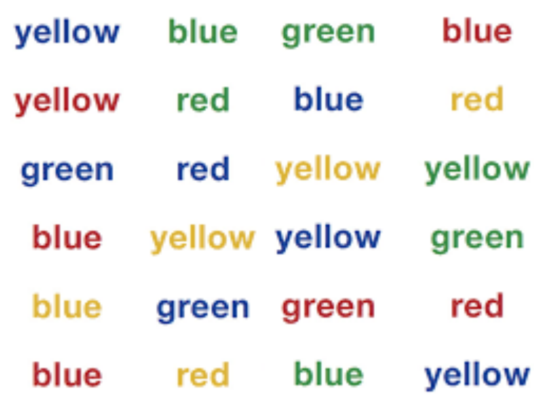


Figure 3: *Stroop's test required the user to say the color the word is printed in, not the word itself.*

Although performance can be improved with training, Stroop found that incongruent instances (meaning and signification are at odds) initially resulted in processing delays of nearly 75% compared with a congruent case (meaning and signification are identical) (Stroop, 1935). Like the flanker test, the Stroop test has also been adapted numerous times, e.g., to allow for the use of images and sound. An example with sound is the Simon test, proposed in 1969 (Simon, 1969). Interestingly, the difference in reaction times for the incongruent condition in the Simon test are not as dramatic as the original color-printed-word test, at about only 30%. This suggests that the inhibition response may be particularly sensitive to visual stimuli (figure 4).

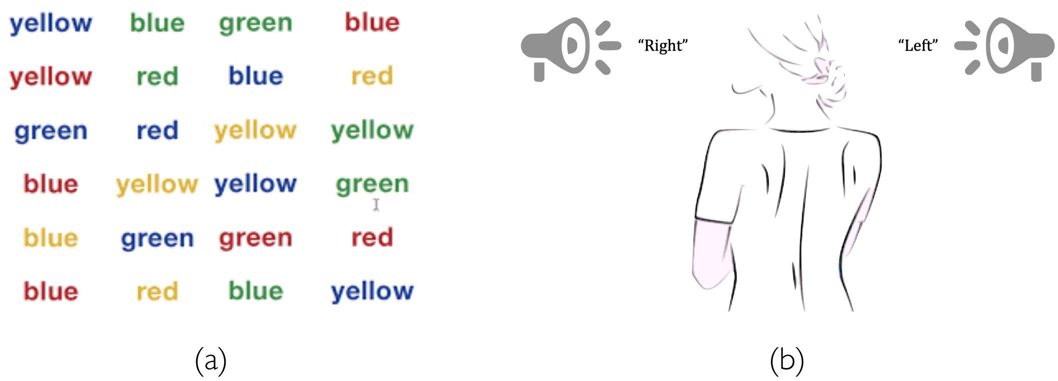


Figure 4: (a) Classic Stroop color/word test; (b) "Simon": Stroop test modified to use audio stimuli

2.4. Cognitive Training Tasks

Over the course of many decades of cognitive research, numerous training tasks have been proposed and tested. Often these tasks are derived from previously validated assessment tests and their aim is consequently to improve performance on those targeted tests. The justification for such training generally uses the following rationale: if the assessment test X is correlated with cognitive functions Y or Z, and the X-derived training task boosts test results in X, perhaps training task X can improve functions Y and Z directly. As previously mentioned (see section 1.1), this training effect is known as the *transfer* effect and can be measured empirically. Transfer is generally further differentiated into *near* and *far* transfer. For example, Jaeggi's 2008 study found a medium strength *near transfer* effect to fluid intelligence measures and a small *far transfer* effect to other intelligence outcomes (S. M. Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). Morrison and Chein, in an extensive literature review, identified numerous reports of strong training-related transfer derived from working memory training (Morrison & Chein, 2011). The extent to which any of these outcomes applies to events in our actual everyday lives is referred to as *ecological validity*. In general, most cognitive training tasks aspire to achieve both near and far transfer gains in their targeted cognitive functions, as well as maximizing the ecological validity of their outcomes.

While an absolute consensus regarding the relationship between cognitive training and ecologically valid transfer effects is still lacking, some tasks stand out simply due to their widespread use and the number of studies that have achieved promising results while employing them. Of these, the following two tasks in particular are essential to the research described in this thesis.

1) N-back:

   The n-back task first appeared in a PhD thesis in 1953 as a visual task to assess age effects on short-term memory and fatigue (Kane, Conway, Miura, & Colflesh, 2007) but was not published in a peer-reviewed journal until 1958 (Kirchner, 1958). It has since been shown to correlate highly with a wide variety of intelligence measures (Friedman et al., 2006, 2008; A. Gevins & Smith, 2000; Salthouse, Pink, & Tucker-Drob, 2008; Shelton, Elliott, Matthews, Hill, & Gouvier, 2010; Van Leeuwen, Van Den Berg, Hoekstra, & Boomsma, 2006; Waiter et al., 2009). Like other tasks whose original purpose was cognitive *assessment*, n-back has now become a popular and widely used *training* task. In the classic n-back task, a subject is presented with a continuous stream of similar stimuli. The subject must then indicate when the current stimulus matches the one from ***n*** steps earlier in the sequence. One important early finding was that while most individuals were able to perform the n-back task with high levels of accuracy at the zero and one-back levels, accuracy at higher levels of n dropped off dramatically, particularly in older individuals (Kane et al., 2007; Mackworth, 1959).

A chief strength of n-back is its adaptability: the stimuli can be letters or other characters, or a series of images. Alternatively, the n-back variable can be a change in the stimulus color, size or position. The classic task can also be easily extended into a multimodal variant. For example, a version of n-back dubbed the "dual n-back" received extensive mainstream coverage in publications such as Wired magazine ("Researchers Develop Software That Makes You Smarter", 2008) when task performance was found to correlate highly with fluid intelligence measures (Susanne M. Jaeggi et al., 2008). While the "dual" nature of Jaeggi's experimental task was initially given credit for the dramatic finding, similar results have since been recorded with just a single mode version of the task (e.g. the classic n-back) (Susanne M. Jaeggi, Buschkuehl, Perrig, & Meier, 2010).

2) UFoV (Useful Field of View) Training:

The concept of a "functional visual field" dates back to 1970 and can be defined as the area of the visual field over which information can be acquired in a brief glance without eye or head movements (Sanders, 1970). Based on Sanders' research, along with other prior visual attention and search studies, Sekuler and Ball developed an assessment test in 1986 that they dubbed the *Useful Field of View* (UFoV) test (Sekuler & Ball, 1986). The test has since been shown to have high ecological validity (Clay et al., 2005) and is highly correlated with important indices of independence and mobility including ambulation, falls, and driving competence (Ball & Owsley, 1993; Broman et al., 2004; Myers, 2000; Owsley et al., 1998; Stalvey, Owsley, Sloane, & Ball, 1999).

The standard version of the test contains several individual subtests to assess processing speed, selective attention and distractor suppression, respectively. The core task requires subjects to identify a previously displayed stimulus from among various similarly shaped distractors. Depending on the subtest, the stimuli may appear in either the central visual area, the peripheral area, or both (see Fig. 5).

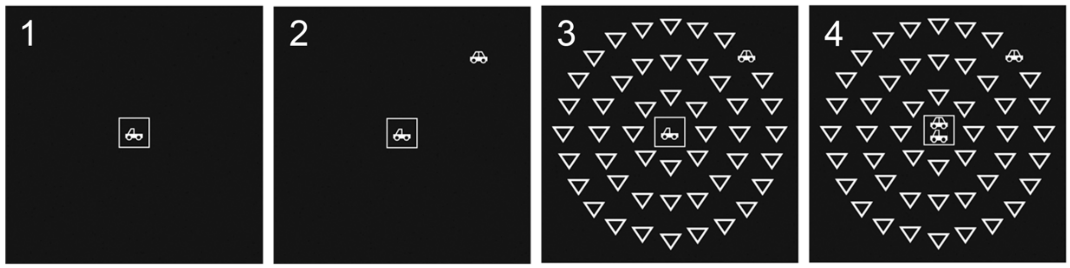


Figure 5*: Original UFoV subtests designed to assess selective attention and distractor suppression capabilities. In subtests two, three and four, both the stimulus type and position in the periphery must be accurately reported to successfully complete the task.*

Like n-back, the UFoV task was originally designed as a clinical assessment tool. In the shift to becoming a *training* task, a number of additional visual and narrative elements are often added to make the task more appealing. These elements include the use of cartoon-like icons, colorful, task-irrelevant background imagery, thematic storylines, scoreboards and others (Fig. 6).

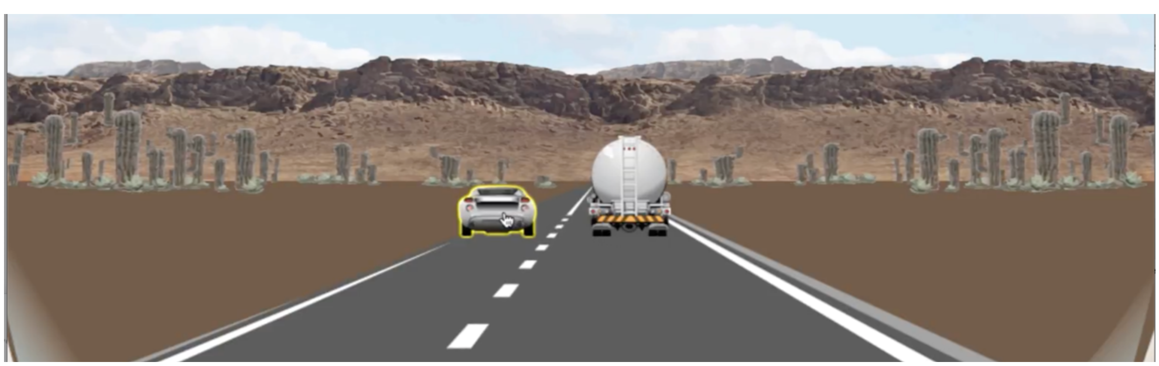


Figure 6*: Commercial UFoV training tasks like Posit Science's "Visual Edge" task often resemble games.*

Such game-like features in training tasks have since become widespread and their use has been hotly debated both inside and outside of the academic community. The following section 2.5 summarizes the debate.

Along with n-back, UFoV training is currently considered among the most promising cognitive training strategies. A comprehensive systematic review and meta-analysis published in 2018 found it to have the largest gains vs. controls of all the computerized training types tested. In particular, UFoV training was found to be linked to improved mobility and reduced medical expenditures, among others (Edwards et al., 2018). Furthermore, studies cited in the review showed that general cognitive gains, as determined by the Activities of Daily Living scale, were still robust up to 7 years following training.

The documented long-term durability of UFoV training seems in part to be linked with the multimodal nature of the task. An early fMRI study found that functional connections among brain regions within the greater cognitive control network were strengthened during UFoV training, suggesting that the neural systems of trained individuals became tuned to more efficiently utilize neural resources (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008). A more recent study also using magnetic resonance imaging technology similarly found increased functional connectivity in multiple regions of interest, including both the executive function and visual areas of the brain (Ross et al., 2019). Crucially for commercial applications, UFoV has additionally been shown to be robust in settings outside of a clinical environment, such as at home. A 2013 study found its effectiveness in home settings comparable to a controlled, clinical environment (Wolinsky, Weg, Howren, Jones, & Dotson, 2013). Most recently, television-based home training with a task similar to UFoV was associated with improvements in several cognitive domains including verbal memory, inhibition and general processing speed (Nouchi, Kobayashi, Nouchi, & Kawashima, 2019).

2.5. Gamified Cognitive Training

Gamification is generally defined as the process of adding digital game elements into non-entertainment settings to increase (intrinsic) motivation and engagement (Vermeir et al., 2020). Given the large number of features that could potentially be classified as game-like, a number of taxonomies have been proposed (e.g., Marczewski's "Periodic Table of Gamification Elements") that further classify features into sub-categories such as socially-oriented or achievement-oriented gamification elements (Marczewski, 2015). Commonly encountered game-like features include thematic imagery, animations, score boards and real-time progress feedback.

When coupled with cognitive tasks specifically designed to maintain or improve one's cognitive abilities, the result may be referred to as *gamified cognitive training*. The recent widespread availability of such gamified training products in the form of "apps" (applications) on smartphones and tablets, along with a growing public awareness of cognitive training in general through mainstream media coverage, have all contributed to the creation of a multi-billion dollar industry (MarketsandMarkets Research, 2020).

The primary objective of gamified training, i.e., increasing subject motivation by incorporating game-like elements, is supported by a large number of studies and surveys in both the cognitive training and games research domains (Boendermaker, Veltkamp, & Peeters, 2017; Mekler, Brühlmann, Brühlmann, Opwis, & Tuch, 2013; Mohammed et al., 2017; Vermeir et al., 2020). Moreover, increased motivation and engagement has been shown to lead to improved quantitative results. For example, research results from 2018 showed that subjects who received explicit rewards for completion of a declarative memory task outperformed individuals who received no such rewards (Prena, Reed, Weaver, & Newman, 2018). Similarly, Gooding and her team compared three different types of computerized cognitive training and concluded that training techniques incorporating motivational strategies appear to provide more benefit than training alone (Gooding et al., 2016).

Where the current state of research is less clear, however, is in our understanding of exactly which game-like elements, or combination of elements, may be responsible for improved outcomes. A recent, comprehensive review looking into the use of gamification strategies in cognitive training and assessment studies overwhelmingly found that while gamified training does boost participant motivation, study heterogeneity impeded the drawing of clear conclusions with regard to training gains (Lumsden, Edwards, et al., 2016). Notably, the study authors identified no fewer than 28 game-like elements employed in the 33 studies surveyed. These included positive/negative task feedback, time-pressure, storylines or narrative elements, interactive status displays and many others.

In addition to a diverse range of independent variables, Lumsden and Edwards also documented a large number of differing experimental outcomes. The majority of studies surveyed sought to measure user satisfaction, engagement and subject retention, while performance and ecological validity (i.e., the transfer of trained skills into real life tasks) were among the least often assessed. Of the studies that did look specifically at the impact of gamification on performance, the game-elements examined were often limited to minor interface adjustments, modified subject instructions, or they were not sufficiently granular to extract meaningful design guidance (Lumsden, Edwards, Lawrence, Coyle, & Munafò, 2016). Mohammed et al., for example, compared two adaptations of an n-back task, including one that in their own words "involves a visual rich display, multiple soundtracks, as well as navigation challenges that can potentially interfere with improving memory per se" (Mohammed et al., 2017). Thus, while some game elements might have contributed to user engagement and task performance, others might have simultaneously served to impede or counteract those very gains.

A similar conclusion was reached in another well-regarded study from 2014 with a sizable subject pool (n=107). Katz et al. found that unneeded stress and new cognitive demands may have been induced by distracting game elements such as persistent score displays, leading to reduced performance. However, rather than individual game elements added one-by-one to an unmodified task, the study design instead removed specific game elements from a larger group of game features. This approach seems to leave the possibility open for the remaining elements to compensate for the removal of a single one, making it difficult to know for sure which element(s) might have specifically accounted for the *new* cognitive demands (B. Katz et al., 2014).

Finally, a second, more recent, systematic review of the effects of gamification on computerized cognitive training was undertaken by Vermeir et al. in 2020. Echoing the results of Lumsden's survey, Vermeir and her colleagues similarly found that of the 49 papers examined, "no study has reported on the effect of a single element alone" and that "game elements were…investigated only in combination, making it impossible to establish whether individual elements had measurable effects" (Vermeir et al., 2020).

From these and other studies, it is clear that one of the primary challenges in designing gamified training tasks is to strike the right balance between motivating the subject on the one hand, and avoiding unintended complication and distraction that might impede task performance on the other. As Katz and others have shown, the danger of inadvertently introducing elements that weaken the primary training goals through the generation of new and unrelated cognitive demands is very real (B. Katz et al., 2014b; Miranda & Palmer, 2013). More accurate differentiation and classification of different gamification elements is one potential way to avoid such issues.

2.6. Action Gaming Research

Complementing the study of gamified tasks, the parallel field of *games research* includes many of the same elements as dedicated cognitive training studies. Rather than employing bespoke tasks that target specific cognitive processes, however, games researchers generally group existing commercial games into specific genres that reflect the characteristics of the game style, and therefore the underlying cognitive processes that are engaged. Within these classifications, the *action game* genre has received particular attention due to its association with increased concentration capacity for active players. Of particular interest for proponents of cognitive training was Bavelier and Green's 2003 publication, "Action video game modifies visual selective attention" in the journal Nature (Green & Bavelier, 2003). This study documented both the aforementioned increased attentional capacity for players of action video games, as well as the likelihood that these differences were not innate but acquired. Notably, the authors demonstrated that similar attentional gains could be acquired by previously non-gamer subjects through training. Bavelier speculated that the cognitive processes inherent to action video games that specifically contributed to cognitive gains included time pressure, a constant need to select targets from among distractors, and divided attention.

In their 2020 paper, "Enhancing attentional control," Cardoso Leite et. al., echoing Bavelier, similarly proposed the following three key mechanics that must be present in an interactive environment in order to foster augmented attentional control:

(1) pacing, or the need for making decisions under time pressure
(2) divided attention, or the need to filter distractors and sustain attention over a large part of one's environment for a significant period of time
(3) the need to switch between modes of processing

If these conditions sound familiar it is probably because are remarkably similar to those proposed by Dr. Edwards as being critical to the success of cognitive training tasks in order to achieve long transfer and ecological validity (Edwards et al., 2018).

Despite striking similarities, research investigating the effects of gamification on cognitive training outcomes occasionally diverges from pure *action games* research. One primary difference has to do with the definition and employment of key terms. For example, the cognitive function literature differentiates between different methods of evaluating attention, including *distraction suppression* and *target recognition*. Traditional cognitive training generally focusses on the former, emphasizing the ability to ignore increasingly salient distractors (Mishra, Kumar, Padmanabhan, & Gulyás, 2021). Action gaming, however, primarily targets the latter, rewarding players who identify potential targets as quickly as possible. As a result, action games training has been found to primarily result in enhanced thresholds on perceptual tasks, such as those that require participants to identify low contrast targets or small letters in crowded visual fields (Green & Bavelier, 2007; Li, Polat, Makous, & Bavelier, 2009).

The experimental findings of the two research disciplines are also used in slightly different contexts. For example, one direct application of the results of action games research is, unsurprisingly, in the field of professional gaming, including elite sports. Indeed, 3D Multiple Object Tracking (3D MOT), one of many tasks inspired by action games research, along with other related measures of attentional control, appears to be a good predictor of a variety of skills in professional athletes (Qiu et al., 2018). Furthermore, these relationships are not just correlational. Dedicated 3D MOT task training has been reported to improve not just attention but also visual information processing speed and working memory in young and older adults (Legault & Faubert, 2012; Parsons et al., 2016; Vartanian, Coady, & Blackler, 2017). It has additionally been shown to directly impact soccer player performance on the field (Romeas, Guldner, & Faubert, 2016). Consequently, training based on MOT is now included in the commercial training products offered by a variety of companies targeting the pro-gamer, sports and fitness industries, including *Neuron Academy*, *Neurotrainer*, and others (Appelbaum & Erickson, 2018).

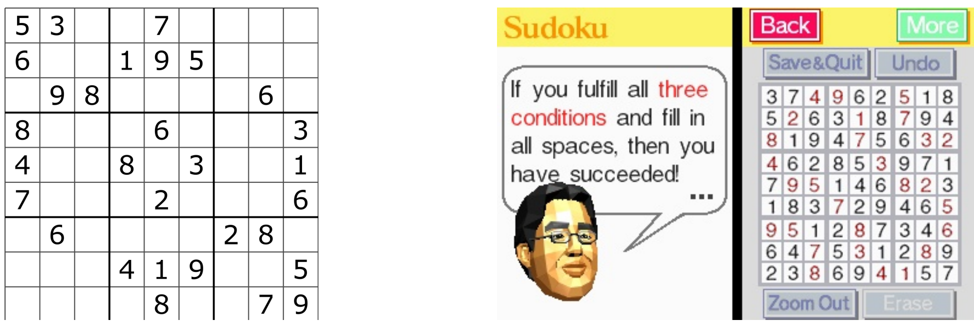
At the time of the first gaming studies in the early 2000s, the action games genre was dominated by simple shooting games that had minimal story lines and whose game objectives rarely went beyond eliminating endless waves of enemies. In contrast, non-action video games such as puzzle or turn-based strategy games did not require decisions to be made under time pressure and focused primarily on resource management and long-term planning. In the intervening two decades, however, over 50 different game genres are now officially recognized ("Category:Video game genres - Wikipedia," retrieved 1/6/22). As Bavelier notes, the action games genre has evolved to incorporate many elements that were formerly associated with non-action genres, including future planning and strategy formulation. The intermingling of genres has in turn increased the complexity of establishing links between specific aspects of gameplay and corresponding cognitive gains (Bavelier & Green, 2019), a challenge for cognitive training research as well.

2.7. Commercial Cognitive Training

Just as action games research findings quickly found their way into commercial products, cognitive training studies have similarly contributed to a dramatic increase in the availability, and public awareness of, commercial "brain training" products. Along the way, an intense legitimacy debate was sparked that continues to this day.

The first commercial screen-based cognitive training app is generally considered to be "Brain Age," released in 2005 for the Nintendo DS game system (Bonnechère, Klass, Langley, & Sahakian, 2021). This initial release included popular logic tasks such as *Sudoku,* but also distractor-avoidance tasks that resembled classic cognitive assessment tasks such as *Stroop* and *Flanker*. Shortly after, in 2012, Posit Science Corporation officially announced the launch of BrainHQ, a commercial cognitive training system ("BrainHQ: Brain Training that Works," 2020). Other companies rapidly followed with competing products and in the following decade the cognitive training industry enjoyed unprecedented growth. Much of this rise was undoubtedly due to the increasingly widespread availability of cognitive training products in the form of game-like, consumer-oriented applications on smartphones, tablets and home entertainment systems. In fact, app-based commercial "brain training" (i.e., regular activities undertaken to maintain or improve one's cognitive abilities) is poised to become a USD 8 billion industry according to some industry experts (MarketsandMarkets Research, 2020). Market leaders Posit Science and Lumosity boast of millions of users and "proven results" ("BrainHQ: Brain Training that Works," 2020; "Lumosity for mobile," 2021).

The cognitive training products produced by these companies, however, often differ significantly in appearance from the paper and pencil versions published in previous decades (Fig 7).



(a) (b)

***Figure 7**: (a) Classic Sudoku task as printed in Dell magazine, ~1979; (b) Brain Age 2 Sudoku, 2005*

No longer dependent on bring printed and distributed on paper, screen-based training apps now include many interactive elements that are usually associated with games, including motivating stories and challenges, colorful buttons, and other game-like features (see previous section: "gamified training"). Since these products now exist for the most part as software titles, they can be installed and accessed from a wide variety of devices. Crucially, they are often designed from the onset for increased convenience and marketed to busy users as products that can be used "on the go":

> *It can fit into your life (even if you are busy): … you can use BrainHQ on almost any computer or mobile device, so you can take it on the go. If you want, you can set up*

personal training goals and have BrainHQ send you training reminders when you want them.

[Brain HQ (https://www.brainhq.com), accessed 9/13/2021].

The rise of commercial training has not been without its share of controversy, however. In part to address overly optimistic claims on the part of commercial makers, an open letter co-authored by scientists at California's Stanford University and the Max Planck research institute was published in 2014. With over 70 co-signers, the letter, entitled "A Consensus on the Brain Training Industry from the Scientific Community" stated unequivocally that "…there is little evidence that playing brain games improves underlying broad cognitive abilities, or that it enables one to better navigate a complex realm of everyday life." ("A Consensus on the Brain Training Industry from the Scientific Community," 2014).

Although a rebuttal later that year ("Brain Training Consensus - Rebuttal," 2014) laid out the case in *favor* of training effectiveness, the United States Federal Trade Commission (FTC), a consumer protection agency, brought lawsuits against several companies in 2015. The charge was that the targeted companies advertised their products using language that suggested academic validity when in fact no consensus as to the effectiveness of cognitive training existed at the time. The offending companies were required to remove claims of efficacy from their marketing materials and pay substantial fines ("Lumosity to Pay $2 Million to Settle FTC Deceptive Advertising Charges for Its 'Brain Training' Program | Federal Trade Commission," 2016).

The debate is unfortunately overshadowed by the fact that a number of prominent research scientists are employed or retained by commercial companies who have a financial incentive to show positive effects for training in general and for specific products in particular. For this and other reasons, the controversy as to whether or not commercial cognitive training can be effective and, if so, under what circumstances, remains ongoing.

2.8. Research Considerations

Based on the prevailing literature, my experimental design and research process was guided by the following considerations:

2.8.1. Using EEG to Measure Task Engagement

My use of EEG technology for the experiments documented in this thesis draws upon the well-documented understanding that the frontal midline theta rhythm can serve as a general predictor of cognitive effort, particularly in tasks involving working memory (see preceding section: "Measuring Cognitive Function with EEG"). In addition, I left open the possibility of unexpected findings by expanding the EEG area of interest across the entire midline, from the frontal to the occipital areas. In line with prior research precedent, collected EEG data was analyzed as four distinct wavelength groups: theta, alpha, low-beta and high-beta. These categories correspond to the frequency ranges 4-8 Hz, 8-13 Hz, 13-20 Hz and 20-28 Hz, as shown in table 1 below.

Table 1: *EEG ranges used in the current research*

Theta	4-8 Hz
Alpha	8-13 Hz
Low Beta	13-20 Hz
High Beta	20-28 Hz

One of the operational challenges of recording EEG data is that the process is extremely sensitive to muscle movements. Unfortunately, whether to actively confirm the position or presence of a stimulus, or to indicate agreement or disagreement with a proposed choice, participation in cognitive assessment and training tasks generally requires physical input from the subject. Such input may take a variety of different forms, including vocal (e.g., verbalizing an answer), tactile (e.g., touching a key or other trigger) and ocular (e.g., directing one's gaze toward a certain agreed area). As all of these response types are likely to produce muscle movements, careful consideration must be given to the experimental design if EEG is to be successfully employed. A portion of my initial research considerations therefore involved first conducting an experiment to see which subject response modalities were the *least* likely to impact the EEG data.

The experiment I designed was centered around the following simple task: an on-screen arrow pointed in the four cardinal directions with the direction changing automatically every 2000 ms. Subjects were instructed to respond when the arrow pointed *down* only. The protocol included two temporally separated control conditions in which the subject was asked to mentally note the response condition but not respond gesturally (no muscle movement). The 4 experimental conditions were as follows: finger tap on lap (relaxed arm), finger tap on keyboard key (slightly tensed arm to reach keyboard placed on table), foot tap on floor trigger device, verbal signal (subject calls out a voiced nonsense syllable "Ge-"). Each condition contained 30 stimulus presentations for a total of 60 seconds per condition, and a total experimental time of 6 minutes. The presentation order of the conditions was randomized for each subject.

EEG electrodes placed in the central and parietal areas captured continuous data during the experiment and time windows of 60 seconds corresponding to each condition were analyzed. The data showed no significant effects for any of the tested conditions, indicating that subject responses incorporating simple muscle movements such as tapping a key on a keyboard or verbalizing a simple answer did not add significant noise to the EEG data stream. The full results can be consulted in appendix A.3.

2.8.2. General Research Environment Considerations

The lack of a consistent training and testing environment is another potentially confounding factor in cognitive research. For example, while the majority of the studies I surveyed in my literature review conducted the entirety of their experiments under lab-like conditions, some allowed subjects to perform tasks in the place and time of their choosing, including at home (Charvet et al., 2017; Lumsden, Skinner, et al., 2016; Mekler et al., 2013). Prior research suggests, however, that environmental variables may play a far larger role in cognitive tasks than previously believed. For instance, Grissmann et al. found working memory load to be sensitive to affective contexts (Grissmann, Faller, Scharinger, Spüler, & Gerjets, 2017), while Mühl et al. documented a negative impact on a working memory task when social stress was introduced (Mühl, Jeunet, & Lotte, 2014). Additionally, Wang et al. found that the amount of ambient light present in the study environment should also be considered. In their 2010 steady state visual evoked potential (SSVEP) study using EEG, the researchers determined that a darkened room had a positive influence on the strength of the alpha rhythm, compared with a brightly lit room (Wang, Qian, Zhuo, & Gao, 2010).

Given the high potential for data contamination due to environmental variables such as audible noise, bright lights or other visual distractions, all experiments described in this thesis were conducted in a quiet, darkened space. The experimental area contained only essential furniture (e.g., desk for the tablet and height-adjustable chair for the subject). Experiments two and three were additionally conducted in a sound-proof, electrically shielded room-within-a-room (aka "shield room"). Subject-researcher interactions such as protocol explanations, informed consent disclosures and task demonstrations, along with essential equipment preparations such as EEG electrode placement, occurred in a separate, lit laboratory environment. After all preparations had been completed, the subject was led into the darkened experiment area and a final check was conducted to ensure all equipment was functioning as expected. The experiment then began with an auditory signal from the researcher.

2.9. Using an HMD to Maximize Operational Control

A relatively recent means of exerting control over the training environment in a wide variety of scenarios is through the use of virtual reality (VR) technology in combination with a head-mounted display (HMD). Such systems, once elaborate, expensive setups requiring a tethered host computer with significant graphics resources, are now mass-produced commercially as game machine add-ons and many models no longer require any additional external processing hardware (Pierce, Young, & Doherty, 2017). Although using an HMD to project a visual environment directly into the user's eyes is currently the most common way to experience VR, the term technically refers to a number of different technologies that can generate such environments. These include specialized televisions with depth display capabilities as well as immersive, multi-planar projection spaces (e.g. CAVE spaces, Fig. 8) that a subject can enter into and walk around in. While on occasion this paper will reference these other technologies, the term *VR* or *virtual space* should be generally understood as involving the use of an HMD.

Figure 8: *CAVE virtual reality environment: synchronized video projectors seamlessly project onto all walls of a room to generate a feeling of spatial immersion.*

Regardless of which modality is being employed, technologies for interacting with virtual environments have become more commonplace in recent years. General interest, as indicated by net searches related to VR, has increased more than three-fold in the past decade ("VR - Explore - Google Trends," 2022). The primary use of these systems, however, is still largely tied to entertainment. As of Winter, 2022, market leader Meta (formerly known as Facebook), which makes the popular Quest and Quest 2 headsets, reports games at the top of both the downloads and sales categories in their eponymous store (Facebook Inc., 2022). Among non-game titles, social platforms and fitness/mindfulness apps dominate the remaining top rankings.

In addition to the individual consumer market, VR is similarly enjoying a robust expansion in a variety of additional fields, including education, e-commerce, professional training and rehabilitation. For example, initial studies looking at the potential role of VR in marketing products and services suggest that the HMD experience accounted for the highest sense of presence and obtained that highest scores for affect, emotions and purchase intentions, compared with standard 2D displays (Martínez-Navarro, Bigné, Guixeres, Alcañiz, & Torrecilla, 2019). These results were corroborated by separate studies that found higher levels of arousal and increased product focus when comparing HMDs to other screen-based presentation technologies (K. Kim, Rosenthal, Zielinski, & Brady, 2014), and even when compared to other types of virtual environments (Juan & Pérez, 2009).

The unique ability of VR to create virtual spaces that nevertheless feel realistic to the user has also made it a natural place to conduct exposure therapy for common phobias including acrophobia (fear of heights) (Boeldt, McMahon, McFaul, & Greenleaf, 2019; Donker et al., 2019) and arachnophobia (fear of spiders) (Miloff, Lindner, Dafgård, Deak, & Garke, 2019; Minns et al., 2019; Rothbaum et al., 1995). Most recently, trauma due to Post Traumatic Stress Disorder (PTSD) has been a subject of intense interest for VR-based intervention as the impracticalities of recreating war zones for the purpose of desensitization make traditional exposure therapy unworkable. Early reports are encouraging. A recent review of nine controlled studies found that VR-based PTSD exposure treatment was just as effective as traditional psychiatric interventions, and significantly more effective than passive controls

(Kothgassner et al., 2019). In another comprehensive literature survey, Deng and his team additionally concluded that gains from VR treatment remained significant at up to 12 months following the intervention and even led to moderately positive effects on related measures such as general depression (Deng et al., 2019). These and similar results led Dr. Raggi to sum up the situation with an essay in "Reviews in the Neurosciences" that declared that VR could soon become a "decisive tool" for neurorehabilitation (Raggi, Tasca, & Ferri, 2017).

Comparison studies examining the use of VR in cognitive tasks are still few, but results are similarly encouraging. A recent comprehensive review of computerized cognitive training studies determined that the largest effect sizes were found in studies that used VR, prompting the authors to speculate that virtual environments might be more stimulating and engaging than traditional computerized cognitive training (CCT) (Hill et al., 2017). This observation was reiterated in a subsequent literature review by Garcia-Betances in 2019: "VR has the capacity to significantly enhance the way computer-based cognitive healthcare interventions are delivered and, thus, the ability to uncover new possibilities for improving their effectiveness" (García-Betances et al., 2017, p.2).

In general, researchers studying the effects of virtual environments report increased engagement and enjoyment when performing a task in VR space, compared to a non-VR condition (Lhemedu-Steinke, Meixner, & Weber, 2018; Martínez-Navarro et al., 2019; McMahan, Parberry, & Parsons, 2015). Studies specifically using biometric outcomes to differentiate VR and non-VR also exist. For instance, in 2015 Slobounov and his team used qualitative and quantitative (EEG) measures to evaluate subject performance with a navigation task. Their results concluded that subjects experienced higher levels of excitement, accompanied by increased theta activity, when navigating a virtual environment. It must be noted, however, that their use of virtual reality was in the form of a CAVE environment, which relied on a large number of precisely positioned video projectors to create the illusion of 3D space. CAVE environments are unique in that they permit free, untethered movement and do not require the use of additional hardware on the part of the user. However, they remain expensive and impractical outside of research environments due to their complex installation and maintenance requirements.

With the introduction of any new technology, there are often a variety of hurdles to overcome before widespread adoption is possible. For example, one common misconception surrounding the use of virtual reality headsets is the assumption that the technology is more physically stressful to use and may be simply rejected outright by older individuals. To address these concerns, a major initiative was undertaken in 2016 by a team from the Nicolaus Copernicus University in Bydgoszcz, Poland to assess the general adoption of new technologies, including HMDs, by geriatric patients. The researchers found that while some patients were initially anxious, once the purpose of the technology was adequately explained they quickly gained familiarity and reported positive emotions in the final assessment. In fact, 83% of the elderly patients involved in the study voluntarily elected to continue rehabilitation using VR (Kujawska et al., 2016). Subsequent studies found similarly high acceptance and satisfaction rates among older patients. For example, Huygelier and her colleagues found that attitudes regarding HMD use went from neutral to positive after exposure to an interactive environment (Huygelier, Schraepen, van Ee, Vanden Abeele, & Gillebert, 2019) and a 2020 study involving seniors with mild cognitive impairment found high rates of satisfaction when the subjects were tasked with collecting ingredients to complete a set of cooking recipes within a VR environment (Yun et al., 2020). Finally, the misconception that HMDs are more physically demanding to use was also found to be unsupported by data. In fact, Guo and his colleagues found the opposite to be true in their 2017 paper investigating visual fatigue in virtual environments, determining that VR is *less* visually stressful than similar content presented on a typical smartphone screen (Guo, Weng, Been-Lirn Duh, Liu, & Wang, 2017).

Nevertheless, real challenges remain. As of 2022, the most popular headsets are still relatively heavy and provide only a limited field of view. Additionally, a still-not-fully-understood phenomenon dubbed "cybersickness" (LaViola, 2000) has the potential to negatively impact research results. For instance, a 2019 study looking at the therapeutic benefits of regular exercise sought to compare a walk in a nature park with an interactive virtual walk (using an HMD and a treadmill) in a simulated digital version of the same park. The researchers reported that a large number of subjects reacted poorly to the HMD, including some who became nauseous, due to a perceived time-lag between their actual movements on the treadmill and the corresponding display of the virtual scenery. It was ultimately recommended that HMDs only be employed in scenarios that do not require subject movement. Although subsequent studies have suggested ways to minimize these effects, for instance through the use of haptics (Peng et al., 2020), concerns remain.

Finally, although most studies designed around the therapeutic use of VR tend to be about what can be inserted *into* a virtual environment, it is also possible to use the technology to exclude or deemphasize detail. In the interests of redirecting focus to a desired stimulus or stimuli, for example, all unrelated imagery in the view field might be intentionally suppressed. I feel it is this latter scenario that is particularly well suited to cognitive training as it effectively allows the subject's perceptual field to be restricted to just the task itself.

Whether adding or subtracting content, however, prior experiments such as those referenced earlier in this section demonstrate without a doubt that the use of VR has the potential to significantly impact the way we conduct cognitive studies in the future. In light of this, there is clearly a need for more comparative studies that examine the effects of VR exposure in a more general capacity, as well as their impact on specific cognitive tasks.

2.10. Experimental Task Considerations

One oft-cited reason for the success or failure of training studies involves the presence or absence of certain features in the underlying training task itself. For example, as Dr. Edwards noted in her survey of studies that made use of the Useful Field of View training task, the following three elements were deemed critical to success: time pressure, divided attention and an adaptive task design (Edwards et al., 2018). These elements, which coincidentally have also been cited as crucial to success in action games studies (Cardoso-Leite, Joessel, & Bavelier, 2020), are now generally accepted among the "best practices" for cognitive research (Hudak et al., 2019; Kalyuga, 2009). As they consequently also formed the basis for the original experimental task I developed for my own research, I would like to address them each in turn:

2.10.1. Time Pressure

Time pressure (TP) can be defined as the difference between the amount of available time and the amount of time required to solve a task. Time pressure has been shown to have a significant effect on cognitive task performance and is a frequently used adaptation to modulate task difficulty (Edwards et al., 2018). Experimental data suggesting that TP results in faster response rates but an increased rate of errors dates back over a half century (Fitts, 1954). This "speed-accuracy trade-off" has since been validated more recently in a 2000 paper on the neurophysiological and behavioral indices of time pressure on visuomotor task performance (S. M. Slobounov et al., 2000). In addition to confirming Fitts' earlier findings, Slobounov and his colleagues also found that the performance gains associated with time pressure remain even when the time pressure component is subsequently removed. Furthermore, significant increases in frontal theta power were found to accompany the presence of TP. Since primary executive functions, particularly working memory, are associated with similar patterns of frontal midline (FM) theta activity, a more intimate relationship between TP and other cognitive functions has also been suggested (Edwards et al., 2018).

Interestingly, the relationship between time pressure and cognitive task performance has been shown to be sensitive to other cognitive processes as well. For example, a 1993 study showed that the impact of time pressure can be affected by subject motivation. The researchers demonstrated that while task performance was initially boosted by the application of time pressure, performance nevertheless dropped dramatically once the task was perceived by the subject as no longer worth completing (Rastegary & Landy, 1993). This phenomenon has been observed by other researchers as well, including Brehm and Wright, whose 1989 proposal for a *Motivational Intensity Model* (MIM) of cognition hypothesizes a similar interdependence between motivation and effort (see section 2.10.3 for a more complete description of the MIM).

For the research presented in this thesis, time pressure was carefully applied in an adaptive manner during the final two experiments in order to:

(1) maximize subject effort

(2) as a means of adaptively pacing the presentation of stimuli according to the current subject's ability

Additionally, care was taken to avoid any dramatic performance drops due to sudden changes in subject motivation or cognitive exhaustion (see chapters 3, 4 for specific implementation details.

2.10.2. Divided Attention

Divided attention can perhaps be best understood by thinking about the modern notion of *multi-tasking*. In divided attention tasks, the subject is required to perform two or more sub-tasks at the same time, with overall success requiring sufficient allocation of attentional resources for *all* subtasks. This may take the form of several simultaneously presented stimuli (e.g., the dual n-back task), counting the number of times a stimulus parameter (e.g., background color) changes, or indeed tasks that employ different modes of concurrent cognition, such as tracking the spatial location of one visual stimulus while simultaneously selecting a previously viewed stimulus from among distractors (e.g., the UFoV task). Divided attention is differentiated from *selective* attention, in which sub-tasks are addressed not simultaneously but sequentially.

fMRI studies have shown substantially increased levels of BOLD signal intensity in cortical activations during divided attention activities, compared with sequential or fixed attention tasks (Beauchamp, Cox, & Deyoe, 1997; Greenlee, 2000). Notably, fMRI results show that divided and selective attention modes differ only in intensity and *not* area of activation. In other words, the reason divided attention tasks appear to contribute to the effectiveness of cognitive training seems to be linked with the increased cognitive stimulation offered by such tasks.

2.10.3. Adaptive Task

In a 2018 follow-up analysis to the original ACTIVE study, co-principle researcher J.D. Edwards published a systematic review of all known research involving Useful Field of View training, the core task used in the ACTIVE study. The goal was ostensibly to see if other researchers had obtained similarly encouraging results when using this task. Among her many conclusions, Edwards suggested that a key element in the success of UFoV training, and by extension cognitive training in general, was the *adaptive* nature of the task (Edwards et al., 2018). The importance of task adaptivity in successful training has since also been confirmed by other researchers, including Kalyuga and his colleagues, who concluded that optimal cognitive efficiency can only be achieved when the difficulty level fits the user's needs and capabilities (Kalyuga, 2009). Most recently, a 2019 study targeting retirement communities had a particularly blunt assessment of the importance of task adaptation: "*Adaptive cognitive training had significant results on memory outcomes; non-adaptive did not…*" (Hudak et al., 2019).

In light of these unambiguous conclusions, developing a task adaptation strategy to better match a subject's native ability and adjust to training gains was high among the requirements for my own study task. In order to adapt a task to a user's natural ability, however, I first needed a method of determining the range of a user's performance, particularly the upper range of maximum cognitive effort. This optimal range is sometimes referred to as "the zone" (Ryan, 2016).

One proposal for determining the upper limit of potential subject performance is found in Brehm and Wright's theory of *Motivational Intensity* (Brehm & Self, 1989; Wright, 2008). The theory states simply that as cognitive demand rises, cognitive effort rises proportionally to match the demand until a physical limit is reached. This point is referred to by Brehm and Wright as the *exhaustion* or *fatigue* point and leads to an immediate drop in resource allocation for the task. The process is summarized in figure 9.

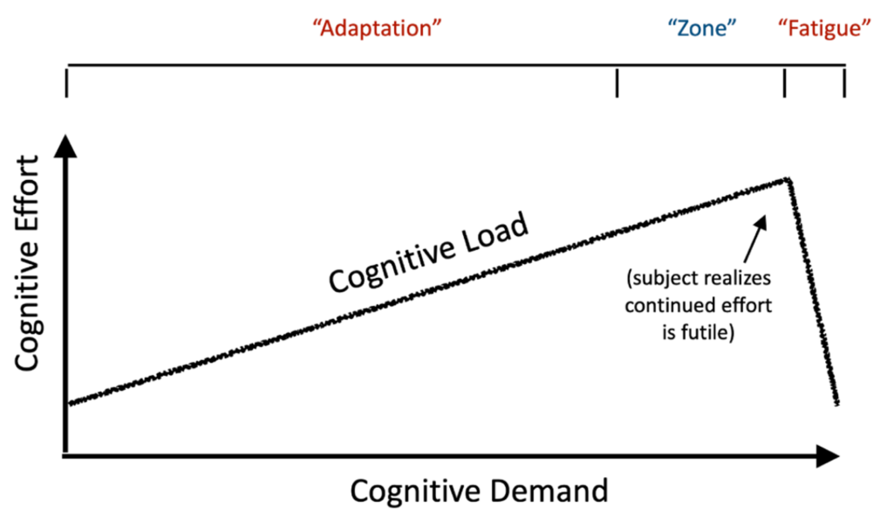


Figure 9: *Summary of Wright and Brehm's Motivational Intensity Model. Cognitive effort rises to meet demand in the adaptation phase. The moment of fatigue, and its accompanying cognitive collapse, is preceded by a brief period of optimal effort, referred to by some as "the zone."*

Although the motivational intensity model (MIM) has been independently validated, I undertook my own experiment to see if the model would enable me to identify the period of maximum task engagement for my own experimental task. Using a visuo-spatial task that had been modified to increase in difficulty with each round, subjects performed the task until they were no longer able to keep up (cognitive fatigue). As predicted by the theory, the telltale drop-off in effort (as measured using EEG) appeared approximately 5 seconds before the end of the task (see appendix A.2 for details and sample data). Ultimately, the predictions of the *Motivational Intensity Model* guided both the EEG assessment of my first experiment, and the task adaptation strategy for the second and third experiments.

2.11. Chapter Summary

In this chapter, I began with an analysis of the current state of cognitive training research and noted that the primary objective of cognitive training is generally understood to be *the prevention or regression of age or illness-related cognitive decline*. I observed that to achieve this goal, training interventions typically target cognitive processes falling within the domain of executive functions, including memory, decision making and problem solving. I then discussed methods for assessing the current state of cognitive function, including biometric technologies such as EEGs as well as standard measures of intelligence and general function such as the MMSE and IADL scales.

This was followed by an in-depth examination of both traditional cognitive training tasks and their corresponding commercial adaptations. I observed that the most salient difference between the two is that the latter group tends to add features resulting in the core task resembling a video game. Pursuant to this observation, relevant studies related to action games research as well as gamified cognitive tasks were examined, including the growing body of virtual reality based cognitive research. Finally, the principles behind the training tasks used in successful studies, including time pressure and adaptive design patterns, were qualified and discussed.

The next chapter begins the documentation of a series of empirical experiments carried out both to validate core principles of the cognitive research discussed in this chapter, as well as to fill in knowledge gaps identified during the course of my research survey.

3. COMPARING BRAIN ACTIVITY IN VIRTUAL AND NON-VIRTUAL ENVIRONMENTS

(Portions of this chapter are based on my conference presentations at IEE 2020 and IMEKO 2021, as well as my publication in Measurement: Sensors, 2021 (Redlinger & Shao, 2021; Redlinger, Shao, & Yagi, 2019))

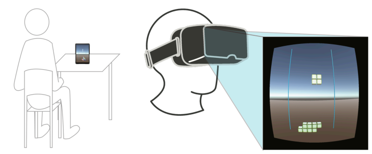
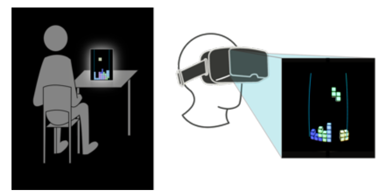
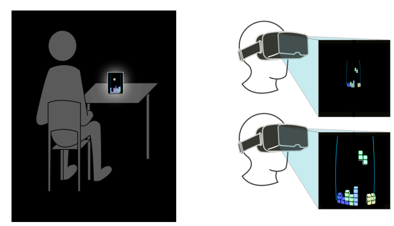
3.1. Chapter Aims

This chapter looks at the feasibility of using a VR head-mounted display together with EEG in light of recent technological advances including lightweight, wireless EEG amplifiers and HMD units. Cognitive research prior to 2018 has largely avoided this combination due to size, movement and comfort constraints with older hardware. I therefore hoped to answer the following questions:

1. Is using an HMD a viable option when conducting EEG experiments?
2. Is there any difference in cognitive activity when performing a standard cognitive task in a VR environment, compared with a non-VR environment?

The chapter also describes the process of gradually refining my experimental protocol. Over the course of the study, three iterations in total were conducted in order to address potentially confounding factors and explore additional research directions. In each case, the experimental task, EEG analysis approach and general experiment design are identical. Except where otherwise noted, study results and discussion are derived from iteration 2 of the protocol, which is considered definitive and the basis for the conference presentation and journal publication listed in the chapter heading. The primary distinctions between protocol iterations are detailed in table 2:

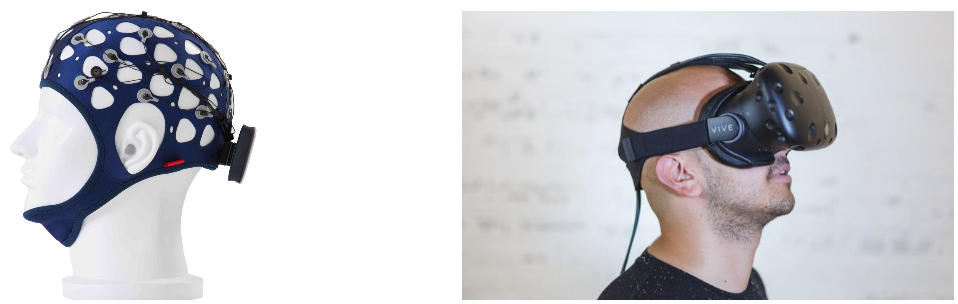
Table 2: Differences between protocol iterations

Iteration 1: two experimental conditions: tablet and HMD study venue = lit room stimulus visual angle: 15°	
Iteration 2: two experimental conditions: tablet and HMD *study venue = darkened room* stimulus visual angle: 15°	
Iteration 3: *three experimental conditions: tablet, small VR, large VR* *study venue = darkened room* *stimulus visual angle: 10° and 15°*	

3.2. Introduction

Recently, technologies for interacting with virtual environments have become more popular and commonplace. Virtual reality (VR) systems, once elaborate, expensive setups requiring a tethered host computer with significant graphics resources, are now produced commercially as game machine add-ons and some models no longer require a host computer (Pierce et al., 2017). Direct comparison between virtual and non-virtual spaces is inherently difficult due to technological challenges that have only recently been ameliorated with improvements to HMD technology. For example, any body movements that occur while recording EEG data has the potential to introduce motion artifacts which must then be laboriously removed using factor analysis or another means. In some cases, it is impossible to acceptably remove movement noise from the signal.

Traditional EEG systems use surface electrodes that are integrated into a skullcap that fits over the head. The electrodes are connected to an amplifier, which in turn is connected to a computer for recording and analysis. Such systems require elaborate preparation and precise adjustment to each electrode to optimize the SNR (signal to noise ratio). Once placed and calibrated, moving the skullcap has a high risk of destabilizing the electrode connections. At the same time, VR headsets themselves must be adjusted on the scalp in order to insure a good fit and optimal positioning of the optical components (Fig. 10).



(a) (b)

Figure 10*: (a) a standard EEG skull cap; (b) a standard VR headset (HTC Vive Pro); (Images courtesy of https://wearablesensing.com and https://stocksnap.io)*

Given these challenges, the majority of studies that heretofore have sought to quantify VR/non-VR differences have made use of non-HMD virtual environments such as 3D televisions (Kober, Kurzmann, & Neuper, 2012) or projection-based CAVE environments (Semyon M. Slobounov, Ray, Johnson, Slobounov, & Newell, 2015). Others have relied only on qualitative outcomes (Lhemedu-Steinke et al., 2018) or focused on psychophysiological metrics such as cardiovascular activity, rather than EEG (Richter, Friedrich, & Gendolla, 2008).

This study took advantage of recent advancements in both HMD technology and portable, compact EEG acquisition systems to quantitatively compare the neural activity in virtual and non-virtual environments. Based on prior research results, my experimental hypothesis was that brain activity related to executive function should be similar or be slightly higher in the VR condition due to the reduced impact of environmental distractions and the relative novelty of the virtual environment itself.

3.3. Materials and Methods

3.3.1. Experimental Task

The experimental task was a rotating-blocks puzzle game based on Tetris, a game frequently used in cognitive research to assess cognitive demand as it simultaneously engages attention, memory and visual spatial functions (Ewing et al., 2016; Franks & Boutcher, 2003; Haier, Karama, Leyba, & Jung, 2009; Maglio, Wenger, & Copeland, 2008; Patsis, Sahli, Verhelst, & De Troyer, 2013). The game requires subjects to rotate and move falling blocks in such a way that they form complete rows, which are then removed automatically by the software. The version implemented in my study was limited only to the underlying cognitive task and did not include any motivational game-like elements such as score or progress displays (Fig. 11).

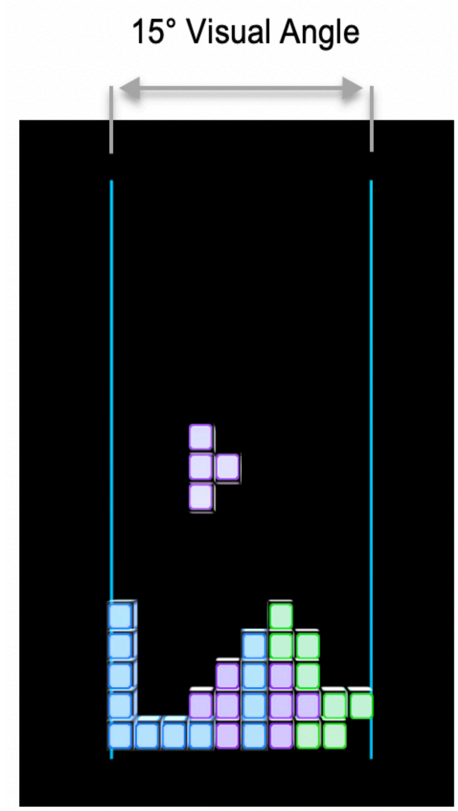


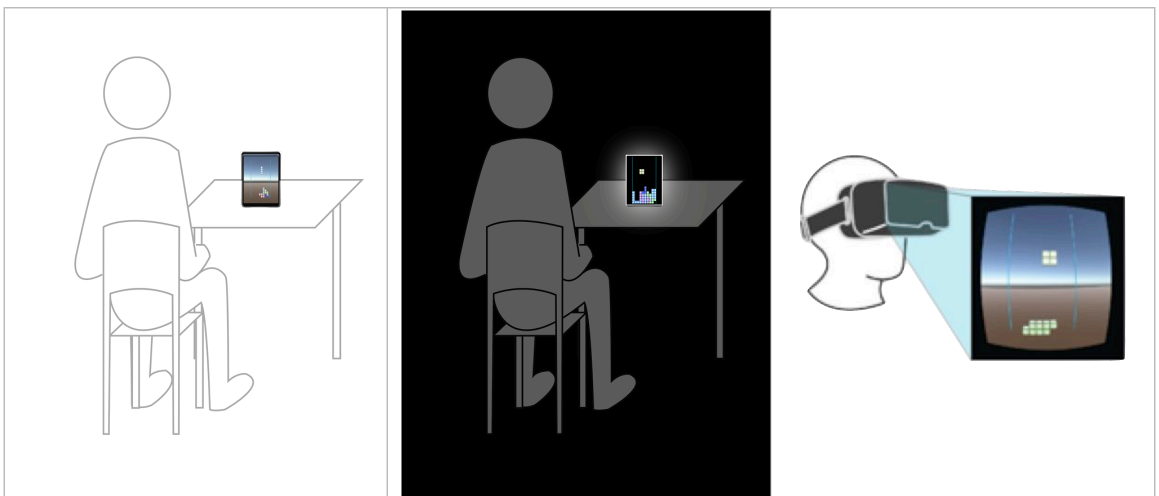
***Figure 11**: Rotating blocks puzzle game. Task visual angle (horizontal axis) was set at 15°.*

Initial task difficulty (i.e. speed at which blocks descended) was set at a level that permitted even inexperienced participants to successfully complete rows. Based on the results of prior motivation studies, in particular Wright's research on motivational intensity (Wright, 2008), I chose to have each completed row lead to an increase in speed until participants were no longer able to keep up. The task ended when the blocks reached the top of the screen.

3.3.2. Test Environment

A standalone, commercial HMD (VIVE Focus, HTC Corp.) was used for the VR condition, and a 10.2-inch tablet computer (iPad, Apple, Inc.) for the non-VR environment. The custom rotating blocks game was created using a cross-platform coding environment (Unity, https://unity3d.com), and subsequently exported for both environments. Care was taken to ensure that both versions performed identically despite the difference in modality (HMD vs tablet) and that the task was presented at the same visual angle. For the non-VR condition in iteration one, a bright, clean, uncluttered environment was used. For iterations two and three, a darkened, sound-proof room with electromagnetic shielding was prepared.

In both cases, the tablet device was positioned just below eye-level, approximately 70 cm from the subject, to insure a 15° visual angle for the experimental task. VR content, as viewed in the HMD, was similarly positioned in virtual space so that it appeared to the subject to be the same size as in the tablet view (identical visual angle). (Fig. 12)



(a)	(b)	(c)

Figure 12: *(a) non-VR condition, iteration one; (b) non-VR condition, iterations two, three; (c) VR condition, all protocol iterations*

While HMD systems typically rely on virtual pointers for user input, such input devices are not appropriate for EEG studies as they have a high risk of introducing muscle-related artifacts. To address this, a touchscreen smartphone was programmed to send network commands to the test environment. A soft, foam overlay approximately 3mm thick with openings corresponding to the locations of the on-screen controls was added so that the subject could identify the controls in a tactile manner without the need to view its screen (Fig. 13).

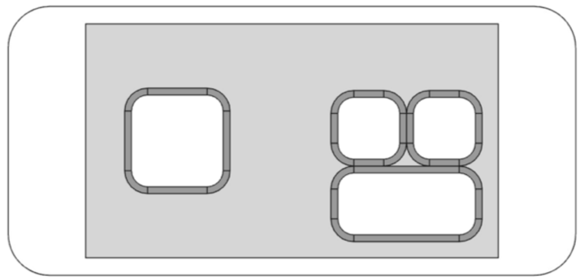


Figure 13: Foam overlay with cut-outs corresponding to "button" locations on a touchscreen smartphone.

3.3.3. Experimental Design

The subject pool comprised twelve healthy individuals, aged 21–48 years (8 males, 4 females) with no history of color vision disorders. All but one subject had previous experience using an HMD but none personally owned a device nor reported any regular weekly use.

Protocol iterations 1 and 2 comprised four phases: non-VR control, non-VR experimental, VR control and VR experimental. Iteration 3 added an additional fifth VR experimental condition containing a modified stimulus (see section 3.3.4 Protocol Evolution). During the control phases, subjects viewed a black background with open eyes for 60 seconds. EEG signals recorded during these phases served as a baseline against which to compare brain activity during the experimental task phases. Every subject performed all conditions, but the order of presentation was randomized and balanced so that half of the subjects began with the non-VR condition and the remaining half with the VR condition. Subjects enjoyed a 5-minute break between modalities during which the VR headset was either removed (VR first group) or installed (non-VR first group).

In the experimental phases, subjects performed the task until they could no longer complete sufficient rows to prevent the falling blocks from reaching the top of the screen. The duration varied according to subjects' ability but averaged approximately 2 minutes.

The use of two separate control phases was to eliminate any bias related to wearing the HMD itself. By comparing each experimental condition's results only to the control data obtained from the same modality, any observed differences could only have resulted from the dependent variable regardless of whether the task was conducted in virtual or non-virtual space.

3.3.4. Protocol Evolution

The goal of this experiment was to compare brain activity in virtual and non-virtual environments. In order to achieve this goal, it was necessary to minimize as much as possible any environmental variables that might contribute to differing results between the two conditions. Unfortunately, a potentially confounding element became apparent only after the conclusion of the first iteration of the experiment when the results revealed an unexpectedly large effect size for the dependent variable. As the non-VR condition was initially performed in a lit room, the presence of ambient light was suspected of having potentially impacted the data since prior research was discovered suggesting the bright spaces can impact cognitive activity (e.g., Wang et al., 2010). It was therefore deemed prudent to redo the experiment in a dark room to correct for this variable. The results of the follow-up iteration were remarkably similar, suggesting that ambient light might not have played a significant role after all. At this point the findings from the second protocol iteration were deemed conclusive and were subsequently compiled, analyzed and published (Redlinger & Shao, 2021).

Continued speculation about the reasons for the large observed difference led to a final iteration of the experiment when additional research came to light suggesting that depth perception might be underestimated in virtual environments (Maruhn, Schneider, & Bengler, 2019; Rousset, Bourdin, Goulon, Monnoyer, & Vercher, 2018). Given this possibility, the subjects could potentially have perceived the experimental stimulus as being larger than intended, with further implications for cognitive load. To test this idea, an additional third protocol iteration was proposed and executed. A comparison of overall results from all three protocol iterations are included in appendix A.6. Data from the final iteration showed that stimulus size does indeed impact EEG power, although probably not substantially enough to explain the observed effect sizes. A more thorough investigation of the stimulus size effect became the basis of my subsequent experiment: Impact of Screen Size on Cognitive Training Task Performance (see chapter 4).

3.3.5. EEG

EEG signals were acquired from the frontal, central, occipital and parietal regions, using a portable, wireless 8 channel EEG amplifier (OpenBCI 32-bit Board Kit, OpenBCI, inc.) with a sampling rate of 250 Hz. The electrode locations were Fz, Cz, Oz and Pz, placed according to the international 10-20 system, and specifically selected in order to capture a broad range of activity along the midline. Gold cup electrodes were attached to the scalp and ear lobes using electro-conductive gel, and an initial impedance of <5 kΩ across all electrode positions was insured. Additional electrodes were affixed above and below the subjects' eyes to record electrooculogram (EOG) signals caused by blinking or other facial movements for later use in noise reduction and signal optimization (Zahan, 2017).

EEG data were recorded throughout the experiment, although only the 30 seconds of activity prior to task conclusion were analyzed for each participant. In addition, since the *Motivational Intensity Model* predicts a collapse in cognitive load as the subject anticipates the end of the task, I chose to inset the analysis epoch by 5 seconds to capture the period of maximum load for each subject. The 5 second value was based on analyses of participants' EEG results while performing a similar task in a prior pilot study. Through the use of this mechanism, I hoped to measure similar levels of engagement across all participants. (Fig. 14)

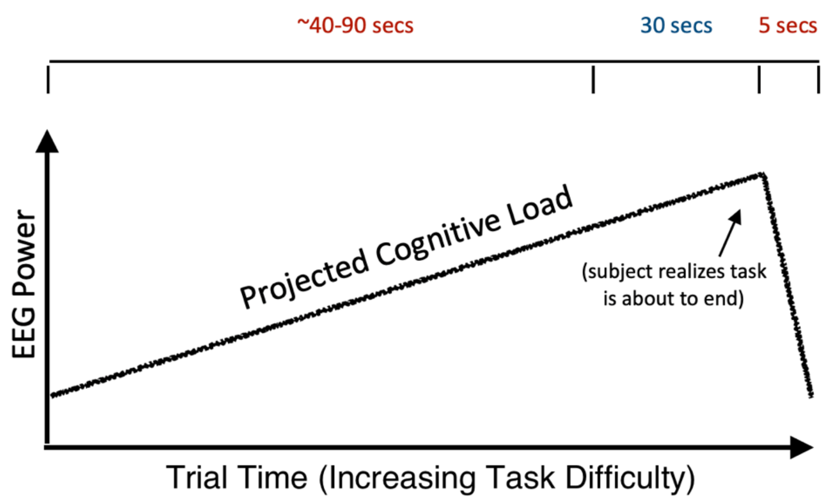


Figure 14: *Determination of analysis window: based on the Motivational Intensity Model, a 30 second window that ends just before the trial finishes should insure maximum cognitive engagement for all subjects. Time region to be analyzed printed in blue ink; discarded regions printed in red ink.*

The decision to focus on just four electrodes along the midline, instead of a more comprehensive, multi-channel EEG system, was made for two primary reasons. First, I wanted to use high-quality, gold-plated electrodes to maximize the sensitivity to potentially slight changes in cognitive activity. Such electrodes are usually applied manually and not generally compatible with standard EEG skullcaps. Second, integrating traditional EEG caps with the straps and harnesses of the HMD in a way that did not further add to the discomfort many users already associate with headsets seemed an unnecessary risk to the study's integrity. Compared to traditional EEG setups, this approach enabled me to limit potential physical or electrical interference between the EEG system and the VR headset.

3.3.6. Analysis Methods

The raw EEG data was notch filtered (50Hz) and high-pass filtered from 4Hz. Muscle artifacts identified from EOG data and other apparent artifacts identified by visual inspection of EEG data plots were excluded from analysis.

Fast Fourier transforms (FFT) were calculated for the following spectral ranges: theta (4-8Hz), alpha (8-13Hz), low beta (13-20Hz) and high beta (20-28Hz) with 30 second windows for each phase of the experiment. The total sum of the power values from each range was divided by the total number of data points. The resulting score was normalized by subtracting the overall population mean and dividing by the standard deviation to obtain a power index. Finally, Wilcoxon signed-rank tests were conducted to determine the significance of any change in power between the controls and their corresponding experimental phases.

3.4. Results

Following feature extraction and normalization, significant spectral power increases for the beta and theta ranges were observed during the experimental phases across all subjects tested. Alpha ranges, in contrast, revealed decreases or no meaningful changes in spectral power between control and experimental states. Overall power during the experimental task condition was strongest in the theta and low beta ranges at position Fz. Results at other electrode locations along the midline mostly mirrored the trends seen at Fz, but generally exhibited higher standard error and were not included in the study results (figure 15).

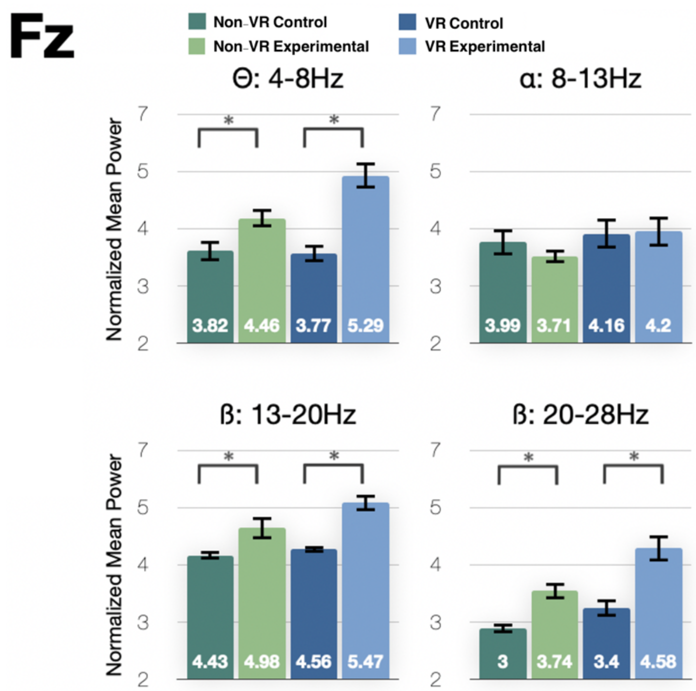


***Figure 15**: Spectral power by experimental condition (Non-VR, VR) and frequency at electrode location Fz; n=12; SE; significance calculated with Wilcoxon rank-sum (* = p < .05).*

The most dramatic power increase that was also significant (Wilcoxon rank-sum; p=.037) was observed in the frontal midline theta rhythm (4-8Hz) as measured at Fz. The average increase of nearly 41% for the VR condition was substantially higher than the modest 17% increase for the non-VR tablet condition. Similar but less dramatic increases can be observed in the low and high beta ranges at Fz, with 20% and 34% increases, respectively (Table 3).

Table 3*: Mean power increase from control by frequency range and condition at Fz; (n=12).*

	Non-VR	VR
Θ (4-8Hz)	16.8%	40.6%
α (8-13Hz)	-7.2%	0.9%
ß (low)	12.2%	20.1%
ß (high)	24.7%	34.4%

As previously mentioned, inexplicably large differences in overall spectral power were observed between the non-VR and VR conditions. An additional experimental condition was therefore added to determine whether differing stimulus sizes might help explain the results.

3.5. Discussion

Observed EEG spectral power during the control phases of both the VR and non-VR conditions was very similar, indicating that the headset itself did not seem to amplify or suppress cognitive activity when no stimulus was present. This is crucial as there was some concern that the innate weight disparity or discomfort that some HMD users report might prove to be a confounding variable in this study.

Interpreting EEG data along the midline, particularly in the frontal area, to determine user engagement and concentration is well-documented and the primary reason I chose to focus on this area. For example, Gevins and Smith showed that a condition of increasing frontal theta and decreasing alpha was highly correlated with mental engagement on certain tasks (Alan Gevins et al., 1998), whereas Yamada found that theta activity in the frontal midline area alone was sufficient to gauge concentration levels in children, when combined with other features such as eye-blinking (Yamada, 1998).

As predicted, my results showed that theta power increased, and alpha power remained flat or decreased slightly, when the subjects were engaged with the experimental task. These results were in line with prior research cited above. Furthermore, this increase was much greater for the VR condition than the non-VR tablet condition, possibly reflecting higher cognitive engagement when the task was performed using an HMD. This effect was also previously documented in studies by Slobounov, Kober and others (Kober et al., 2012; S. M. Slobounov et al., 2000) who similarly found higher levels of task engagement in VR environments. While I had originally hypothesized increased cognitive activity during the VR condition, I was surprised by the degree of the difference. The nearly 41% increase in frontal theta power for the HMD was more than double that of the tablet. While I do not yet fully understand the reasons for this, I would like to offer several theories.

One of the characteristic features of head-mounted displays is their ability to impart on the user a feeling of detachment from the outside world. This increased sense of isolation may have resulted in improved concentration. While not a part of the current study, a follow-up questionnaire specifically asking subjects about their subjective experiences with using an HMD compared to the tablet might have helped to shed light on this possibility and is highly recommended for any future comparison studies.

In addition, it has been well documented that stimulus size can play a role in modulating the cognitive response to a wide variety of tasks. For example, Chen and McManus found a substantial positive correlation between the perceived stimulus size and the resulting EEG power in their 2019 SSVEP study (Chen, McManus, Valsecchi, Harris, & Gegenfurtner, 2019). Recent e-learning studies have also demonstrated a positive effect related to larger screen sizes on students' ability to recall information immediately after a pre-recorded lecture (Al Ghamdi et al., 2016; Park, Han, Kim, Cho, & Del Pobil, 2018) and longer-term retention of learned foreign language vocabulary (D. Kim & Kim, 2012). Finally, Zheng et al. found larger stimulus sizes corresponded to improved signal to noise ratios in their 2013 EEG study (Zheng et al., 2013). In light of this, I took great care to insure a similar visual angle between the two experimental conditions to reduce this effect as much as possible. However, it remains a possibility that while technically identical, the size of the stimulus in the VR environment may have been *perceived* as greater due to the recently reported phenomenon of distorted depth perception in virtual environments. Indeed, Rousset found that "observers immersed in virtual environments perceive virtual space as compressed relative to the real world, resulting in systematic underestimations of egocentric distance" (Rousset et al., 2018). Although a later study by Maruhn and his colleagues found that these effects can be mitigated with additional subject preparation and training (Maruhn et al., 2019), I feel the phenomenon merits further investigation.

3.6. Conclusion

The results of this experiment unambiguously demonstrate the viability of implementing EEG recording capabilities simultaneously while using an HMD. Furthermore, the findings of the current study demonstrate an increased cognitive response in the frontal midline theta rhythm when performing a complex cognitive task while wearing an HMD. This in turn may indicate a generalized increase in user engagement in virtual environments. If validated, this finding would have significant implications for those currently pursuing commercial and e-learning opportunities in VR.

4. IMPACT OF SCREEN SIZE ON COGNITIVE TRAINING TASK PERFORMANCE

(This chapter is based on my publication in Int. J. Psychophysiol. (Redlinger, Glas, & Rong, 2021))

4.1. Chapter Aims

This chapter documents my attempts to reconcile the chief findings of the previous experiment and my overall research goals by focusing exclusively on the potential impact of screen size. Therefore, my principle aims were to:

1. Examine whether today's smaller screens might be impacting our basic cognitive processes.
2. Compared with the previous experiment, further limit any potentially confounding environment variables by moving the entirety of the experiment into VR space where stimulus size, brightness and background effects can be precisely controlled.
3. Develop a novel cognitive task that is easy to learn, incorporates the most promising features from previous successful studies, and can be readily adapted to match each subject's native abilities.
4. Monitor task performance (reaction time and accuracy) outcomes in addition to cognitive activity.

4.2. Introduction

In the previous chapter, I documented an increase in cognitive activity when performing a visuo-spatial task in virtual space compared with the same task in non-virtual space. The large effect size for the VR condition was unexpected and the exact reasons for the outcome discrepancy is still unknown. In the subsequent analysis, several limitations of the experiment were noted, including issues related to potential background disturbances in the non-VR condition, a tendency for subjects to overestimate distances in virtual space, the novelty of the VR environment itself and prior familiarity with the task (based on the classic game Tetris) for some subjects.

In order to address these limitations, I designed my follow-up experiment to be conducted exclusively in virtual space, which eliminated any confounding variables related to moving between environments. In addition, since the previous experimental task was derived from a popular video game, the possibility existed that some subjects were already familiar with the task mechanism, potentially impacting the experimental effect.

In order to address this, a new task was needed. It was decided that the new task needed to meet the following criteria:

1. Be novel, but easy to learn.
2. Include the primary elements that previous studies had recommended as "best practices" for cognitive training.
3. Be easily scalable to different screen sizes and visual modalities.

A survey of prior research into the impact of screen size was undertaken, which revealed a general consensus that perceived stimulus size and cognitive activity are positively correlated. For example, Chen and McManus' steady state visually evoked potential (SSVEP) study found increased SSVEP power when the stimulus size was increased (Chen et al., 2019). Recent e-learning studies have similarly demonstrated a positive effect on information recall related to larger screen sizes (Al Ghamdi et al., 2016; Park et al., 2018). Surprisingly, even though commercial cognitive training is trending towards smaller, not larger, displays in order to better accommodate personal digital devices, to the best of my knowledge no research study focusing specifically on cognitive training performance as a function of screen size has yet been undertaken.

Additionally, recent research has shown that environmental variables may play a far larger role in cognitive training than previously believed. For instance, Grissmann et al. found working memory load to be sensitive to affective contexts (Grissmann et al., 2017), while Mühl et al. found a similar impact on a working memory task when social stress was introduced (Mühl et al., 2014).

Using an HMD for cognitive training addresses both of these issues. Furthermore, maintaining optimal levels for display brightness and stimulus size is more easily accomplished in an HMD environment compared with traditional displays. Taken together, these advantages indicate the use of HMDs for visual cognitive training tasks, a conclusion that has also been reached by other researchers (García-Betances et al., 2017).

4.3. Materials and Methods

4.3.1. Study Design & Sample Size Considerations

This study examined the impact of screen size with two primary outcomes: cognitive load as measured using EEG, and raw task performance as determined by accuracy and reaction time. The experimental task was a modified n-back working memory task. In the classic n-back task, a subject is presented with a continuous stream of similar stimuli. The subject must then indicate when the current stimulus matches the one from *n* steps earlier in the sequence. For this study, I wrote software that generates an n-back working memory task for an HMD environment rather than a traditional monitor screen. Utilizing an HMD serves two purposes: 1) to precisely control the display brightness as well as the task visual angle across subjects and experimental conditions, and 2) to suppress external stimuli to minimize environmental effects.

A preliminary feasibility study (n=6) showed me that I could expect a relatively large effect size (>.5) between experimental conditions. This was true for both EEG power and task performance. I also undertook several measures to bolster statistical power. First, I implemented an adaptive task design. This reduced variability between subjects by equating performance levels across observers through automated manipulations of task difficulty. The exact method is described in more detail in subsection 4.3.4. Secondly, an intra-subject protocol design was implemented in which each subject performed the task in all experimental conditions. This enabled the use of repeated-measures ANOVA and Wilcoxon signed-rank sum tests, known to be particularly robust at establishing significance in small-n situations. Both of these strategies were cited by Smith & Little as ways of improving predictive power in studies with smaller subject pools (Smith & Little, 2018). With this study design, I determined that a sample size of n=20 should enable me to achieve sufficient statistical power at the 5% confidence level.

4.3.2. Test environment

An HMD (HTC Vive Focus, HTC Corp.) in its default configuration was chosen for the test environment. The cognitive training task was created in Unity 3D, a programming environment commonly used for creating 3D visual content for virtual reality headsets (Unity 3D, https://unity3d.com).

HMD systems typically rely on hand-held pointers for user input. Such input devices are not appropriate for EEG studies as they could introduce muscle-related artifacts. To address this, a touchscreen smartphone was programmed to send network commands to the HMD wirelessly. As with my previous experiments, a soft, foam overlay with holes corresponding to the locations of the on-screen virtual buttons was added to the screen (Fig. 16). With this combination, the subjects could identify the smartphone controls in a tactile manner without any need to view the screen. This is crucial as the subject cannot see the smartphone screen while wearing the headset.

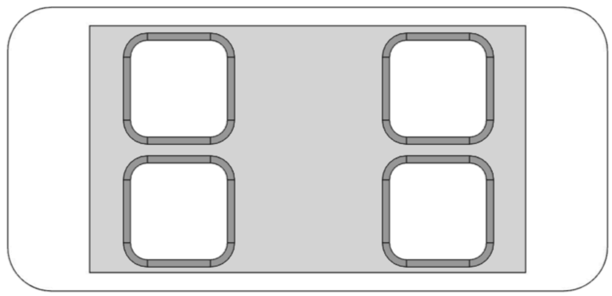


Figure 16: Answer choice locations on a touchscreen smartphone are made tactile by using a 2mm foam overlay. Button locations correspond to appropriately sized openings in the overlay.

During the experiment, subjects were seated and instructed to hold the smartphone controller in their laps, cradled by both hands. The experimental task was performed by tapping the virtual buttons on the screen with either the right-hand or left-hand thumbs while minimizing other body movements. Since the position of answer choices on the screen was randomly assigned, neither left nor right-handed subjects enjoyed an advantage from this arrangement (Fig. 17).

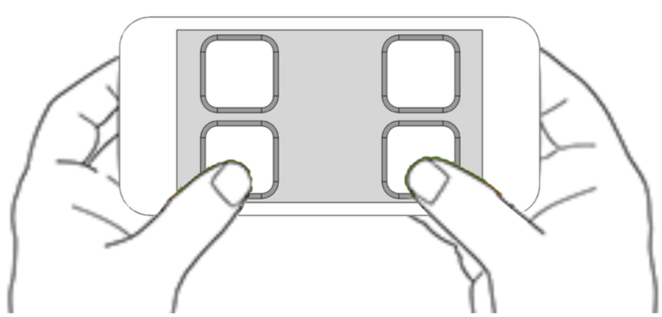


Figure 17: Using the input controller to make answer choice selections.

4.3.3. Experimental Task

As noted earlier, the experimental task was adapted from the classic n-back working-memory task. In my modified design, subjects were asked to focus on a sequence of stimuli placed in the center of the HMD screen. With each new trial, the previously displayed stimulus was shown as one of four answer choices in the corners of the display and a new stimulus took its place in the center. To proceed to the next trial, the subject was asked to tap the virtual button on the smartphone screen corresponding to the location of the stimulus that was previously in the center of the display. Once a choice was made by the subject, the answer choices disappeared and the stimulus currently at the center of the display was reassigned to one of the four corners, etc. (Fig. 18).

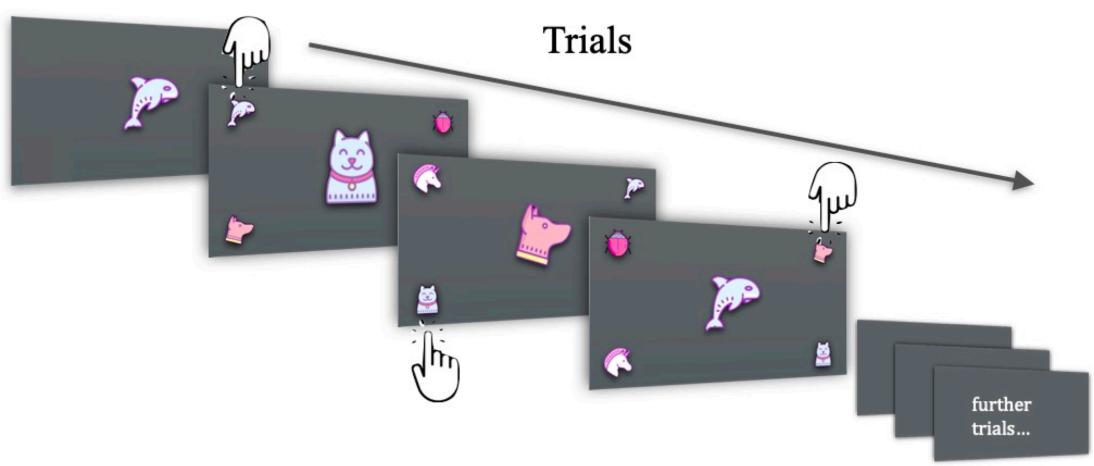


Figure 18: *Sample trials showing the center stimulus and peripheral answer choices: the current answer choices (small images) and the next stimulus (large image) are displayed simultaneously. Preceding the first trial, only the initial stimulus is displayed. Subjects select correct answers in the subsequent trials, as demonstrated above with cartoon hands. In each case, the correct answer corresponds to the center stimulus from the previous slide.*

In each trial, the center stimulus and the incorrect answer choices were selected at random by the software in such a way that no duplicate images ever appeared together. Trials lasted approximately 800 milliseconds on average (SD = 68.1ms) and were designed to elicit continuous cognitive load as both the current answer choices and the following stimulus were displayed simultaneously. An additional aim of this approach was to discourage the subject from moving their gaze away from the center stimulus, which served as a de-facto fixation point. Saccades or physical movements of the head carry with them the risk not only of negatively impacting task performance but also of introducing noise in the form of muscle artifacts into the EEG signal path.

The figures themselves are from a set of 20 cartoon animal images, all drawn in a similar style but differing in shape and color. The image collection was licensed for non-commercial use from a popular internet vendor. It was chosen for its design similarity to prevailing commercial cognitive training product designs, which frequently employ a similar cartoon design aesthetic.

4.3.4. Adaptive Task

An adaptive model was chosen for the experimental task to ensure similar engagement levels for all participants. As the experiment progressed, task difficulty was increased incrementally until the subject failed to respond within the allotted time window or made 2 or more sequential mistakes. The task difficulty level was reflected in the amount of time available for the subject to choose an answer. As the difficulty level rose, this amount of time decreased. Similarly, if the difficulty level decreased, more time was made available to complete each trial. The prevailing task difficulty level impacted the experiment in the following ways:

> 1) A visible countdown timer just below the task area reflected the amount of time allocated for making a selection. As the trial time progressed, the bar's contents filled incrementally from left to right, incentivizing the subject to answer as quickly as possible. The bar was purposefully designed to be as unobtrusive as possible so as not to distract from the primary task.

2) Failure to make a selection within the allotted time resulted in the trial being marked incorrect and the next stimulus was presented. Making any selection (whether correct or incorrect) resulted in the timer pausing briefly (200 ms) before being reset for the next trial. At the end of each trial, the response (or failure to respond) was evaluated by the software and the reaction speed and accuracy were recorded.

4.3.5. Experimental Protocol

Twenty subjects, aged 21–48 years of age (M = 28.6, SD = 7.7 years), participated in the experiment after signing an informed consent. This included 6 women and 14 men with no history of color vision disorders. The experiment was conducted in the following order: task training, EEG baseline activity measurement, experimental phases. The EEG baseline measurement phase (60 seconds) involved viewing a black background with open eyes to provide nominal cognitive activity against which to compare any changes during the experimental phases.

The phases consisted of performing the experimental task at 3 different visual angles, 10°, 20° and 40°, respectively. The visual angle (VA) was calculated using the standard formula:

VA = (S * 57.29) / D, where S is the size of the object and D is the distance from the observer.

These angles corresponded to the outer edges of the answer choices, measured horizontally (Fig. 19).

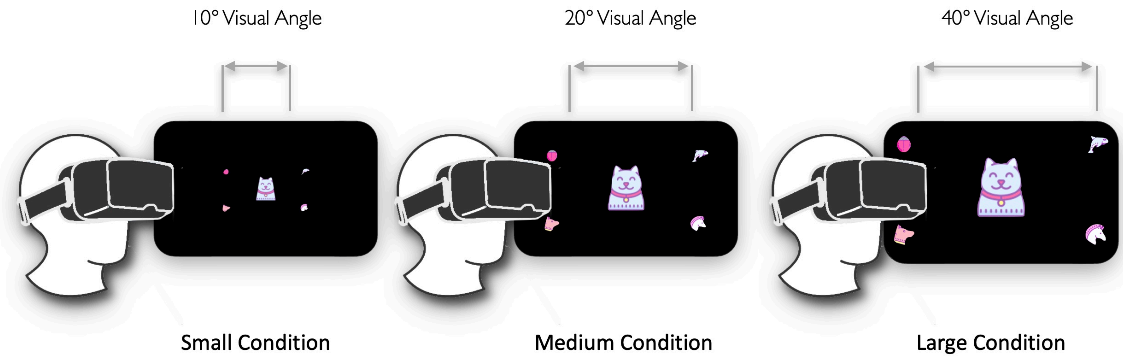


Figure 19: *The three experimental conditions emulating small, medium and large screen sizes. The cognitive task is identical in each condition.*

All visual elements were precisely placed at a virtual distance of one meter from the user as viewed within the HMD. In the Unity 3D programming environment, one unit of space is equivalent to one meter of perceived distance. Therefore, to set the visual angle for each experimental condition, I specified the desired VA and solved the equation above for S. The value of S was applied to the visual task automatically by the software with each new experimental condition, before the presentation of the first task trial.

Subjects were presented with a series of 75 task trials for each experimental condition. Every subject performed each condition twice, for a total of 6 sets per subject. A 30 second break (black screen; no visual stimulus) separated the training and baseline measurement phases to prevent contamination of the baseline data by lingering arousal from training. Between each set of trials there were additional 10 second rest breaks.

Body and particularly eye movements have a high possibility of introducing movement artifacts into the EEG data. Therefore, subjects were instructed to blink and adjust their posture as needed during these rest breaks but to refrain from doing so during the trial sets themselves.

Visual text messages in the display announced the beginning and end of these break periods. The ending message flashed off 2 seconds before the start of the following trial set. The total time required to complete each set varied according to subject ability, but lasted approximately 60 seconds on average. This resulted in an overall experimental protocol time of 9-10 minutes (Fig. 20).

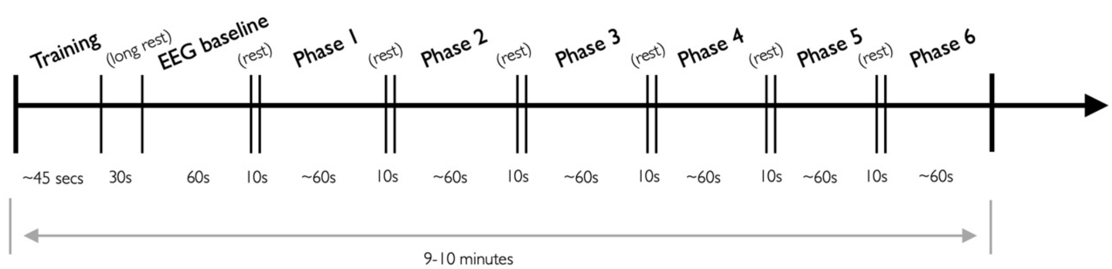


***Figure 20**: Protocol flow: following training and EEG baseline recording, 6 experimental phases, each containing 75 trials, were conducted. A 10 second rest separated each experimental phase. The content of the experimental phases was randomly selected from the 3 condition types (10°, 20° or 40°) and balanced so that each subject experienced each condition twice. Total time to complete the protocol varied from 9 to 10 minutes per subject.*

The order of the conditions was randomized for each subject but balanced so that any given size condition did not immediately follow the identical condition from the preceding set. Thus, an example order might be: small, large, medium, large, small, medium.

Before the start of the protocol, each subject was granted time to practice the task until they were able to achieve a 75% average accuracy rate for at least 10 trials. This training period lasted an average of 45 seconds, with some subjects mastering the task more quickly than others.

4.3.6. EEG

EEG signals (microvolts) were acquired from the frontal, central, occipital and parietal regions, using an 8 channel EEG amplifier (OpenBCI 32-bit Board Kit, OpenBCI, inc.) with a sampling rate of 250 Hz. The primary electrode locations were Fz and Pz, placed according to the international 10-20 system. Data from Cz and Oz were also recorded in an exploratory capacity, but the results did not offer any novel findings and were not included in the published paper. Complete data for all four electrodes is provided in Appendix A.7, however.

Gold cup electrodes were attached to the scalp and ear lobes using electro-conductive gel, and an initial impedance of <5 kΩ across all electrode positions was insured. Additional electrodes were affixed above and below the subjects' eyes to record electrooculogram (EOG) signals caused by blinking or other facial movements for later use in noise reduction and signal optimization (Zahan, 2017).

EEG data was recorded throughout the experiment, although only the final 30 seconds of activity was analyzed for each phase. This was to insure that the task difficulty was maximally adapted to each subject. The use of this mechanism made it possible to measure similar levels of cognitive engagement across all participants and all conditions.

4.3.7. Task Performance

Overall reaction time and task accuracy were calculated for each phase and averaged across all trials for a given size condition.

4.3.8. Analysis Method

The raw EEG data was notch filtered (50Hz) and high-pass filtered from 4Hz. Muscle artifacts identified from comparison with the EOG data, as well as other apparent artifacts identified by visual inspection of EEG data plots, were removed in full from the time series prior to analysis.

Fast Fourier transforms (FFT) were calculated for the following spectral ranges: theta (4-8Hz), alpha (8-13Hz), low beta (13-20Hz) and high beta (20-28Hz) with 30 second windows for each phase of the experiment. The total sum of the power values from each range was divided by the total number of data points. The resulting score was normalized by subtracting the overall population mean and dividing by the standard deviation to obtain a power index. For this study, the population mean refers to the combined EEG data of all subjects divided by number of subjects.

Statistical significance was determined with a repeated measures ANOVA, followed by a non-parametric, Wilcoxon signed rank test to determine the significance of any changes in power between the experimental phases. The choice of Wilcoxon was due to the large individual differences in performance observed between subjects, and justified by the within-subjects nature of the study.

Task performance data was averaged to obtain an overall accuracy and reaction-time value for each subject per task condition. Individual results were averaged, and similar Wilcoxon signed-rank tests conducted.

4.4. Results

4.4.1. EEG Data

Cognitive load increases corresponding to stimulus size were observed in the theta and beta ranges for all electrodes. Changes in the theta range followed a relatively linear pattern of increasing power, while the low beta results reflected a more logarithmic arc.

At electrode Fz, one-way repeated measures ANOVA showed statistically significant differences between the experimental conditions for the theta range ($f()=24.1$, $p<.00001$) and low beta range ($f()=10.86$, $p<.00001$). Post hoc Wilcoxon signed-rank tests revealed that for theta power, all increases between baseline and experimental conditions were highly significant ($p < .01$). In addition, the difference between the 10° VA and 20° VA condition was significant at the 1% threshold ($n = 20$, $Z = -3.3973$, $p < 0.01$). In contrast, changes in the low beta range were highly significant only between baseline and the 20° and 40° conditions, as well as between the 20° and 40° VA stimulus conditions (all: $p < .01$).

Unexpectedly, alpha ranges revealed relatively little change in spectral power between control and experimental phases, regardless of EEG position. Power trends for electrode position Pz were similar, but with slightly higher standard error measurements (Fig. 21). (N.B.: please see appendix A.7 for complete electrode data).

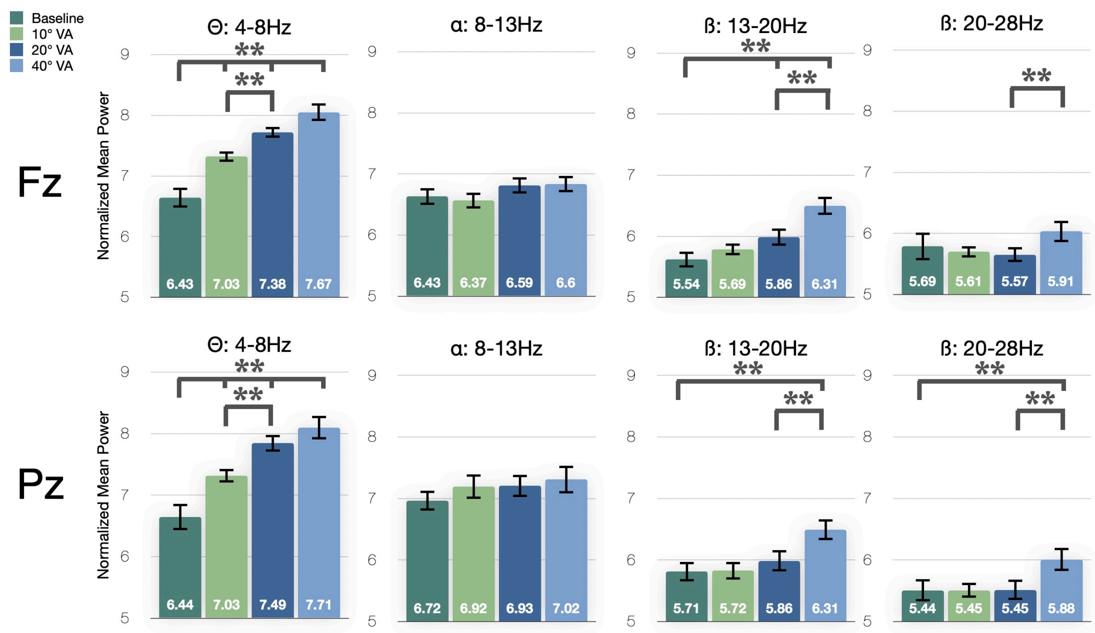


***Figure 21**: Spectral power by experimental condition (control, 10°, 20°, 40° VA) and frequency at EEG locations Fz & Pz; n=20; SE; statistical significance calculated with Wilcoxon signed-rank test (* = p < .05, ** = p < .01)*

A comparison of the average power increase from baseline further highlights these observations. In the low beta range, the percentage increase from 20° to 40° is approximately 4 times the percentage increase from 10° to 20° for all electrodes. Theta power increases, however, are more linear, with each increase in simulated screen size contributing about 5% additional spectral power to the overall cognitive load. It is noteworthy that these trends in theta activity are not concentrated in one physical area; similar, double-digit or near double-digit increases in theta power are seen across all midline electrodes (Table 4).

Table 4: Average power increase from baseline by condition and EEG position. Theta is defined as the frequency range from 4-8Hz, Alpha, 8-13Hz, low-Beta, 13-20Hz and high-Beta, 20-28Hz. Double-digit increases are shaded green.

EEG position Fz

	10° VA	20° VA	40° VA
Θ	9.2%	14.7%	19.2%
α	-0.9%	2.5%	2.7%
ß (low)	2.6%	5.8%	13.9%
ß (high)	-1.4%	-2.1%	3.8%

EEG position Pz

	10° VA	20° VA	40° VA
Θ	9.2%	16.3%	19.8%
α	3%	3.2%	4.5%
ß (low)	0.2%	2.7%	10.4%
ß (high)	0%	0.1%	8%

4.4.2. Performance Data

Changes in performance for all subjects between conditions is presented in Fig. 22.

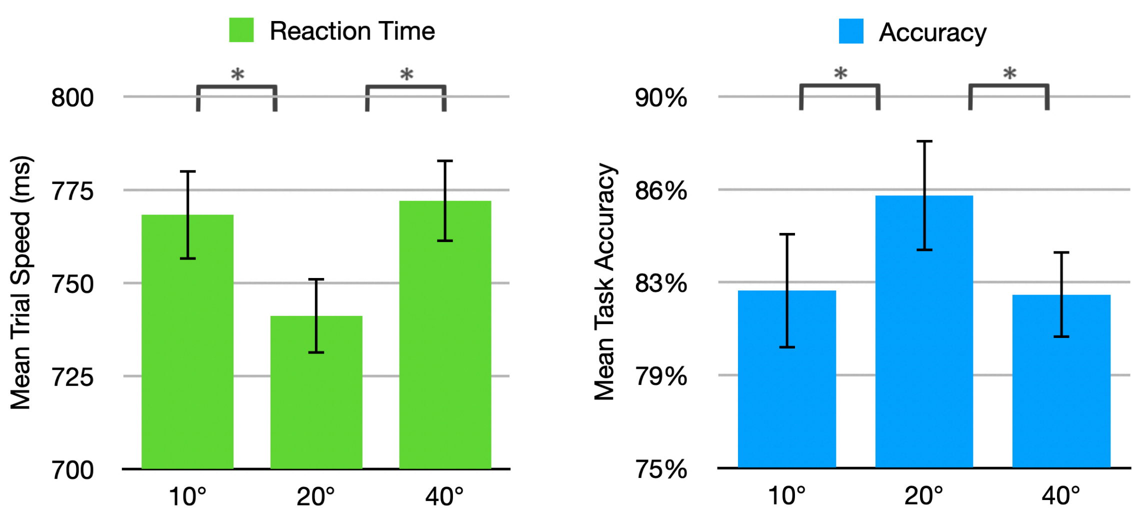


***Figure 22**: Mean task performance (reaction time and accuracy) by condition (10°, 20°, 40°); n=20; SE; statistical significance calculated with Wilcoxon signed-rank test (* = p < .05*

Overall task performance, including both reaction time and accuracy, was highest in the 20° condition. A one-way repeated measures ANOVA comparing the three conditions showed significance both for task speed (f(2)=5.13, p=.014) and accuracy (f(2)=5.734, p=.009). Additionally, post-hoc Wilcoxon signed-rank tests showed significant increases in performance for the 20° condition compared with the 10° condition (reaction time: $p=.016$, accuracy: $p=.018$). There was also a marked decrease in task performance, particularly in reaction time, for the 40° condition. When moving from the 20° condition to the 40° condition, a significant downward performance trend was observed in both reaction time and accuracy (reaction time: $p=.002$, accuracy: $p=.011$).

4.4.3. Performance & Cognitive Load

Cronbach's coefficient α was used to calculate the internal consistency coefficients of EEG power for each condition, along with the two performance outcomes. Alpha values of .71 (EEG + Reaction Time), .73 (EEG + Accuracy) and .76 (EEG + Accuracy + Reaction Time) showed a satisfactory level of internal consistency.

Unfortunately, regressing spectral power onto performance results on a per-subject basis yielded no meaningful correlations. These results were unexpected as an upward trend in performance from 10° VA to 20° VA conditions appears to mirror similar increases in EEG theta power for the averaged data sets. Similarly, the oppositional trend between 20° VA and 40° VA conditions (higher EEG power but lower performance) was marked by such significant individual differences that correlations were not meaningful (see appendix A.8).

4.5. Discussion

This study examined the role of screen size in a cognitive training task by employing several novel methods:

(1) The entirety of the experiment was conducted in an HMD environment which eliminated environmental distractions as a confounding variable.

(2) The effects of both very small and very large stimuli were examined to view the cognitive response on a size continuum.

Generally, my results were in keeping with previous studies examining cognitive impact of stimulus size manipulations. It also revealed some limitations of relying on surface EEG to predict cognitive responses to visual stimuli.

4.5.1. EEG

Short term memory processing and concentration is commonly associated with frontal cortex activity, and specifically with the frontal midline theta rhythm (Yamada, 1998). I therefore expected the largest response to the cognitive task to be in the frontal area (Fz) within the 4-8Hz theta range, regardless of experimental condition. In fact, I was surprised to find that spectral power was remarkably consistent across all electrodes tested. The parietal area revealed unexpectedly high power in the theta rhythm, often matching or surpassing the results in the frontal region.

As expected, the larger stimulus conditions generally elicited a greater cognitive response in most ranges tested. Only the alpha rhythm failed to conform to this trend, a result that was fully expected as alpha and theta rhythms often respond oppositionally during periods of mental concentration (Scharinger, Soutschek, Schubert, & Gerjets, 2017).

In general, I found that theta rhythms rose approximately 5% in spectral intensity with each increase in simulated screen size, giving the resulting graph a linear appearance. The beta rhythms, on the other hand, increased disproportionately with each change in stimulus size. This effect was particularly noticeable at electrode Pz, where the results show a clear logarithmic trend: the increase in spectral power when moving from the 20° to the 40° VA condition was approximately four times the increase in power from 10° to 20°. Since the increase in surface area when moving from a 20° to a 40° visual angle is also exactly 4x, it seems the beta rhythm may be primarily reflecting increases in visual processing load rather than changes in load related to the cognitive task per se. It is worth noting that electrode Pz is close to the visual cortex and previous research has implicated the beta rhythm in a variety of assistive roles with regards to visual perception (Kloosterman et al., 2015).

Possible reasons for the increases in the theta rhythm are less clear. The study results were obtained with an HMD in a sound-proofed, electrically shielded room. Great effort was made to minimize any impact of external visual or auditory distractions that might cause extraneous cognitive load. It is therefore tempting to attribute all changes in EEG power to an increased cognitive effort to solve the primary task. If this were the case, however, all experimental conditions that revealed significant changes in task performance should be accompanied by corresponding changes in cognitive load. Unfortunately, attempts to support this hypothesis in this study with correlations of cognitive load and performance were unsuccessful.

An additional possibility is that the adaptive nature of the study task may have contributed to differences in cognitive load. For example, during periods of downward correction (i.e., the task difficulty is reduced to accommodate subject performance) the perceived lull may have had an effect on cognitive response. However, an analysis of correction events did not reveal them to have occurred in greater numbers during any one condition type. Differences in average adaption level achieved per condition were well within the standard error and not found to be significant.

Similarly, the difference in total mean response times between the 20° and 40° conditions was approximately 2 seconds over the course of a 75-trial set. This effectively allowed for more trials to be analyzed during the 20° condition. The continuous-load design of the experimental task, however, was chosen in part to minimize any peaks and valleys in cognitive load that might accompany more traditional cognitive tasks that alternate between stimulus presentation and subject response. It is therefore unlikely that this detail alone accounted for the unexpected lack of correlation.

Overall, though some past researchers have found strong correlations between cognitive load and performance in visual tasks (Ewing et al., 2016; Alan Gevins et al., 1997), my results suggest instead that cognitive load alone cannot be relied upon to accurately predict cognitive task performance, particularly when the stability of display parameters cannot be assured. Certain types of external variables, such as stimulus size, appear to have such an outsized effect on EEG power that it becomes extremely challenging to sufficiently differentiate whether an observed increase in cognitive power was due to executive function alone or a combination of executive function and additional visual processing.

4.5.2. Task Performance

Previous studies have documented a positive correlation between stimulus size and performance in tasks dealing with visual stimuli. However, the largest size conditions in those studies were generally limited by the size of a traditional computer monitor or television screen. In contrast, this study was able to leverage the advantages of an HMD to test the size/performance conformity hypothesis with extremely large stimuli to determine if there was an upper limit to the previously observed correlation.

My results indicated that working memory task performance increases as a factor of stimulus size, but only up to a certain point. Given the "bigger is better" conclusions of past studies, it came as a surprise that the size condition that elicited the best performance in the visual memory task was not the largest, but rather the more moderately-sized 20° visual angle condition. When the task visual angle exceeded this size, both accuracy and reaction time became negatively impacted.

One possible explanation for this result may be connected with our visual processing physiology. The inner part of the retina closest to the optic nerve is densely packed with cone cells, allowing for rapid and clear discrimination of visual stimuli. This macular zone corresponds to a visual angle of approximately 18° (Remington, 2011). I therefore posit that when the majority of the experimental visual stimuli lie just within this area (i.e., the 20° VA size condition in this study), visual processing is optimal, and the brain can allocate maximum resources to the primary executive task at hand (in this case: working memory).

For the 10° visual angle condition, although well-placed in the macular area, the smaller stimulus has the disadvantage of reaching fewer receptors. In this case, the brain has to "do more with less" when evaluating the stimuli. At the 40° visual angle, however, the answer choices reside entirely in the *peripheral* area of the view field. This area is furnished with a much lower density of cone cells than the macular area (Remington, 2011). Thus, although the stimuli might cover a larger surface area of the retina, the receptor density is less than optimal and may have contributed to the reduced task performance observed.

4.6. Conclusion

The current results indicate that the size of the visual stimulus clearly and significantly impacts performance in a traditional working memory task. While these results are insufficient to draw exact specifications for an optimal stimulus size, it is clear that the small sizes of current smartphone screens are not ideal for the delivery of visual content if the goal is to maximize working memory performance.

Previous studies have shown the potential for head-mounted displays to suppress environmental distractions and increase user engagement. The current study confirms this and additionally highlights the performance benefits of an optimally sized visual stimulus, which is more easily achieved in an HMD environment compared with traditional displays. Taken together, these benefits support the use of HMDs for visual cognitive training tasks.

5. IMPACT OF VISUAL GAME-LIKE FEATURES ON COGNITIVE PERFORMANCE IN A VIRTUAL REALITY WORKING MEMORY TASK

(This chapter is based on my publication in JMIR Serious Games (Redlinger, Glas, & Rong, 2022) and my conference presentation "Enhanced Cognitive Training using Virtual Reality: Examining a Memory Task Modified for Use in Virtual Environments" at ICISPC /AIVR 2021)

5.1. Chapter Aims

This chapter documents an experiment that examined the role of gamification in the context of virtual reality-based cognitive training. The impetus for the study was a realization that the majority of cognitive tasks used in commercial training products resemble games. Nevertheless, previous research into the effects of such gamified training frequently did not control for environmental distractions during training and tended to lump many different game-like elements into a single experimental condition. As a result, study results were often contradictory and inconclusive.

The principle aim of this study was therefore to contribute to the body of gamification research by including these novel features:

1. Isolate the subject from potentially confounding environmental distractions by using an HMD for all experimental conditions.
2. Addition of 3D depth, a feature commonly found in VR-based applications, as one of the visual features examined in the context of gamified cognitive training.
3. Individually compare cognitive activity and performance for each of several gamified conditions with an unmodified, "bare-bones" version of the task
4. Examine whether changes in cognitive activity (EEG) elicited by game-like visual features correlate with task performance

5.2. Introduction

The previous chapter established that a visual angle of approximately 20° seems to be ideal for screen-based cognitive tasks as it maximizes the user performance while minimizing task-irrelevant cognitive activity. Although the previous study demonstrated robust, significant effect sizes for the experimental conditions, the task's visual design (e.g., plain stimuli against a black background) did little more than present the underlying visual memory mechanism. In contrast, cognitive training tasks intended for consumers generally contain colorful backgrounds and motivational game elements to encourage the user to stay engaged in the training process. In short, commercial cognitive training tasks often resemble games.

The embrace of gamification as a method of increasing user engagement and enjoyment of otherwise dull, repetitive tasks is indeed supported by a large number of studies (Boendermaker et al., 2017; Mekler et al., 2013; Mohammed et al., 2017; Vermeir et al., 2020). The full picture regarding the potential impact of gamification on cognitive performance, however, is less conclusive. Two recent, comprehensive reviews looking into the use of gamification strategies in brain training and general cognitive assessment studies overwhelmingly found that while gamified training does appear to boost participant motivation, study heterogeneity impeded the drawing of clear conclusions with respect to performance or ecological validity (i.e., the degree to which experimental results are generalizable to real life situations) (Lumsden, Edwards, et al., 2016; Vermeir et al., 2020). For example, the authors in the first study identified no fewer than 28 game-like elements employed in the 33 studies surveyed. These included positive/negative task feedback, time-pressure, storylines or narrative elements, interactive status displays and many others. The second survey, from 2020, found that of the 49 papers examined, no study has reported on the effect of a single element alone and that game elements were investigated only in combination, making it impossible to establish whether individual elements had measurable effects (Vermeir et al., 2020).

A typical example is Mohammed et al.'s 2017 study, which compared two adaptations of an n-back task. The two experimental conditions consisted of a stripped-down version which exposed only the core task, and a gamified version that contained a visually rich display combined with multiple audio soundtracks. While the authors found no significant performance differences between the gamified and non-gamified versions of the task, they acknowledged that the lack of granularity in the game condition ultimately made it difficult to draw useful conclusions (Mohammed et al., 2017).

Another study with a sizable subject pool (n=107) found *negative* correlations between certain game-elements and task performance. The authors speculated that unneeded stress and new cognitive demands may have been induced by distracting game elements such as persistent score displays, leading to reduced performance. However, rather than individual game elements added to a bare-bones task, the study design instead removed specific game elements from a larger group of game features. This approach seems to leave the possibility open for the remaining elements to compensate for the removal of a single one, making it difficult to know for sure which element(s) might have specifically accounted for the "new" cognitive demands (B. Katz et al., 2014).

In summary, as gamification encompasses a great number of individual elements, a lack of precision and homogeneity between studies has hampered the ability to draw consensus conclusions regarding which game-elements, if any, may impact task performance. Additionally, while motivational features such as scoreboards and real-time performance feedback have been widely studied (B. Katz et al., 2014b; Landers & Landers, 2015; Lumsden, Skinner, Coyle, Lawrence, & Munafo, 2017; Lumsden, Skinner, et al., 2016; Mekler et al., 2013; Ninaus et al., 2015), the specific impact of certain purely visual features, such as 3D depth and colorful, immersive backgrounds, is less well documented, despite being increasingly encountered in consumer products such as game systems and dedicated VR headsets.

The study described in this chapter aims therefore to add to the body of knowledge regarding the use of game-like visual design elements by specifically examining the application of two particular visual features: immersive, colorful backgrounds and the use of 3D depth. These features were specifically chosen because of their under-representation in previous studies and their frequent use in virtual and augmented reality technology, a rapidly growing consumer market segment that also contains cognitive training products. We hypothesize that task performance may be adversely affected by the additional visual processing demands but that the motivational effects documented by previous researchers may in turn compensate or reverse those effects. Finally, through the use of EEG as an additional quantitative outcome, we hope to gain insight into possible neural correlates for any observed performance impact.

5.3. Materials and Methods

5.3.1. Study Design & Sample Size Considerations

Two primary outcomes will be used to examine the impact of gamified design elements on cognitive training task performance. Cognitive activity will be broadly measured along the midline using EEG (see subsection 5.3.7 for details regarding the use of EEG in this experiment). Raw task performance will be assessed by analyzing task accuracy and subject response time. The experimental task is a simple visual memory task that requires the subject to pick out the previously displayed stimulus from among several distractors. In order to better control the testing environment, the task was coded for display in a head-mounted display environment rather than a traditional monitor screen (see subsection 5.3.2).

The use of an HMD serves two purposes: 1) to precisely control the display brightness and task visual angle across subjects and experimental conditions, and 2) to minimize potentially distracting external stimuli. For these and other reasons, several recent papers have recommended the use of HMDs, describing them as among the "most fitting platforms for applying nonpharmacological computerized neurocognitive assessments" (García-Betances et al., 2017, p. 55) and a "frontier for neurorehabilitation" (Raggi et al., 2017, p. 587).

The current experimental task was previously used in a related study exploring changes in size and position of visual stimuli and showed a robust effect size (>.5) between conditions (Redlinger et al., 2021). For this study, I undertook several additional modifications to further boost statistical power. First, to reduce between-subject variability an adaptive task design was employed in which task difficulty was automatically modulated to insure maximum subject engagement. The precise method is described in more detail in subsection 5.3.4.

Second, an intra-subject protocol design exposed each subject to all experimental conditions. This enabled the use of repeated-measures ANOVA and Wilcoxon signed-rank sum tests, known to be particularly robust at establishing significance in small-n situations (Normand, 2016; Smith & Little, 2018). With this study design, I used the G-Power algorithm to determine that a sample size of n=20 should be sufficient to achieve adequate statistical power at the 5% confidence level (Faul, Erdfelder, Buchner, & Lang, 2009).

5.3.2. Test environment

An HMD (HTC Vive Focus, HTC Corp.) in its default configuration was chosen for the test environment. The cognitive training task was created in Unity 3D, a programming environment commonly used for creating 3D visual content for virtual reality headsets (Unity 3D, https://unity3d.com).

HMD systems typically rely on hand-held pointers for user input. Such input devices are not appropriate for EEG studies, however, as they could introduce muscle-related artefacts. To address this, a touchscreen smartphone was programmed to send network commands to the HMD wirelessly. A soft, foam overlay with holes corresponding to the locations of the on-screen virtual buttons was added to the screen. With this combination, the subjects could identify the smartphone controls in a tactile manner using only their hands, without any need to view the screen. This is crucial as the subject cannot see the smartphone screen while wearing the headset (see previous figure 16).

During the experiment, subjects were seated and instructed to hold the smartphone controller in their laps, cradled by both hands. The experimental task was performed by tapping the virtual buttons on the screen with both thumbs while minimizing other body movements. Since answer positions were assigned randomly, neither left nor right-handed subjects enjoyed an advantage from this arrangement (see previous figure 17).

5.3.3. Experimental Task

In order to emulate a typical commercial cognitive training task, I designed the core task to incorporate a number of cognitive processes drawn from both gaming research and the cognitive training literature. These include visual memory recognition, divided attention, perceived time pressure and distractor avoidance. The experimental task required subjects to focus on a sequence of stimuli located in the center of the HMD screen. With the start of each new trial, the previously displayed center stimulus was moved to one of the four corners of the display and a new stimulus took its place in the center. Three randomly chosen images were placed in the remaining three corners so that the screen always contained one center image and four images in the outer corners. To proceed to the next trial, the subject was asked to identify the stimulus that was previously in the center of the display. Subjects did this by tapping the virtual button on the smartphone screen corresponding to the location of the object they wished to select. Once a choice was made by the subject, the answer choices disappeared and the stimulus currently at the center of the display was reassigned to one of the four corners, a new stimulus took its place in the center, etc. (Fig. 23).

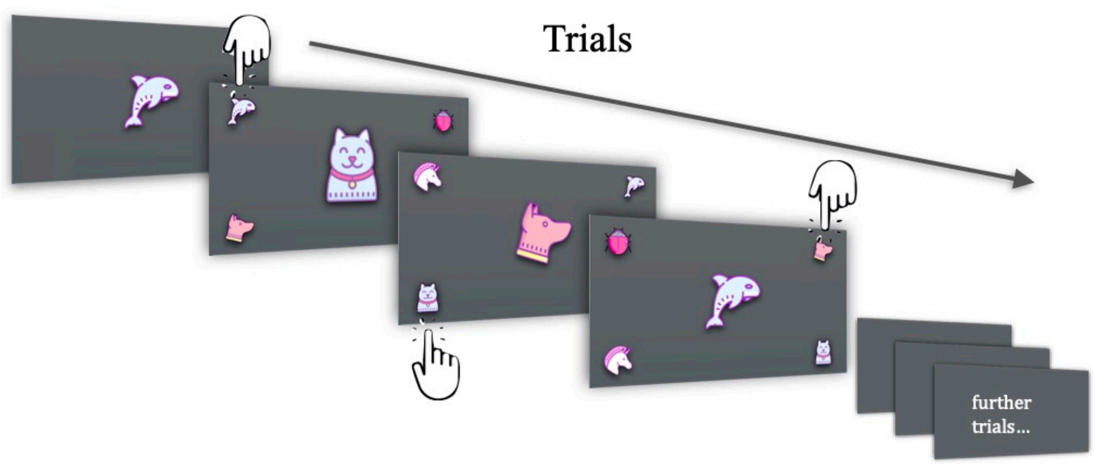


***Figure 23**: Novel visual memory task to insure continuous cognitive load: the answer choices (small images) and the stimulus from the following trial (large image) are displayed simultaneously.*

A trial was also considered ended if the allotted time elapsed before a selection was made. Please see subsequent sections 5.3.4 ("adaptive task") and 5.3.5 ("experimental protocol") for specific details related to trial times and durations.

In each trial, the center stimulus and the incorrect answer choices were selected at random by the software in such a way that no duplicate images ever appeared together. Trials lasted approximately 1.2 seconds on average (SD = 116ms) and were designed to elicit continuous cognitive load as both the current answer choices and the following stimulus were displayed simultaneously. This was to minimize the usual "peaks and valleys" in cognitive activity that often accompany tasks that alternate between stimulus presentation and subject response.

The goal in choosing this experimental task was to create a minimally complex task that could nevertheless reliably elicit cognitive load with little prior task training. While the basic mechanism is inspired by the classic n-back task, I restricted the task to 1-back in order to minimize the individual differences in performance ability commonly associated with higher degrees of *n* (Kirchner, 1958).

The figures themselves are from a set of 20 cartoon animal images, all drawn in a similar style but differing in shape and color. The image collection was licensed for non-commercial use from a popular internet vendor. It was chosen for its design similarity to prevailing commercial cognitive training product designs, which frequently employ a similar cartoon design aesthetic.

5.3.4. Adaptive Task

An adaptive model was chosen for the experimental task to ensure similar engagement levels for all participants. As the experiment progressed, task difficulty was increased incrementally until the subject failed to respond within the allotted time window or made two or more sequential mistakes. The task difficulty level was reflected in the amount of time available for the subject to choose an answer. As the difficulty level rose, this amount of time decreased in 50 millisecond intervals. Conversely, if the difficulty level decreased, more time (50ms) was made available to complete each trial. The prevailing task difficulty level impacted the experiment in the following two ways:

> 1) A visible countdown timer just below the task area reflected the amount of time allocated for making a selection. As the trial time progressed, the bar's contents filled incrementally from left to right, reminding the subject to answer as quickly as possible. The bar was purposefully designed to be as unobtrusive as possible so as not to distract from the primary task.
>
> 2) Failure to make a selection within the allotted time resulted in the trial being marked incorrect and the next stimulus was presented. Making any selection (whether correct or incorrect) resulted in the timer pausing briefly (200 ms) before being reset for the next trial.

At the end of each trial, the response (or failure to respond) was evaluated by the software and the reaction speed and accuracy were recorded. Only trials in which the subject actively made a selection were included in the reaction time assessment.

5.3.5. Experimental Protocol

Twenty subjects, aged 21–48 years of age (M = 28.6, SD = 7.7 years), were recruited from among students and staff at the Tokyo Institute of Technology and agreed to participate in the experiment after signing an informed consent. This included 6 women and 14 men, all right-handed, with no prior history of color vision disorders. In addition, all subjects reported having had previous experience using an HMD. The experimental protocol was approved by the ethics board of the Tokyo Institute of Technology (IRB: 2019059).

The protocol was executed in the following order: task training, EEG baseline activity measurement, experimental conditions. The EEG baseline measurement phase (60 seconds) involved viewing a black background with open eyes to record nominal cognitive activity with no visual stimuli.

The experimental conditions consisted of four distinct visual representations of the same core task: unmodified (the stimuli are simply placed on a flat plane against a black background), background distractor (stimuli + task-irrelevant background image), 3D depth distractor (stimuli presented at different virtual distances from the subject) and game distractor (dynamic motivational features in addition to the two previous distractors) (Fig. 24).

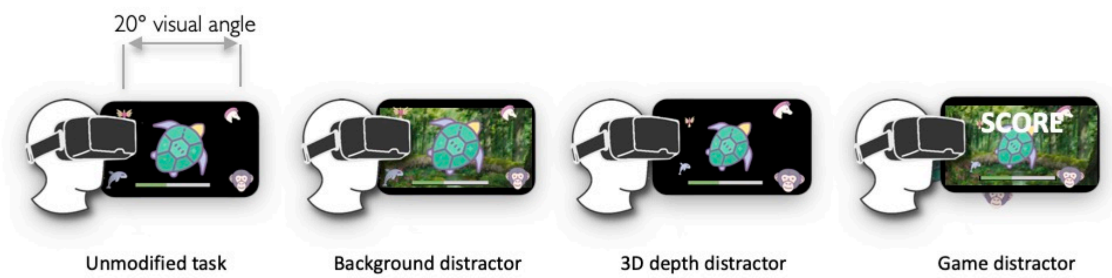


Figure 24*: The four experimental conditions: the unmodified task on a black background; the task performed atop an irrelevant, colorful background; the task performed in 3-dimensional space; the task with both background and depth distractors plus an interactive scoreboard and user feedback. Horizontal dimensions of task limited to a 20° visual angle.*

The image used in the "background distractor" condition was a cartoon forest scene, obtained from the same provider as the stimulus images. The colors, detail level and visual style are similar to that of the stimuli but there is no other obvious contextual connection. The game condition's dynamic features comprised a scoreboard and real-time performance feedback. The performance feedback was implemented as follows: an incorrect user response caused the selected answer choice to briefly shake back and forth to indicate "no", whereas a correct choice caused the item to gently pulse outward towards the user. These animations lasted exactly 200ms. In addition, a scoreboard at the top of the display indicated the current accuracy rate and the total score for the current trial set.

All experimental conditions were each repeated twice in randomized order, for a total of 8 sets per subject. Each set contained 50 trials and lasted approximately 60 seconds. Between the training and baseline phases, a 30 second break (black screen; no visual stimulus) was imposed. This was to prevent contamination of the baseline EEG data by lingering arousal from training. Between each set of trials there were additional 10 second rest breaks.

The task visual angle for all conditions was set at 20°, corresponding to the outer edges of the answer choices, measured horizontally. The visual angle (VA) was calculated using the standard formula:

> $VA = (S * 57.29) / D$, where S is the size of the object and D is the distance from the observer.

This visual angle was shown in a previous experiment to be optimal for maximizing task training performance (Redlinger et al., 2021). With the exception of the *3D depth distractor* and *game distractor* conditions, all visual task elements were precisely placed at a virtual distance of two meters from the user as viewed within the HMD. In the conditions that made use of 3D depth, the answer choices (and colorful background) remained at the same virtual distance of two meters, but the primary central stimulus moved forward to appear at a distance of one meter from the user. In the Unity 3D programming environment, one unit of space is equivalent to one perceived meter of distance. To set the visual angle for each experimental condition, we specified the desired VA and solved the equation above for S. The value of S was applied to the visual task automatically by the software with each new experimental condition, before the presentation of the first task trial.

Body and particularly eye movements have a high possibility of introducing movement artifacts into the EEG data. Subjects were therefore instructed to blink and adjust their posture as needed during the rest breaks but to refrain from doing so during the trial sets themselves.

Visual text messages in the display announced the beginning and end of these break periods. The latter message flashed off 2 seconds before the start of the following set. The total time required to complete each set of trials varied according to subject ability (as dictated by the rules of the adaptive task), but lasted approximately 60 seconds on average. This resulted in an overall experimental protocol time of 11-12 minutes (Fig. 25).

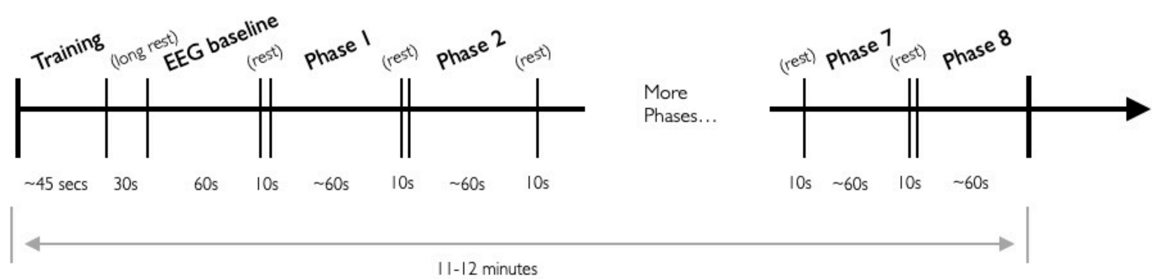


***Figure 25**: Protocol flow: following training and EEG baseline recording, 8 experimental phases, each containing 50 trials, were conducted. A 10 second rest separated each experimental phase. The content of the experimental phases was randomly selected from the 4 condition types (Unmodified task, Background distractor, 3D depth, Game distractor) and balanced so that each subject experienced each condition twice. Total time to complete the protocol varied from 11 to 12 minutes per subject.*

5.3.6. Training

Before the start of the protocol, the task rules were explained, and each subject was granted time to practice the task until they were able to achieve a 75% average accuracy rate for at least 10 trials. Some subjects mastered the task more quickly than others such that the training period lasted between 30 and 90 seconds, with an average of 44 seconds. Since the adaptive mechanism was also engaged during the training period, the training process additionally served to establish the starting difficulty for that subject for the following experimental trial sets.

5.3.7. EEG

EEG signals (microvolts) were acquired from the frontal, central, occipital and parietal regions, using a wireless 8 channel EEG amplifier (OpenBCI 32-bit Board Kit, OpenBCI, inc.) with a sampling rate of 250 Hz. The electrode locations were Fz, Cz, Oz and Pz, placed according to the international 10-20 system, and specifically selected in order to capture a broad range of activity along the midline. Electrode positions Fz and Cz were of particular interest due to their frequently cited relationship with concentration and cognitive load, while Oz and Pz were chosen due to their proximity to the visual cortex and prior association with both attention and complex visual decoding (Ewing et al., 2016; Alan Gevins et al., 1997; Sauseng et al., 2005; Yamada, 1998). Gold cup electrodes were attached to the scalp and ear lobes using electro-conductive gel, and an initial impedance of <5 kΩ across all electrode positions was ensured. Additional electrodes were affixed above and below the subjects' eyes to record electrooculogram (EOG) signals caused by blinking or other facial movements for later use in noise reduction and signal optimization (Zahan, 2017).

EEG data was recorded throughout the experiment, although only the final 30 seconds of activity was analyzed for each phase. This was to ensure that the task adaptation algorithm had been given sufficient time to adjust difficulty levels for each subject before reaching the analysis time window. Time markers for determination of analysis epochs were embedded in the EEG data stream directly using real-time network packets generated by the experimental task. Through the use of this mechanism, we hoped to be able to precisely measure similar levels of cognitive engagement for each participant.

5.3.8. Task Performance

Overall reaction time and task accuracy were calculated for each phase and averaged across all trials for a given experimental condition.

5.3.9. Analysis Method

The software used for EEG data preprocessing and analysis was MATLAB R2019b (The MathWorks, Inc., Natick, MA, USA). The raw EEG data was notch filtered (50Hz) and high-pass filtered from 4Hz using built-in butterworth and bandpass filters in MATLAB. As noted earlier, EOG data was recorded in tandem with EEG for each subject. This enabled us to create customized artefact recognition routines that were individually applied during the data preprocessing phase for each subject. Additional muscle artefacts identified from a visual inspection of EEG data plots were also removed in full from the time series prior to analysis.

Fast Fourier transforms (FFT) were calculated for the following spectral ranges: theta (4-8Hz), alpha (8-13Hz), low beta (13-20Hz) and high beta (20-28Hz) with 30 second windows for each phase of the experiment. The total sum of the power values from each range was divided by the total number of EEG data samples. The resulting score was normalized by subtracting the overall population mean (combined EEG data of all subjects divided by number of subjects) and dividing by the standard deviation to obtain a power index. FFT and statistical analysis was also performed using built-in MATLAB functions.

Shapiro-Wilk tests showed that we could not necessarily operate under an assumption of normally distributed data. Statistical significance was therefore determined with a repeated measures ANOVA, followed by a non-parametric, Wilcoxon signed rank test to determine the significance of any changes in power between the experimental phases. The choice of Wilcoxon was due to the large individual differences in performance observed between subjects, non-normally distributed data, and the within-subjects nature of the study.

Task performance data were averaged to obtain an overall accuracy and reaction-time value for each subject per task condition. Individual results were averaged, and similar Wilcoxon signed-rank tests conducted.

When looking at the preliminary data, it became quickly apparent that performance levels varied significantly from subject to subject. Some individuals were able to complete the task quickly and accurately while others struggled to respond and made frequent mistakes. This contributed to a large standard deviation in the overall results that could potentially complicate the drawing of meaningful conclusions. To address this, subjects were additionally sub-classified into high and low performance groups for further analysis. The selection criteria was based on the average overall task difficulty level achieved by each subject.

5.4. Results

5.4.1. EEG Data

The presence of gamified visual features led to observable changes in spectral power at all EEG locations. In particular, increases in beta EEG power for the 3D depth distractor condition were observed at all electrode locations. One-way repeated measures ANOVA confirmed significant differences for the high-beta range across the entire midline (Fz: (f(3, 76)=3.75, p=.015; Cz: f(3, 76)=4.09, p=.011; Pz: f(3, 76)=2.82, p=.046; Oz: f(3, 76)=2.97, p=.039), with post hoc Wilcoxon signed-rank tests showing moderate to high significance for the 3D depth condition at Oz in particular, compared with the unmodified task ($p < .05$). Differences between individual experimental conditions were, however, not significant.

For the theta range, only the results at Oz displayed significant variation (f(3, 76)=3.20, p=.029) and only one individual experimental condition, the background distractor, proved to be highly significant ($n = 20$, $Z = -2.81$, $p < 0.01$) in the post hoc analysis. The game condition, which also includes the background distractor, was found to be moderately significant ($p = .02$). Changes in the alpha rhythm did not prove to be significant at any electrode position (Fig. 26). For clarity, only the results for Fz and Oz are displayed here but full electrode data is provided in appendix A.9

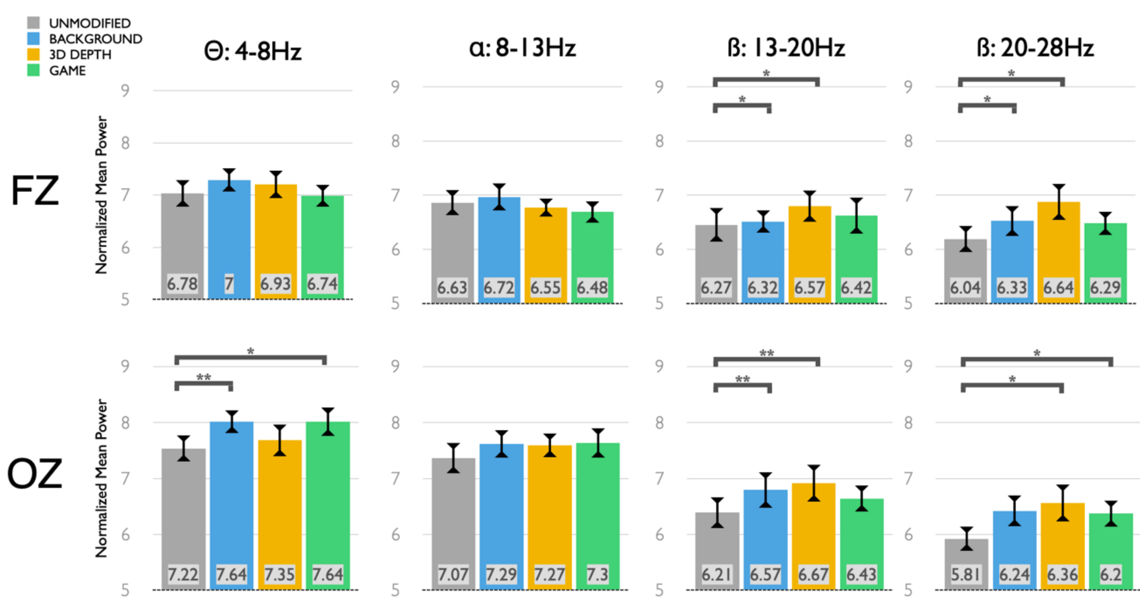


***Figure 26**: Spectral power by condition and frequency at EEG locations FZ, CZ, PZ & OZ; n=20, SE; statistical significance calculated with Wilcoxon Signed-Rank test (* = p < .05, ** = p < .01)*

It is noteworthy that the game condition, which also includes the 3D depth distractor, did not reach the same levels of cognitive activity as the depth-only condition for the beta range. This may indicate that the presence of the additional distractions in the game condition could be inhibiting the overall impact of the 3D depth effect. In the theta range, however, the presence of the background distraction in both the background and game conditions led to similar cognitive responses.

5.4.2. Performance Data

One-way repeated measures ANOVA comparing the four conditions showed no significance for either task speed ($f(3, 72)=1.21$, $p=.31$) nor accuracy ($f(3, 72)=0.143$, $p=.93$). In general, the presence of colorful, task-irrelevant backgrounds led to slight reductions in accuracy but had little impact on performance speed. Conversely, the presence of 3D depth cues seems to have slightly impacted reaction time but not accuracy (Fig. 27).

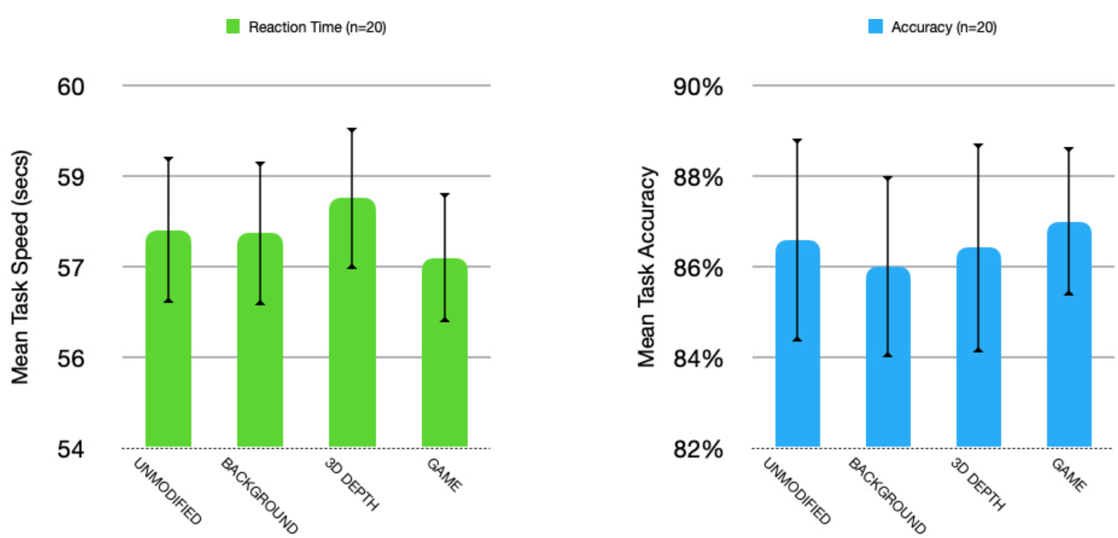


Figure 27: Mean task performance (reaction time & accuracy) by condition and group; SE

As noted earlier, I took the additional step of separating subjects into high performance and low performance groups according to ability (average maximum task difficulty achieved during all trial sets). This was due to a large standard error observed in the performance data that I felt had the potential to mask underlying trends. While the resulting subgroups are too small in size to deliver meaningful statistical power, the results did reveal several nuances and present a potentially interesting direction for a follow-up investigation.

For task accuracy the additional visual distractions present in the multiple-distraction game condition appear to have had a cumulative negative impact on the high-performers. However, a seemingly opposite effect was observed with the low performance group, who cumulatively achieved their *highest* accuracy in this condition.

In terms of task completion speed, the results do not show any significant differences between conditions, even when observing only the more internally homogenous high-performance subgroup (Fig. 28).

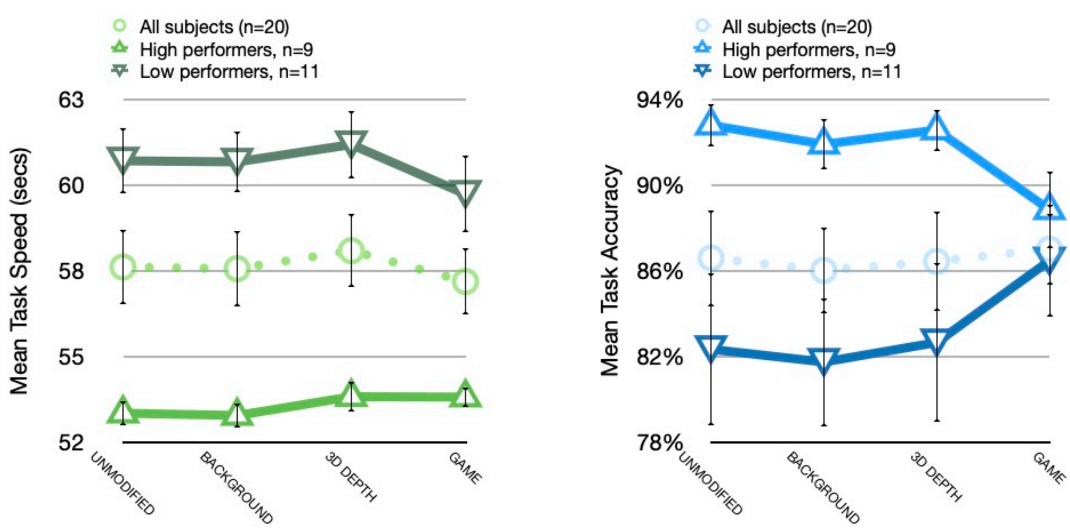


Figure 28: *Mean task performance (reaction time & accuracy) by condition and high/low performance subgroup; SE*

5.4.3. Performance/EEG compared

Perhaps due to a lack of significant differences in performance between experimental conditions, regressing EEG spectral power onto performance results produced no meaningful correlations for either the overall group nor either of the subgroups. Large individual differences in subject performance likely also contributed to the lack of significant results.

5.5. Discussion

The objective of the current study was to examine the impact of visual, game-like elements on task performance and cognitive activity in a visual working memory task. For both reaction time and task accuracy, no significant differences in performance could be determined. Nevertheless, certain performance trends can be observed that seem to leave open the possibility that specific types of visual distractions may impact some aspects of cognitive performance while leaving others unaffected. For example, our data show that visually distracting backgrounds had no observable impact on reaction speed but a slight impact on accuracy. Conversely, 3D depth decoding appears to have slightly affected speed of processing but not task accuracy.

The EEG power analysis similarly revealed no significant differences in the crucial frontal theta rhythm at Fz, which often serves as a proxy for subject concentration and task engagement (Ewing et al., 2016; Yamada, 1998). On the other hand, significant differences between conditions were observed in the beta band, and in the theta band at the occipital electrode. While these results return to insignificance if one corrects for multiple comparisons using Bonferroni or a similar method, the question nevertheless arises as to what might have caused these observed effects in the beta and theta rhythms, particularly given the lack of correlation with performance. For instance, the higher theta power observed at Oz was actually accompanied by slightly *reduced* accuracy in the background distractor condition. The proximity of Oz to the occipital area and the visual cortex suggests that rather than being directly linked with cognitive effort related to the task, perhaps the theta rhythm is instead simply more sensitive to certain underlying ocular processes required by the visually rich background employed in this condition.

For example, although the current experimental task is designed to discourage voluntary eye movements by encouraging the subject to continually focus on a center stimulus, the presence and frequency of *involuntary* eye movements such as micro-saccades were unfortunately not recorded as part of the current experimental design. Indeed, there is evidence to suggest that saccades may be highly correlated with theta power during periods of memory encoding (Sato & Yamaguchi, 2008) and that micro-saccades in particular are a source of sizable activity in the human visual cortex, with specific neural correlates in the theta and delta ranges (Dimigen, Valsecchi, Sommer, & Kliegl, 2009). More recently, Martinez-Conde and her colleagues similarly documented the presence of large potentials in the occipital cortex following micro-saccades (Martinez-Conde, Otero-Millan, & MacKnik, 2013).

Other studies have observed links between increased cognitive stress related to memory tasks and elevated saccadic frequency and duration (Di Stasi et al., 2013; Zargari Marandi, Madeleine, Omland, Vuillerme, & Samani, 2018). Thus, the possibility that the background condition may have elicited a disproportionate amount of ocular activity, and along with it increased theta power, represents one hypothesis for the observed results.

At the same time, increased high beta (20-28Hz) spectral power in the 3D depth condition was accompanied by generally slower reaction times. Here, although previous research has implicated the beta rhythm in a variety of assistive roles with regards to visual perception (Kloosterman et al., 2015), studies that specifically examine 3D decoding are less conclusive. For example, while some researchers found 3D environments elicited greater cognitive activity compared to their 2D counterparts, particularly in the beta range (Bilgin, Agres, Robinson, Wai, & Guan, 2019), Dan et al. found a *reduction* in EEG power during the 3D condition vs. the 2D condition in their experiment involving a learning task (Dan & Reiner, 2017). However, the latter study involved complex "reality-like" visuals and focused on the Fz theta/Pz alpha ratio for EEG-feature classification, rather than a broad-spectrum analysis, and did not specifically target the beta range. The possibility therefore remains that, as with the theta band, underlying cognitive demands related to visual processes may have obscured task-related activity. As noted earlier, the cognitive task employed incorporates a number of cognitive processes, including visual working memory and divided attention. This multi-modality presents a further challenge when trying to determine the exact reason for unexpected EEG results as it is difficult to ascertain which cognitive process was responsible for the observed effects.

The supplementary analysis of performance by subject ability, while not statistically meaningful, nevertheless revealed an unexpected trend with regard to task accuracy. The performance results from the high group appear to be cumulatively reduced by successive layers of distractions, with the game condition eliciting the lowest average accuracy levels. The poorer performers, however, paradoxically appeared to perform best during this condition. It must be noted, however, that the average degree of accuracy obtained in the *low* group was still well below that of the average overall performance from the *high* group.

I would like to propose two possibilities: throughout the experiment, the low performance group may have suffered from a form of performance anxiety that led to generally slower decision-making and lower overall accuracy. The presence of multiple additional visual elements in the game-condition, however, may have provided a certain degree of reassurance and encouragement, an effect of gamified design documented by previous researchers (García-Betances et al., 2017; Mekler et al., 2013). Similarly, the inclusion of a score board and positive/negative response feedback after every trial in the game condition may have helped to refocus subject attention and encourage less experienced or more easily distracted subjects to improve their performance.

Finally, it is worth noting some limitations to the current results. First, as the context of this study was potential users of commercial cognitive training products, I used a broadly inclusive criteria for subject selection, which resulted in a wide range of ages and an uneven gender balance. This may have impacted the study results in unexpected ways. Secondly, although all experimental conditions differed significantly from the unmodified task in the high beta range (except the game condition at Fz), they did not differ significantly from each other. This lack of precision reinforces the possibility that any visual novelty, whether it be the presence of 3D depth or a colorful background, triggers an increased cognitive response in the high-beta range. Greater EEG channel density and separating the multi-modal task into its component cognitive processes could potentially help isolate and differentiate the observed responses.

Lastly, the experimental conditions in experiment three represent a compromise between a fully gamified product, comprising potentially dozens of individual game elements, and an unembellished clinical task. Since the study objective was to assess the impact of individual elements, even the comparatively feature-rich final condition contained only four specific theme and feedback features. It is, however, possible that the true effects of gamification on task performance are only to be found in aggregate, that is to say, with a full complement of features. If so, the task employed in this study perhaps failed to reach a minimum level of entertainment needed to affect motivation and performance. The inclusion of additional acknowledged gamification elements such as *narrative*, *rewards*, *consequences* or *social pressure* (Marczewski, 2015) might have resulted in a different study outcome.

5.6. Conclusion

In isolation, a small performance impact was incurred by the inclusion of task-irrelevant visual backgrounds and the use of 3D depth elements. That impact was mitigated or reversed for some subjects, however, when combined with "motivating" features such as real-time feedback and scoreboards. Overall, the primary finding of the current study is that performance in simple memory tasks of the kind that are frequently found in commercial cognitive training apps is not significantly affected by the use of visually distracting backgrounds or 3D depth, nor by common motivational game elements such as scoreboards and real-time performance feedback. Particularly in light of the user engagement and motivational advantages of gamification documented by previous researchers, the observed impacts may not be substantial enough to warrant the creation of specific design patterns or the redesigning of existing gamified cognitive tasks unless the specific goal is to maximize speed and/or accuracy, in which case the current findings may provide some useful guidance.

6. GENERAL DISCUSSION

6.1. Chapter Aims

This thesis is the culmination of three years of laboratory research examining the current state of app-based cognitive training. In light of recent advances in consumer technologies, I was particularly interested to see whether today's tiny smartphone screens and the general trend towards game-like experiences had any noticeable impact on cognitive task performance, or whether next-generation head-mounted displays paired with a minimally distractive core-task-only design might be a more effective option. Along the way, I conducted an extensive review of the cognitive training literature, developed my own tasks and VR-based platform for deploying cognitive training, and conducted three quantitative experiments making use of current advances in portable, wireless EEG technology. This chapter synthesizes the results I have obtained and reflects on my findings in the context of the broader debate about cognitive training. It has five aims:

1) Answer and contextualize the central question of my thesis: is the current state of commercial cognitive training, namely gamified apps "played" on smartphones, sufficient to achieve robust cognitive training?
2) Discuss the pros and cons of using HMDs in the context of cognitive training
3) Address the limitations of my research and specifically the use of EEG in cognitive training studies
4) Discuss the challenges remaining in the field and suggest directions for future research
5) Summarize the various findings of this thesis into a pragmatic list of "best practices" to follow when designing cognitive training apps

6.2. The Verdict on App-based Cognitive Training

The needs of science (e.g., academic rigor, quantifiable results, etc.) are often in opposition to the needs of the consumer (e.g., convenience, affordability). Unsurprisingly, reaching a verdict on the relative merits of today's convenient, app-based cognitive training products is complicated and involves making difficult trade-offs.

On the one hand, it seems increasingly clear that small typefaces on small screens are not the best way to achieve optimal performance with any cognitive process that involves working memory. This was determined by recent e-learning experiments that compared screen-sizes, along with my own findings of statistically significant performance improvements with larger (but not extremely large) stimuli.

At the same time, the prevailing evidence shows that gamified cognitive tasks *do* increase the quality of the experience and possibly the desire to continue performing a task (Boendermaker et al., 2017; Mekler et al., 2013; Mohammed et al., 2017; Vermeir et al., 2020). Since the success of cognitive training is generally believed to be dependent on a regimen of repeated, spaced training (Chiu et al., 2017; Edwards et al., 2018) the potentially improved user engagement found with gamified training apps may be essential if there is to be any expectation of benefit from casual training. Furthermore, certain game features that focus on motivation and progress tracking rather than the presentation of task-irrelevant visual elements do not appear to significantly impact performance. In such cases, their use is not only justified but should be encouraged.

6.3. To Be or Not to Be (…Virtual)

The use of head-mounted displays for training, learning and entertainment represents the current vanguard of human-computer interaction. As with gamified training apps, however, the use of HMDs also presents both advantages and disadvantages.

Of particular interest for cognitive research and consumers alike is the HMD's inherent advantage of reduced peripheral distraction and potential for optimized stimulus presentation. The binocular nature of standard VR headsets enables the precise positioning of stimuli at virtual distances that correspond perceptually to their real-world equivalents. As reported previously in this thesis, one area of concern when attempting to use clinical research to support claims of commercial training benefits is that the environmental circumstances under which clinical training takes place often differ greatly from those of casual environments, such as a typical home. Use of an HMD for both clinical studies and casual home use has the potential to increase ecological validity of study results by reducing this discrepancy.

A chief disadvantage of HMDs, as of Spring, 2022, remains their awkward size and weight. In order to dispense with the need for a high-performance computer to power the headset, today's HMDs instead integrate the necessary optics, processors and power-supply into the headset itself, greatly increasing their mass and physical profile. In addition, the latency between the user's movements in the real world and the delayed or non-existent representation of those movements in virtual space often leads to a physiological queasiness that is referred to as "VR-sickness" or "cybersickness" (García-Betances et al., 2017; LaViola, 2000). To compensate for these effects, ample back and neck support and limiting the amount of movement needed to complete a task (all techniques I employed in my thesis research) are recommended procedures to manage the circumstances that reportedly lead to cybersickness.

In sum, while their use in clinical settings is extremely promising for visual stimulus studies, phobia desensitization and other similar research, the cumbersome nature and discomfort associated with the current generation of consumer HMDs must be adequately addressed before their value as a platform for commercial cognitive training can be fully realized.

6.4. Thesis Limitations

One stated objective of my research was to collect both performance-based and bio-potential-based quantitative data in order to data-mine the results for potential correlations. The nature of bio-potentials is that while they are able to provide us precise information about physiological changes in our organism, they are generally not able to specify an exact relationship with any particular behavior or cognition. Attempting to make causal inferences based only on biometric data introduces the risk of misleading and un-scientific conclusions of the sort I introduced earlier in this thesis (see section 2.2 "We Literally Love our iPhones"). It was therefore my hope that by using a simple cognitive task in which performance and EEG were both dependent variables subjected to the same stimulus manipulations I would be able to gain valuable insight into the interrelation between the two outcomes.

Unfortunately, the challenge of separating EEG signals related to visual processing from those related to the core cognitive task proved to be daunting and represents one of the primary limitations of the research documented in this thesis. While spectral power changes due to manipulations of task difficulty (e.g., increased time pressure or higher degrees of n in an n-back task) have been demonstrated to be readily observable in previous studies (Ewing et al., 2016; Fairclough, Gilleade, Ewing, & Roberts, 2013), I found in my own experiments that even slight manipulations of the visual stimulus often resulted in such dramatic EEG fluctuations that the more subtle changes in spectral power related to the experimental task tended to be obscured. In other words, while I regularly documented changes in EEG power during my various experimental conditions, those changes were not always reflected in task performance. On the contrary, they tended to correlate instead with increased visual complexity or larger stimulus sizes. Conversely, seemingly clear trends in subject performance failed to correlate with corresponding EEG data.

The sensitivity of EEG results to a wide variety of potential cognitive activity, and the resulting difficulty in establishing direct causality, has been documented by other researchers as well. For instance, Grissmann et al.'s 2017 EEG study attempted to find a meaningful link between cognitive task performance and affective mental states using EEGs and other outcomes. While they did succeed in determining that negative valence states were correlated with increased working memory load and reduced task performance, they noted that the EEG results typically associated with changes in working memory load failed to be observed. The authors speculated that the very act of presenting affect-inducing stimuli, in the form of high/low valence images, may have generated cognitive activity that obscured or suppressed the expected EEG fluctuations (Grissmann et al., 2017). Mühl and Jeunet reached similar conclusions while researching the effects of social stress on cognitive task performance, finding that workload estimation via EEG at times failed to match actual performance due to interference from the stress itself (Mühl et al., 2014).

A further limitation of this thesis involved a conscious trade-off between protocol efficiency and subject comfort on the one hand, and EEG coverage and data resolution on the other.

For example, the relative novelty of combining surface EEG technology together with a head-mounted display was only possible because of a conscious choice to limit the number and placement of the EEG electrodes. The use of HMDs generally requires a not-insignificant amount of adjustment to straps and optics to achieve good focus and a comfortable fit for the subject. All of this must necessarily occur after the placement and calibration of the EEG electrodes on the surface of the scalp has concluded. With such a strong potential for damaging the fragile electrode connections in the process, initial placement and wiring must be carefully considered. Consequently, while there is no doubt that a more comprehensive EEG skullcap with a larger number of electrodes and greater surface coverage might have enabled a better separation and understanding of the various EEG spectral effects observed in these experiments, its use may have in turn contributed to increased user discomfort, protocol complexity, and a greater risk to data integrity.

6.5. Challenges and Future Work

In light of the recent pandemic, the search for ways to conduct cognitive research without the need to convene subjects in a clinical setting has intensified. As previously noted, the use of crowd-sourced subject pools has the advantage of simplified, relatively affordable access to large numbers of subjects (Lumsden, Skinner, et al., 2016). The types of experiments that can be run in this distributed manner, however, is greatly limited by the inability to insure a consistent and comparable experimental environment for all subjects. The systematic implementation of HMDs has the potential to revolutionize crowd-sourced cognitive studies by solving this exact problem: placing all subjects in a virtual, controlled environment regardless of their actual physical or social context.

Compared to global smartphone sales of more than 1 billion units in 2020, unit sales of popular VR headsets such as Meta Platforms, Inc.'s Quest 1 and Quest 2 is still tiny at approximately 2 million units (Alsop, 2021; "Number of smartphones sold to end users worldwide from 2007 to 2021," 2021). Nevertheless, forthcoming releases of third-generation headsets from leading technology companies such as HTC, Meta and Apple will likely have a dramatic impact on VR adoption. It may indeed soon be realistic to require online subjects to use an approved headset for cognitive experiments. Furthermore, announced features of third-generation units include thinner lenses and the use of external batteries, resulting in dramatically less weight and improved user comfort.

In terms of future research, one direction I hope to pursue involves the use of eye-tracking as an alternate means of gauging task engagement and cognitive effort, potentially in combination with EEGs. While EEG is among the gold-standard technologies for measuring cognitive function, the limitations documented in this thesis have led me to seek out potential alternatives. Eye-tracking technology represents just such an alternative: a robust research discipline comprising decades of rigorous academic study with a wide variety of assessment techniques and device manufacturers. Furthermore, eye-tracking metrics such as saccade dynamics are highly correlated to a wide variety of cognitive functions (Di Stasi et al., 2013; Zargari Marandi et al., 2018) and are sensitive to many of the same task manipulations, including time pressure, as EEG (McCarley, 2009). In terms of practicality, certain research-grade HMDs already have eye-tracking technology built-in and some next-generation consumer headsets are also rumored to contain this feature. The possibility of pairing an eye-tracking enabled HMD with EEG, heart rate variability (HRV) or other biopotential outcomes constitutes a very promising direction for future quantitative research.

6.6. Conclusion / Cognitive Training Task Design Best Practices

One recurring point that has hopefully been made clear in this thesis is that small differences in protocol or stimulus presentation can have a dramatic impact on the results of cognitive studies. Sometimes it can even mean the difference between meaningful effect sizes and none at all. This fact has been particularly apparent in the field of cognitive training research, where confusing and contradictory study results have gone so far as to result in diametrically opposed schools of thought.

Nevertheless, I feel it has become increasingly difficult to hold the opinion that cognitive training is ineffective *in all cases* – respected, longitudinal studies such as ACTIVE have hopefully addressed this. The challenge for future cognitive scientists and makers of cognitive training products alike is rather how to best use the knowledge gained from a myriad of clinical studies and our growing understanding of the psychophysics of cognition to maximize the cognitive benefit from training. To this end, I would like to conclude by proposing the following best practices for cognitive training task design:

1) Design for larger screens. Recommend that your users train on larger displays. A tablet is better than a handheld phone. When the choice is available, increase the font-size of the text. Large, clear stimuli that are easy to read without excessive eye or head movements will allow the brain to focus on the underlying task, rather than wasting energy to first parse the stimulus.

2) Use game elements that facilitate task engagement, but with a light touch. Visual indicators that periodically show success and progress are beneficial, but are best presented after the completion of a task rather than as permanently visible

elements. Use of colorful graphical elements should be restricted to meaningful areas of the task, such as the stimuli themselves, rather than unrelated background decoration. There is some leeway, however (e.g., if the graphics specifically encourage prolonged participation or directly promote task engagement).

3) Make the task adaptive. Adaptive designs have been found to improve efficiency and performance not just in cognitive training but in learning and memory tasks as well. Simple adaptations that are easily implemented include variable time pressure and decreasing stimulus display duration.

4) Encourage your users to use an HMD or else seek out a quiet, shaded, non-social area to train. Research has shown that the amount of ambient sound, light and a variety of psychosocial factors impact concentration and cognitive performance. If you use an HMD, take the time to ensure maximum comfort and optical clarity before beginning.

5) Stick with it! For healthy adults, plan for at least three or more sessions per week with a combined weekly training time of about two hours, if possible. In addition, studies show that gains appear only after approximately ten total hours of training. For maximum long-term training effectiveness, a training regimen should be maintained for at least eight weeks.

All of these suggestions are derived from either my original experimental results or well-substantiated and documented conclusions reached by previous researchers. Because what is at stake in the brain training debate is nothing short of our collective cognitive health, the process of identifying and vetting potentially promising interventions is necessarily a complicated and painstaking one. Competing interests from commercial and academic actors, along with technological and methodological difficulties, all contribute to the challenge of reaching a clear consensus. Nevertheless, the need for a comprehensive strategy to address cognitive decline has never been greater. We might yet see the day when a universally acknowledged cognitive training practice becomes as commonplace as performing a standing squat or mounting a treadmill.

REFERENCES

A Consensus on the Brain Training Industry from the Scientific Community. (2014). Retrieved July 9, 2020, from http://longevity3.stanford.edu/blog/2014/10/15/the-consensus-on-the-brain-training-industry-from-the-scientific-community-2/

Al Ghamdi, E., Yunus, F., Da'ar, O., El-Metwally, A., Khalifa, M., Aldossari, B., & Househ, M. (2016). The Effect of Screen Size on Mobile Phone User Comprehension of Health Information and Application Structure: An Experimental Approach. *Journal of Medical Systems*, *40*(1), 1–8. https://doi.org/10.1007/s10916-015-0381-5

Alsop, T. (2021). VR headset unit sales worldwide 2019-2024. Retrieved November 29, 2021, from https://www.statista.com/topics/2532/virtual-reality-vr/#dossierKeyfigures

Antonenko, P. D. (2007). *The effect of leads on cognitive load and learning in a conceptually rich hypertext environment*. https://doi.org/10.31274/RTD-180813-16756

Appelbaum, L. G., & Erickson, G. (2018). Sports vision training: A review of the state-of-the-art in digital training techniques. *International Review of Sport and Exercise Psychology*, *11*(1), 160–189. https://doi.org/10.1080/1750984X.2016.1266376

Artuso, C., Carretti, B., & Palladino, P. (2019). Short-term training on working memory updating and metacognition in primary school: The effect on reading comprehension. *School Psychology International*, *40*(6), 641–657. https://doi.org/10.1177/0143034319881671

Ball, K., & Owsley, C. (1993). The useful field of view test: a new technique for evaluating age-related declines in visual function. *Journal of the American Optometric Association*, *64*(1), 71–79. Retrieved from https://www.unboundmedicine.com/medline/citation/8454831/The_useful_field_of_view_test:_a_new_technique_for_evaluating_age_related_declines_in_visual_function_

Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (2001). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *39*(2–3), 241–248. https://doi.org/10.1016/S0167-8760(00)00145-8

Bauco, C., Borriello, C., Cinti, A. M., Martella, S., Zannino, G., Rossetti, C., … Marigliano, V. (1998). Correlation between MMSE performance, age and education in centenarians. *Archives of Gerontology and Geriatrics*, *26*(SUPPL.1), 23–26. https://doi.org/10.1016/S0167-4943(98)80004-6

Bavelier, D., & Green, C. S. (2019). Enhancing Attentional Control: Lessons from Action Video Games. *Neuron*, *104*(1), 147–163. https://doi.org/10.1016/J.NEURON.2019.09.031

Beauchamp, M. S., Cox, R. W., & Deyoe, E. A. (1997). Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *Journal of Neurophysiology*, *78*(1), 516–520. https://doi.org/10.1152/JN.1997.78.1.516/ASSET/IMAGES/LARGE/JNP.JY20F3.JPEG

Bilgin, P., Agres, K., Robinson, N., Wai, A. A. P., & Guan, C. (2019). A comparative study of mental states in 2D and 3D virtual environments using EEG. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, *2019-Octob*, 2833–2838. https://doi.org/10.1109/SMC.2019.8914326

Boeldt, D., McMahon, E., McFaul, M., & Greenleaf, W. (2019). Using Virtual Reality Exposure Therapy to Enhance Treatment of Anxiety Disorders: Identifying Areas of Clinical Adoption and Potential Obstacles. *Frontiers in Psychiatry*, *10*, 773. https://doi.org/10.3389/FPSYT.2019.00773/BIBTEX

Boendermaker, W. J., Veltkamp, R. C., & Peeters, M. (2017). Training Behavioral Control in Adolescents Using a Serious Game. *Games for Health Journal*, *6*(6), 351–357. https://doi.org/10.1089/g4h.2017.0071

Bonnechère, B., Klass, M., Langley, C., & Sahakian, B. J. (2021). Brain training using cognitive apps can improve cognitive performance and processing speed in older adults. *Scientific Reports 2021 11:1*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-91867-z

Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working Memory Training in Older Adults: Evidence of Transfer and Maintenance Effects. *Psychology and Aging*, *25*(4), 767–778. https://doi.org/10.1037/A0020683

Brain Training Consensus - Rebuttal. (2014). Retrieved December 9, 2021, from https://www.cognitivetrainingdata.org/studies-cognitive-training-benefits/

BrainHQ: Brain Training that Works. (2020). Retrieved July 9, 2020, from https://www.brainhq.com/

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*, 109–131. https://doi.org/10.1146/ANNUREV.PS.40.020189.000545

Broman, A. T., West, S. K., Muñoz, B., Bandeen-Roche, K., Rubin, G. S., & Turano, K. A. (2004). Divided Visual Attention as a Predictor of Bumping while Walking: The Salisbury Eye Evaluation. *Investigative Ophthalmology & Visual Science*, *45*(9), 2955–2960. https://doi.org/10.1167/IOVS.04-0219

Cardoso-Leite, P., Joessel, A., & Bavelier, D. (2020). Games for enhancing cognitive abilities. *Handbook of Game-Based Learning*, 347–365. Retrieved from https://orbilu.uni.lu/handle/10993/46517

Charvet, L. E., Yang, J., Shaw, M. T., Sherman, K., Haider, L., Xu, J., & Krupp, L. B. (2017). Cognitive function in multiple sclerosis improves with telerehabilitation: Results from a randomized controlled trial. *PLOS ONE*, *12*(5), e0177177. https://doi.org/10.1371/JOURNAL.PONE.0177177

Chen, J., McManus, M., Valsecchi, M., Harris, L. R., & Gegenfurtner, K. R. (2019). Steady-state visually evoked potentials reveal partial size constancy in early visual cortex. *Journal of Vision*, *19*(6), 1–15. https://doi.org/10.1167/19.6.8

Chi, H., Agama, E., & Prodanoff, Z. G. (2017). Developing serious games to promote cognitive abilities for the elderly. *2017 IEEE 5th International Conference on Serious Games and Applications for Health, SeGAH 2017*. https://doi.org/10.1109/SEGAH.2017.7939279

Chiu, H.-L., Chu, H., Tsai, J.-C., Liu, D., Chen, Y.-R., Yang, H.-L., & Chou, K.-R. (2017). The effect of cognitive-based training for the healthy older people: A meta-analysis of randomized controlled trials. *PLOS ONE*, *12*(5), e0176742. https://doi.org/10.1371/JOURNAL.PONE.0176742

Clay, O. J., Wadley, V. G., Edwards, J. D., Roth, D. L., Roenker, D. L., & Ball, K. K. (2005). Cumulative meta-analysis of the relationship between useful field of view and driving performance in older adults: current and future implications. *Optometry and Vision Science : Official Publication of the American Academy of Optometry*, *82*(8), 724–731. https://doi.org/10.1097/01.OPX.0000175009.08626.65

Collective. (n.d.). Category:Video game genres - Wikipedia. Retrieved January 6, 2022, from https://en.wikipedia.org/wiki/Category:Video_game_genres

Conklin, H. M., Ashford, J. M., Clark, K. N., Martin-Elbahesh, K., Hardy, K. K., Merchant, T. E., … Zhang, H. (2017). Long-Term Efficacy of Computerized Cognitive Training Among Survivors of Childhood Cancer: A Single-Blind Randomized Controlled Trial. *Journal of Pediatric Psychology*, *42*(2), 220–231. https://doi.org/10.1093/JPEPSY/JSW057

Dan, A., & Reiner, M. EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. , 122 International Journal of Psychophysiology § (2017).

Deng, W., Hu, D., Xu, S., Liu, X., Zhao, J., Chen, Q., … Li, X. (2019). The efficacy of virtual reality exposure therapy for PTSD symptoms: A systematic review and meta-analysis. *Journal of Affective Disorders*, *257*, 698–709. https://doi.org/10.1016/J.JAD.2019.07.086

Di Stasi, L. L., Mccamy, M. B., Catena, A., Macknik, S. L., Cañas, J. J., & Martinez-Conde, S. (2013). Microsaccade and drift dynamics reflect mental fatigue. *European Journal of Neuroscience*, *38*(3), 2389–2398. https://doi.org/10.1111/ejn.12248

Dimigen, O., Valsecchi, M., Sommer, W., & Kliegl, R. (2009). Human Microsaccade-Related Visual Brain Responses. *Journal of Neuroscience*, *29*(39), 12321–12331. https://doi.org/10.1523/JNEUROSCI.0911-09.2009

Donker, T., Cornelisz, I., van Klaveren, C., van Straten, A., Carlbring, P., Cuijpers, P., & van Gelder, J. (2019). Effectiveness of Self-guided App-Based Virtual Reality Cognitive Behavior Therapy for Acrophobia: A Randomized Clinical Trial. *JAMA Psychiatry*, *76*(7), 682–690. https://doi.org/10.1001/JAMAPSYCHIATRY.2019.0219

Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*(3), 99–105. https://doi.org/10.1016/J.TICS.2008.01.001

Edwards, J. D., Fausto, B. A., Tetlow, A. M., Corona, R. T., & Valdés, E. G. (2018, January 1). Systematic review and meta-analyses of useful field of view cognitive training. *Neuroscience and Biobehavioral Reviews*, Vol. 84, pp. 72–91. https://doi.org/10.1016/j.neubiorev.2017.11.004

Edwards, J. D., Xu, H., Clark, D. O., Guey, L. T., Ross, L. A., & Unverzagt, F. W. (2017). Speed of processing training results in lower risk of dementia. *Alzheimer's & Dementia (New York, N. Y.)*, *3*(4), 603–611. https://doi.org/10.1016/j.trci.2017.09.002

Eramudugolla, R., Kiely, K. M., Chopra, S., & Anstey, K. J. (2017). Effect of Speed of Processing Training on Older Driver Screening Measures. *Frontiers in Aging Neuroscience*, *9*. https://doi.org/10.3389/fnagi.2017.00338

Ewing, K. C., Fairclough, S. H., & Gilleade, K. (2016). Evaluation of an adaptive game that uses EEG measures validated during the design process as inputs to a Biocybernetic loop. *Frontiers in Human Neuroscience*, *10*(MAY2016). https://doi.org/10.3389/fnhum.2016.00223

Facebook Inc. (2022). Top Selling | Oculus. Retrieved January 6, 2022, from https://www.oculus.com/experiences/quest/section/274907549851488/#/?_k=138hoi

Fairclough, S. H., Gilleade, K., Ewing, K. C., & Roberts, J. (2013). Capturing user engagement via psychophysiology: Measures and mechanisms for biocybernetic adaptation. *International Journal of Autonomous and Adaptive Communications Systems*, Vol. 6, pp. 63–79. https://doi.org/10.1504/IJAACS.2013.050694

Fairclough, S. H., & Venables, L. (2006). Prediction of subjective states from psychophysiology: a multivariate approach. *Biological Psychology*, *71*(1), 100–110. https://doi.org/10.1016/J.BIOPSYCHO.2005.03.007

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G * Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods 2009 41:4*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381–391. https://doi.org/10.1037/H0055392

Fitzgerald, S. (2017). Is Successful Brain Training Fake News? *Neurology Today*, *17*(7), 8. https://doi.org/10.1097/01.NT.0000516006.46242.F6

Fournier, L. R., Wilson, G. F., & Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: Manipulations of task difficulty and training. *International Journal of Psychophysiology*, *31*(2), 129–145.

https://doi.org/10.1016/S0167-8760(98)00049-X

Franks, P. W., & Boutcher, S. H. (2003). Cardiovascular response of trained preadolescent boys to mental challenge. *Medicine and Science in Sports and Exercise*, *35*(8), 1429–1435. https://doi.org/10.1249/01.MSS.0000078935.78895.8E

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/J.1467-9280.2006.01681.X

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual Differences in Executive Functions Are Almost Entirely Genetic in Origin. *Journal of Experimental Psychology. General*, *137*(2), 201. https://doi.org/10.1037/0096-3445.137.2.201

García-Betances, R. I., Cabrera-Umpiérrez, M. F., & Arredondo, M. T. (2017). Computerized neurocognitive interventions in the context of the brain training controversy. *Reviews in the Neurosciences*, *29*(1), 55–69. https://doi.org/10.1515/revneuro-2017-0031

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex*, *10*(9), 829–839. https://doi.org/10.1093/CERCOR/10.9.829

Gevins, Alan, Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors*, *40*(1), 79–91. https://doi.org/10.1518/001872098779480578

Gevins, Alan, Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, *7*(4), 374–385. https://doi.org/10.1093/cercor/7.4.374

Gooding, Alc., Choi, J., Fiszdon, J., Wilkins, K., Kirwin, P., van Dyck, C., … Mindt, Rivera, M. (2016). Comparing three methods of computerised cognitive training for older adults with subclinical cognitive decline. *Neuropsychological Rehabilitation*, *26*(5–6), 810–821. https://doi.org/10.1080/09602011.2015.1118389

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. https://doi.org/10.1038/NATURE01647

Green, C. S., & Bavelier, D. (2007). Action-Video-Game Experience Alters the Spatial Resolution of Vision. *Psychological Science*, *18*(1), 88. https://doi.org/10.1111/J.1467-9280.2007.01853.X

Greenlee, M. W. (2000). Human cortical areas underlying the perception of optic flow: brain imaging studies. *International Review of Neurobiology*, *44*, 269–292. https://doi.org/10.1016/S0074-7742(08)60746-1

Grissmann, S., Faller, J., Scharinger, C., Spüler, M., & Gerjets, P. (2017). Electroencephalography Based Analysis of Working Memory Load and Affective Valence in an N-back Task with Emotional Stimuli. *Frontiers in Human Neuroscience*, *11*, 616. https://doi.org/10.3389/fnhum.2017.00616

Guo, J., Weng, D., Been-Lirn Duh, H., Liu, Y., & Wang, Y. (2017). Effects of using HMDs on visual fatigue in virtual environments. *Proceedings - IEEE Virtual Reality*, 249–250. https://doi.org/10.1109/VR.2017.7892270

Haier, R. J., Karama, S., Leyba, L., & Jung, R. E. (2009). MRI assessment of cortical thickness and functional activity changes in adolescent girls following three months of practice on a visual-spatial task. *BMC Research Notes*, *2*. https://doi.org/10.1186/1756-0500-2-174

Hill, N. T. M., Mowszowski, L., Naismith, S. L., Chadwick, V. L., Valenzuela, M., & Lampit, A. (2017, April 1). Computerized cognitive training in older adults with mild cognitive impairment or dementia: A systematic review and meta-analysis. *American Journal of Psychiatry*, Vol. 174, pp. 329–340. https://doi.org/10.1176/appi.ajp.2016.16030360

Holm, A., Lukander, K., Korpela, J., Sallinen, M., & Müller, K. M. I. (2009). Estimating brain load from the EEG. *TheScientificWorldJournal*, *9*, 639–651. https://doi.org/10.1100/tsw.2009.83

Hudak, E. M., Bugos, J., Andel, R., Lister, J. J., Ji, M., & Edwards, J. D. (2019). Keys to staying sharp: A randomized clinical trial of piano training among older adults with and without mild cognitive impairment. *Contemporary Clinical Trials*, *84*. https://doi.org/10.1016/J.CCT.2019.06.003

Huygelier, H., Schraepen, B., van Ee, R., Vanden Abeele, V., & Gillebert, C. R. (2019). Acceptance of immersive head-mounted virtual reality in older adults. *Scientific Reports 2019 9:1*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-41200-6

Inanaga, K. (1998). Frontal midline theta rhythm and mental activity. *Psychiatry and Clinical Neurosciences*, *52*(6), 555–566. https://doi.org/10.1046/J.1440-1819.1998.00452.X

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833. https://doi.org/10.1073/PNAS.0801268105)

Jaeggi, Susanne M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0801268105

Jaeggi, Susanne M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10081–10086. https://doi.org/10.1073/pnas.1103228108

Jaeggi, Susanne M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory (Hove, England)*, *18*(4), 394–412. https://doi.org/10.1080/09658211003702171

Juan, M. C., & Pérez, D. (2009). Comparison of the Levels of Presence and Anxiety in an Acrophobic Environment Viewed via HMD or CAVE. *Presence: Teleoperators and Virtual Environments*, *18*(3), 232–248. https://doi.org/10.1162/PRES.18.3.232

Kable, J. W., Caulfield, M. K., Falcone, M., McConnell, M., Bernardo, L., Parthasarathi, T., … Lerman, C. (2017). No effect of commercial cognitive training on brain activity, choice behavior, or cognitive performance. *Journal of Neuroscience*, *37*(31), 7390–7402. https://doi.org/10.1523/JNEUROSCI.2832-16.2017

Kahana, M. J., Seelig, D., & Madsen, J. R. (2001). Theta returns. *Current Opinion in Neurobiology*, *11*(6), 739–744. https://doi.org/10.1016/S0959-4388(01)00278-1

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction*, *19*(5), 402–410. https://doi.org/10.1016/J.LEARNINSTRUC.2009.02.003

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(3), 615–622. https://doi.org/10.1037/0278-7393.33.3.615

Katz, B., Jaeggi, S., Buschkuehl, M., Stegman, A., Shah, P., Voss, M. W., … Cosman, J. (2014a). *Differential effect of motivational features on training improvements in school-based cognitive training*. https://doi.org/10.3389/fnhum.2014.00242

Katz, B., Jaeggi, S., Buschkuehl, M., Stegman, A., Shah, P., Voss, M. W., … Cosman, J. (2014b). Differential effect of motivational features on training improvements in school-based cognitive training. *Frontiers in Human Neuroscience*, *8*(1 APR), 242. https://doi.org/10.3389/fnhum.2014.00242

Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function. *JAMA*, *185*(12), 914–919. https://doi.org/10.1001/JAMA.1963.03060120024016

Kim, D., & Kim, D. J. (2012). Effect of screen size on multimedia vocabulary learning. *British Journal of Educational Technology*, *43*(1), 62–70. https://doi.org/10.1111/J.1467-8535.2010.01145.X

Kim, K., Rosenthal, M. Z., Zielinski, D. J., & Brady, R. (2014). Effects of virtual environment platforms on emotional responses. *Computer Methods and Programs in Biomedicine*, *113*(3), 882–893. https://doi.org/10.1016/J.CMPB.2013.12.024

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358.

https://doi.org/10.1037/H0043688

Klimesch, W. (1996). Memory processes, brain oscillations and EEG synchronization. *International Journal of Psychophysiology*, *24*(1–2), 61–100. https://doi.org/10.1016/S0167-8760(96)00057-8

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research. Brain Research Reviews*, *29*(2–3), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3

Kloosterman, N. A., Meindertsma, T., Hillebrand, A., van Dijk, B. W., Lamme, V. A. F., & Donner, T. H. (2015). Top-down modulation in human visual cortex predicts the stability of a perceptual illusion. *Journal of Neurophysiology*, *113*(4), 1063–1076. https://doi.org/10.1152/jn.00338.2014

Kober, S. E., Kurzmann, J., & Neuper, C. (2012). Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: An EEG study. *International Journal of Psychophysiology*, *83*(3), 365–374. https://doi.org/10.1016/j.ijpsycho.2011.12.003

Kothgassner, O. D., Goreis, A., Kafka, J. X., Eickels, R. L. Van, Plener, P. L., & Felnhofer, A. (2019). Virtual reality exposure therapy for posttraumatic stress disorder (PTSD): a meta-analysis. *European Journal of Psychotraumatology*, *10*(1), 1654782. https://doi.org/10.1080/20008198.2019.1654782

Kujawska, A., Kujawski, S., Gajos, M., Perkowski, R., Androsiuk, J., Jarecka, J., … Topka, W. (2016). NEW WAYS OF TREATMENT IN GERIATRICS AND PERCEPTION OF THOSE METHODS BY PATIENTS. *Journal of Education Culture and Society*, *7*(1), 75–91. https://doi.org/10.15503/JECS20161.75.91

Landers, R. N., & Landers, A. K. (2015). An Empirical Test of the Theory of Gamified Learning: The Effect of Leaderboards on Time-on-Task and Academic Performance. *Http://Dx.Doi.Org/10.1177/1046878114563662*, *45*(6), 769–785. https://doi.org/10.1177/1046878114563662

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, *32*(1), 47–56. https://doi.org/10.1145/333329.333344

Lawton, M. P., & Brody, E. M. (1969). Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. *The Gerontologist*, *9*(3_Part_1), 179–186. https://doi.org/10.1093/GERONT/9.3_PART_1.179

Legault, I., & Faubert, J. (2012). Perceptual-cognitive training improves biological motion perception: evidence for transferability of training in healthy aging. *Neuroreport*, *23*(8), 469–473. https://doi.org/10.1097/WNR.0B013E328353E48A

Lhemedu-Steinke, Q., Meixner, G., & Weber, M. (2018). Comparing VR Display with Conventional Displays for User Evaluation Experiences. *25th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018 - Proceedings*, 583–584.

https://doi.org/10.1109/VR.2018.8446076

Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience*, *12*(5), 549. https://doi.org/10.1038/NN.2296

Linares, R., Borella, E., Teresa Lechuga, M., Carretti, B., & Pelegrina, S. (2019). Nearest transfer effects of working memory training: A comparison of two programs focused on working memory updating. *PLoS ONE*, *14*(2). https://doi.org/10.1371/journal.pone.0211321

Lindstrom, M. (2011). You Love Your iPhone. Literally. Retrieved May 1, 2022, from The New York Times website: https://www.nytimes.com/2011/10/01/opinion/you-love-your-iphone-literally.html

Logie, R. H., Horne, M. J., & Pettit, L. D. (2014). When cognitive performance does not decline across the lifespan. *Working Memory and Ageing*, 37–63. https://doi.org/10.4324/9781315879840-8

Lumosity for mobile. (2021). Retrieved August 6, 2021, from https://www.lumosity.com/en/

Lumosity to Pay $2 Million to Settle FTC Deceptive Advertising Charges for Its "Brain Training" Program | Federal Trade Commission. (2016). Retrieved January 6, 2022, from https://www.ftc.gov/news-events/press-releases/2016/01/lumosity-pay-2-million-settle-ftc-deceptive-advertising-charges

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. *JMIR Serious Games*, *4*(2), e11. https://doi.org/10.2196/games.5888

Lumsden, J., Skinner, A., Coyle, D., Lawrence, N., & Munafo, M. (2017). Attrition from Web-Based Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques. *Journal of Medical Internet Research*, *19*(11). https://doi.org/10.2196/JMIR.8473

Lumsden, J., Skinner, A., Woods, A. T., Lawrence, N. S., & Munafò, M. (2016). The effects of gamelike features and test location on cognitive test performance and participant enjoyment. *PeerJ*, *2016*(7), e2184. https://doi.org/10.7717/peerj.2184

Mackworth, J. F. (1959). Paced memorizing in a continuous task. *Journal of Experimental Psychology*, *58*(3), 206–211. https://doi.org/10.1037/H0049090

Madrigal, A. (2008). Forget : Researchers Develop Software That Makes You Smarter | WIRED. Retrieved November 5, 2021, from https://www.wired.com/2008/04/forget-researchers-develop-software-that-makes-you-smarter/

Maglio, P., Wenger, M., & Copeland, A. (2008). Evidence for the role of self-priming in epistemic action: expertise and the effective use of memory. *Acta Psychologica*, *127*(1), 72–88. https://doi.org/10.1016/J.ACTPSY.2007.02.001

Marczewski, A. (2015). User Types. In *Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design* (1st editio, pp. 65–80). CreateSpace Independent Publishing Platform.

MarketsandMarkets Research. (2020). Cognitive Assessment and Training Market by Assessment Type ... Global Forecast to 2021. Retrieved July 9, 2020, from https://www.marketsandmarkets.com/Market-Reports/cognitive-assessment-market-1039.html

Martinez-Conde, S., Otero-Millan, J., & MacKnik, S. L. (2013). The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience 2013 14:2*, *14*(2), 83–96. https://doi.org/10.1038/nrn3405

Martínez-Navarro, J., Bigné, E., Guixeres, J., Alcañiz, M., & Torrecilla, C. (2019). The influence of virtual reality in e-commerce. *Journal of Business Research*, *100*, 475–482. https://doi.org/10.1016/J.JBUSRES.2018.10.054

Maruhn, P., Schneider, S., & Bengler, K. (2019). Measuring egocentric distance perception in virtual reality: Influence of methodologies, locomotion and translation gains. *PLOS ONE*, *14*(10), e0224651. https://doi.org/10.1371/journal.pone.0224651

McCarley, J. S. (2009). Effects of speed - Accuracy instructions on oculomotor scanning and target recognition in a simulated baggage X-ray screening task. *Ergonomics*, *52*(3), 325–333. https://doi.org/10.1080/00140130802376059

McMahan, T., Parberry, I., & Parsons, T. D. (2015). Evaluating Electroencephalography Engagement Indices During Video Game Play. *Technical Report LARC-2015-03 University of North Texas*, (Fdg), 4–8. Retrieved from https://www.researchgate.net/publication/277711650_Evaluating_Electroencephalography_Engagement_Indices_during_Video_Game_Play

Meiran, N., Dreisbach, G., & von Bastian, C. C. (2019). Mechanisms of working memory training: Insights from individual differences. *Intelligence*, *73*, 78–87. https://doi.org/10.1016/j.intell.2019.01.010

Mekler, E. D., Brühlmann, F., Brühlmann, B., Opwis, K., & Tuch, A. N. (2013). Do Points, Levels and Leaderboards Harm Intrinsic Motivation? An Empirical Analysis of Common Gamification Elements. *Gamification '13 – First International Conference on Gameful Design, Research, and Applications*. https://doi.org/10.1145/2583008.2583017

Miloff, A., Lindner, P., Dafgård, P., Deak, S., & Garke, M. (2019). Automated virtual reality exposure therapy for spider phobia vs. in-vivo one-session treatment: A randomized non-inferiority trial. *Behaviour Research and Therapy*, *118*, 130–140. https://doi.org/10.1016/J.BRAT.2019.04.004

Minns, S., Levihn-Coon, A., Carl, E., Smits, J. A. J., Miller, W., Howard, D., … Powers, M. B. (2019). Immersive 3D exposure-based treatment for spider fear: A randomized controlled trial. *Journal of Anxiety Disorders*, *61*, 37–44. https://doi.org/10.1016/J.JANXDIS.2018.12.003

Miranda, A. T., & Palmer, E. M. (2013). Intrinsic motivation and attentional capture from gamelike features in a visual search task. *Behavior Research Methods 2013 46:1*, *46*(1), 159–172. https://doi.org/10.3758/S13428-013-0357-7

Mishra, S., Kumar, A., Padmanabhan, P., & Gulyás, B. (2021). Neurophysiological Correlates of Cognition as Revealed by Virtual Reality: Delving the Brain with a Synergistic Approach. *Brain Sciences*, *11*(1), 1–18. https://doi.org/10.3390/brainsci11010051

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Mohammed, S., Flores, L., Deveau, J., Cohen Hoffing, R., Phung, C., M. Parlett, C., … R. Seitz, A. (2017). The Benefits and Challenges of Implementing Motivational Features to Boost Cognitive Training Outcome. *Journal of Cognitive Enhancement*, *1*(4), 491–507. https://doi.org/10.1007/s41465-017-0047-y

Monroe, T., & Carter, M. (2012). Using the Folstein Mini Mental State Exam (MMSE) to explore methodological issues in cognitive aging research. *European Journal of Ageing*, *9*(3), 265. https://doi.org/10.1007/S10433-012-0234-8

Morrison, A., & Chein, J. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*(1), 46–60. https://doi.org/10.3758/S13423-010-0034-0

Mühl, C., Jeunet, C., & Lotte, F. (2014). EEG-based workload estimation across affective contexts. *Frontiers in Neuroscience*, *8*(8 JUN), 114. https://doi.org/10.3389/fnins.2014.00114

Myers, R. S. (2000). Relation of useful field of view and other screening tests to on-road driving performance. *Perceptual and Motor Skills*, *91*(5), 279. https://doi.org/10.2466/PMS.91.5.279-290

Naumann, L., Schultze-Kraft, M., Dähne, S., & Blankertz, B. (2017). Prediction of difficulty levels in video games from ongoing EEG. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9961 LNCS*, 125–136. https://doi.org/10.1007/978-3-319-57753-1_11

Ninaus, M., Pereira, G., Stefitz, R., Prada, R., Paiva, A., Neuper, C., & Wood, G. (2015). Game elements improve performance in a working memory training task. *International Journal of Serious Games*, *2*(1). https://doi.org/10.17083/ijsg.v2i1.60

Normand, M. P. (2016). Less Is More: Psychologists Can Learn More by Studying Fewer People. *Frontiers in Psychology*, *7*(JUN), 934. https://doi.org/10.3389/fpsyg.2016.00934

Nouchi, R., Kobayashi, A., Nouchi, H., & Kawashima, R. (2019). Newly developed TV-based cognitive training games improve car driving skills, cognitive functions, and mood in healthy older adults: Evidence from a randomized controlled trial. *Frontiers in Aging Neuroscience*, *11*(MAY). https://doi.org/10.3389/FNAGI.2019.00099/FULL

Number of smartphones sold to end users worldwide from 2007 to 2021. (2021). Retrieved November 29, 2021, from https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., … Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–778. https://doi.org/10.1038/NATURE09042

Owsley, C., Ball, K., McGwin, G., Sloane, M. E., Roenker, D. L., White, M. F., & Overley, E. T. (1998). Visual processing impairment and risk of motor vehicle crash among older adults. *JAMA*, *279*(14), 1083–1088. https://doi.org/10.1001/JAMA.279.14.1083

Park, E., Han, J., Kim, K. J., Cho, Y., & Del Pobil, A. P. (2018). Effects of screen size in mobile learning over time. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3164541.3164625

Parsons, B., Magill, T., Boucher, A., Zhang, M., Zogbo, K., Bérubé, S., … Faubert, J. (2016). Enhancing Cognitive Function Using Perceptual-Cognitive Training. *Clinical EEG and Neuroscience*, *47*(1), 37–47. https://doi.org/10.1177/1550059414563746

Patsis, G., Sahli, H., Verhelst, W., & De Troyer, O. (2013). Evaluation of Attention Levels in a Tetris Game Using a Brain Computer Interface. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7899 LNCS*, 127–138. https://doi.org/10.1007/978-3-642-38844-6_11

Peng, Y.-H., Yu, C., Liu, S.-H., Wang, C.-W., Taele, P., Yu, N.-H., & Chen, M. Y. (2020). WalkingVibe: Reducing Virtual Reality Sickness and Improving Realism while Walking in VR using Unobtrusive Head-mounted Vibrotactile Feedback. *Proceedings of ACM CHI Conference on Human Factors in Computing Systems*, *20*, 1–12. https://doi.org/10.1145/3313831.3376847

Pierce, M. B., Young, P. A., & Doherty, S. M. (2017). Engagement and competence in VR and non-VR environments. *Proceedings of the Human Factors and Ergonomics Society*, *2017-October*, 2082–2085. https://doi.org/10.1177/1541931213602003

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*. https://doi.org/10.1016/0301-0511(95)05116-3

Prena, K., Reed, A., Weaver, A. J., & Newman, S. D. (2018). Game Mechanics Matter:

Differences in Video Game Conditions Influence Memory Performance. *Https://Doi.Org/10.1080/08824096.2018.1428545*, *35*(3), 222–231. https://doi.org/10.1080/08824096.2018.1428545

Puma, S., Matton, N., Paubel, P. V., Raufaste, É., & El-Yagoubi, R. (2018). Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology*, *123*, 111–120. https://doi.org/10.1016/J.IJPSYCHO.2017.10.004

Qiu, F., Pi, Y., Liu, K., Li, X., Zhang, J., & Wu, Y. (2018). Influence of sports expertise level on attention in multiple object tracking. *PeerJ*, *2018*(9), e5732. https://doi.org/10.7717/PEERJ.5732/SUPP-1

Raggi, A., Tasca, D., & Ferri, R. (2017, July 26). A brief essay on non-pharmacological treatment of Alzheimer's disease. *Reviews in the Neurosciences*, Vol. 28, pp. 587–597. https://doi.org/10.1515/revneuro-2017-0002

Rastegary, H., & Landy, F. J. (1993). The Interactions among Time Urgency, Uncertainty, and Time Pressure. In *Time Pressure and Stress in Human Judgment and Decision Making* (pp. 217–239). https://doi.org/10.1007/978-1-4757-6846-6_15

Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/COGP.1999.0735

Redlinger, E., Glas, B., & Rong, Y. (2021a). Enhanced Cognitive Training using Virtual Reality: Examining a Memory Task Modified for Use in Virtual Environments. *ACM International Conference Proceeding Series*, 1–8. https://doi.org/10.1145/3480433.3480435

Redlinger, E., Glas, B., & Rong, Y. (2021b). Impact of screen size on cognitive training task performance: An HMD study. *International Journal of Psychophysiology*, *166*, 166–173. https://doi.org/10.1016/j.ijpsycho.2021.06.003

Redlinger, E., & Shao, C. (2021). Comparing brain activity in virtual and non-virtual environments: A VR & EEG study. *Measurement: Sensors*, *18*, 100062. https://doi.org/10.1016/J.MEASEN.2021.100062

Redlinger, E., Shao, C., & Yagi, T. (2019). Quantifying Engagement in Virtual Environments - a VR & EEG Study. *The Papers of Technical Meeting on "Medical and Biological Engineering,"* 53–57. IEE JAPAN.

Remington, L. A. (2011). *Clinical Anatomy of the Visual System - 2nd Edition* (2nd ed.). Retrieved from https://www.elsevier.com/books/clinical-anatomy-of-the-visual-system/9780750674904

Ressler, K. J. (2010). Amygdala Activity, Fear, and Anxiety: Modulation by Stress. *Biological Psychiatry*, *67*(12), 1117. https://doi.org/10.1016/J.BIOPSYCH.2010.04.027

Richter, M., Friedrich, A., & Gendolla, G. H. E. (2008). Task difficulty effects on cardiac activity. *Psychophysiology*, *45*(5), 869–875. https://doi.org/10.1111/j.1469-8986.2008.00688.x

Romeas, T., Guldner, A., & Faubert, J. (2016). 3D-Multiple Object Tracking training task improves passing decision-making accuracy in soccer players. *Psychology of Sport and Exercise*, *22*, 1–9. https://doi.org/10.1016/J.PSYCHSPORT.2015.06.002

Ross, L. A., Webb, C. E., Whitaker, C., Hicks, J. M., Schmidt, E. L., Samimy, S., … Visscher, K. M. (2019). The Effects of Useful Field of View Training on Brain Activity and Connectivity. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *74*(7), 1152–1162. https://doi.org/10.1093/geronb/gby041

Rothbaum, B., Hodges, L., Kooper, R., Opdyke, D., Williford, J., & North, M. (1995). Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *The American Journal of Psychiatry*, *152*(4), 626–628. https://doi.org/10.1176/AJP.152.4.626

Rousset, T., Bourdin, C., Goulon, C., Monnoyer, J., & Vercher, J. L. (2018). Misperception of egocentric distances in virtual environments: More a question of training than a technological issue? *Displays*, *52*, 8–20. https://doi.org/10.1016/j.displa.2018.02.004

Ryan, R. M. (2016). The empirical study of human autonomy using self-determination theory. *Scientists Making a Difference: One Hundred Eminent Behavioral and Brain Scientists Talk about Their Most Important Contributions*, 311–315. https://doi.org/10.1017/CBO9781316422250.068

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, *35*(11), 991–1009. https://doi.org/10.1016/J.ERGON.2005.04.005

Sala, G., & Gobet, F. (2019, January 1). Cognitive Training Does Not Enhance General Cognition. *Trends in Cognitive Sciences*, Vol. 23, pp. 9–20. https://doi.org/10.1016/j.tics.2018.10.004

Salthouse, T. A., Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, *36*(5), 464–486. https://doi.org/10.1016/J.INTELL.2007.10.003

Sanders, A. F. (1970). Some aspects of the selective process in the functional visual field. *Ergonomics*, *13*(1), 101–117. https://doi.org/10.1080/00140137008931124

Sato, N., & Yamaguchi, Y. (2008). Computational model-based prediction of human episodic memory performance based on eye movements. *IEICE Transactions on Communications*, *E91-B*(7), 2142–2143. https://doi.org/10.1093/ietcom/e91-b.7.2142

Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., … Birbaumer, N. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *European Journal of Neuroscience*, *22*(11), 2917–2926.

https://doi.org/10.1111/J.1460-9568.2005.04482.X

Schacter, D. L. (1977). EEG theta waves and psychological phenomena: A review and analysis. *Biological Psychology*, *5*(1), 47–82. https://doi.org/10.1016/0301-0511(77)90028-X

Scharinger, C., Soutschek, A., Schubert, T., & Gerjets, P. (2017). Comparison of the working memory load in N-back and working memory span tasks by means of EEG frequency band power and P300 amplitude. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00006

Schwarz, N., Diener, E., & Kahneman, D. (1999). *Well-being : the foundations of hedonic psychology*. 593.

Sekuler, R., & Ball, K. (1986). Visual localization: age and practice. *Journal of the Optical Society of America A*, *3*(6), 864. https://doi.org/10.1364/JOSAA.3.000864

Shelton, J. T., Elliott, E. M., Matthews, R. A., Hill, B. D., & Gouvier, W. D. (2010). The Relationships of Working Memory, Secondary Memory, and General Fluid Intelligence: Working Memory Is Special. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(3), 813–820. https://doi.org/10.1037/A0019046

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, *81*(1), 174–176. https://doi.org/10.1037/H0027448

Slobounov, S. M., Fukada, K., Simon, R., Rearick, M., & Ray, W. (2000). Neurophysiological and behavioral indices of time pressure effects on visuomotor task performance. *Cognitive Brain Research*, *9*(3), 287–298. https://doi.org/10.1016/S0926-6410(00)00009-4

Slobounov, Semyon M., Ray, W., Johnson, B., Slobounov, E., & Newell, K. M. (2015). Modulation of cortical activity in 2D versus 3D virtual reality environments: An EEG study. *International Journal of Psychophysiology*, *95*(3), 254–260. https://doi.org/10.1016/j.ijpsycho.2014.11.003

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin and Review*, *25*(6), 2083–2101. https://doi.org/10.3758/s13423-018-1451-8

Souders, D. J., Boot, W. R., Blocker, K., Vitale, T., Roque, N. A., & Charness, N. (2017). Evidence for Narrow Transfer after Short-Term Cognitive Training in Older Adults. *Frontiers in Aging Neuroscience*, *0*(FEB), 41. https://doi.org/10.3389/FNAGI.2017.00041

Stalvey, B. T., Owsley, C., Sloane, M. E., & Ball, K. (1999). The life space questionnaire: A measure of the extent of mobility of older adults. *Journal of Applied Gerontology*, *18*(4), 460–478. https://doi.org/10.1177/073346489901800404

Sterman, M. B., & Mann, C. A. (1995). Concepts and applications of EEG analysis in aviation performance evaluation. *Biological Psychology*, *40*(1–2), 115–130. https://doi.org/10.1016/0301-0511(95)05101-5

Stojanoski, B., Lyons, K. M., Pearce, A. A. A., & Owen, A. M. (2018). Targeted training: Converging evidence against the transferable benefits of online brain training on cognitive function. *Neuropsychologia*, *117*, 541–550. https://doi.org/10.1016/j.neuropsychologia.2018.07.013

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/H0054651

Studer-Luethi, B., & Meier, B. (2020). Is Training with the N-Back Task More Effective Than with Other Tasks? N-Back vs. Dichotic Listening vs. Simple Listening. *Journal of Cognitive Enhancement*, 1–15. https://doi.org/10.1007/s41465-020-00202-3

Tiego, J., Testa, R., Bellgrove, M. A., Pantelis, C., & Whittle, S. (2018). A hierarchical model of inhibitory control. *Frontiers in Psychology*, *9*(AUG). https://doi.org/10.3389/fpsyg.2018.01339

Turunen, M., Hokkanen, L., Bäckman, L., Stigsdotter-Neely, A., Hänninen, T., Paajanen, T., … Ngandu, T. (2019). Computer-based cognitive training for older adults: Determinants of adherence. *PLOS ONE*, *14*(7), e0219541. https://doi.org/10.1371/journal.pone.0219541

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1016/J.COGPSYCH.2014.01.003

Van Leeuwen, M., Van Den Berg, S. M., Hoekstra, R. A., & Boomsma, D. I. (2006). *Endophenotypes for intelligence in children and adolescents*. https://doi.org/10.1016/j.intell.2006.09.008

Vartanian, O., Coady, L., & Blackler, K. (2017). 3D Multiple Object Tracking Boosts Working Memory Span: Implications for Cognitive Training in Military Populations. *APA*, *28*(5), 353–360. https://doi.org/10.1037/MIL0000125

Vermeir, J. F., White, M. J., Johnson, D., Crombez, G., & Van Ryckeghem, D. M. L. (2020). The Effects of Gamification on Computerized Cognitive Training: Systematic Review and Meta-Analysis. *JMIR Serious Games*, *8*(3), e18644. https://doi.org/10.2196/18644

VR - Explore - Google Trends. (2022). Retrieved January 6, 2022, from https://trends.google.com/trends/explore?cat=3&date=all&geo=JP&q=VR

Waiter, G. D., Deary, I. J., Staff, R. T., Murray, A. D., Fox, H. C., Starr, J. M., & Whalley, L. J. (2009). Exploring possible neural mechanisms of intelligence differences using processing speed and working memory tasks: An fMRI study. *Intelligence*, *37*(2), 199–206. https://doi.org/10.1016/J.INTELL.2008.09.008

Wang, N., Qian, T., Zhuo, Q., & Gao, X. (2010). Discrimination between idle and work states in BCI based on SSVEP. *Proceedings - 2nd IEEE International Conference on Advanced Computer Control, ICACC 2010*, *4*, 355–358. https://doi.org/10.1109/ICACC.2010.5486907

Wilson, G. F., & Russell, C. A. (2003). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors*, *45*(4), 635–644. Retrieved from https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=00187208&v=2.1&it=r&id=GALE%7CA114169679&sid=googleScholar&linkaccess=fulltext

Wolinsky, F. D., Vander Weg, M. W., Howren, M. B., Jones, M. P., & Dotson, M. M. (2013). A randomized controlled trial of cognitive training using a visual speed of processing intervention in middle aged and older adults. *PloS One*, *8*(5), e61624. https://doi.org/10.1371/journal.pone.0061624

Wolinsky, F. D., Weg, M. W. Vander, Howren, M. B., Jones, M. P., & Dotson, M. M. (2013). A Randomized Controlled Trial of Cognitive Training Using a Visual Speed of Processing Intervention in Middle Aged and Older Adults. *PLoS ONE*, *8*(5). https://doi.org/10.1371/JOURNAL.PONE.0061624

Wolpaw, J. R., & Wolpaw, E. W. (2012). Brain-Computer Interfaces: Principles and Practice. *Brain-Computer Interfaces: Principles and Practice*, 1–424. https://doi.org/10.1093/ACPROF:OSO/9780195388855.001.0001

Wright, R. A. (2008). Refining the Prediction of Effort: Brehm's Distinction between Potential Motivation and Motivation Intensity. *Social and Personality Psychology Compass*, *2*(2), 682–701. https://doi.org/10.1111/j.1751-9004.2008.00093.x

Yamada, F. (1998). Frontal midline theta rhythm and eyeblinking activity during a VDT task and a video game: Useful tools for psychophysiology in ergonomics. *Ergonomics*, *41*(5), 678–688. https://doi.org/10.1080/001401398186847

Yun, S. J., Kang, M.-G., Yang, D., Choi, Y., Kim, H., Oh, B.-M., & Seo, H. G. (2020). Cognitive Training Using Fully Immersive, Enriched Environment Virtual Reality for Patients With Mild Cognitive Impairment and Mild Dementia: Feasibility and Usability Study. *JMIR Serious Games*, *8*(4). https://doi.org/10.2196/18127

Zahan, S. (2017). Removing EOG artifacts from EEG signal using noise-assisted multivariate empirical mode decomposition. *ICECTE 2016 - 2nd International Conference on Electrical, Computer and Telecommunication Engineering*. https://doi.org/10.1109/ICECTE.2016.7879634

Zając-Lamparska, L., Wiłkość-Dębczyńska, M., Wojciechowski, A., Podhorecka, M., Polak-Szabela, A., Warchoł, Ł., … Izdebski, P. (2019). Effects of virtual reality-based cognitive training in older adults living without and with mild dementia: a pretest–posttest design pilot study. *BMC Research Notes 2019 12:1*, *12*(1), 1–8. https://doi.org/10.1186/S13104-019-4810-2

Zargari Marandi, R., Madeleine, P., Omland, O., Vuillerme, N., & Samani, A. (2018). Reliability of Oculometrics during a Mentally Demanding Task in Young and Old Adults. *IEEE Access*, *6*, 17500–17517. https://doi.org/10.1109/ACCESS.2018.2819211

Zheng, W., Vialatte, F. B., Adibpour, P., Chen, C., Gaume, A., & Dreyfus, G. (2013). Effect of stimulus size and shape on steady-state visually evoked potentials for brain-computer interface optimization. *IJCCI 2013 - Proceedings of the 5th International Joint Conference on Computational Intelligence*, 574–577. https://doi.org/10.5220/0004667705740577

APPENDICES

A.1 Theta and Alpha Response During Periods of Rest and Cognitive Activity

Sample spectral activity for rest periods and cognitively active periods (visual n-back task) for a typical subject are displayed in figure 1 below. Note the increases in alpha activity (dotted red line) and corresponding decreases in theta activity (dotted blue line) during the rest periods, compared with the active periods (Fig 1).

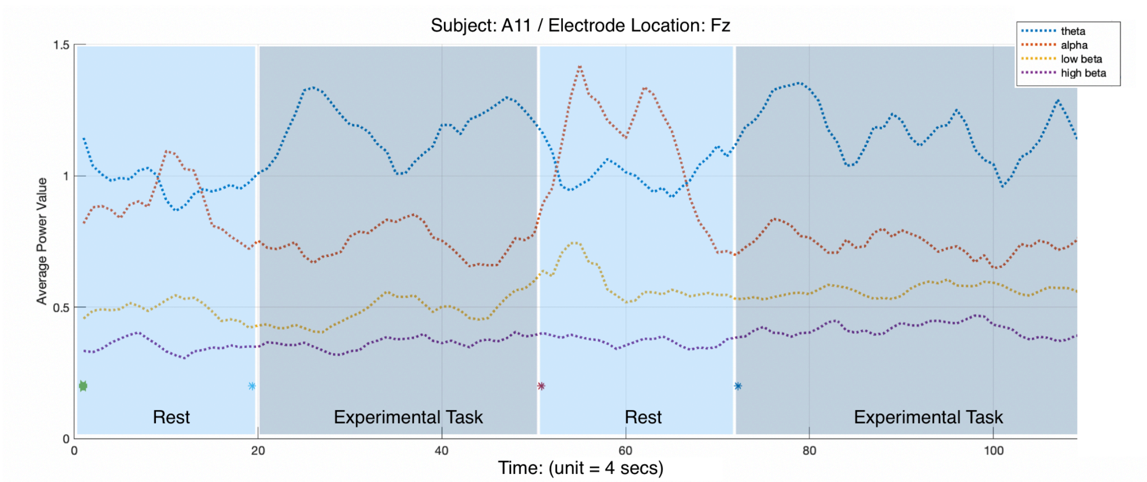


Figure 1: *Sample spectral activity over time for a typical subject. The light-blue areas denote rest periods while the grey shaded areas demarcate subject engagement with the cognitive task.*

A.2 Validating the Motivational Intensity Model

According to the *Motivational Intensity Model*, the moment the subject realizes the futility of continuing to pursue a cognitive task is accompanied by a dramatic decrease in task engagement. In figure 1 below, the subject realizes they cannot maintain sufficient performance to prevent blocks from reaching the top of the window at about the 27 second point (A). A dramatic drop in theta activity in the frontal area can be seen in the EEG data stream during the following 5 second window (B). The visual state of the task at the moment of cognitive capitulation is approximated in inset block C.

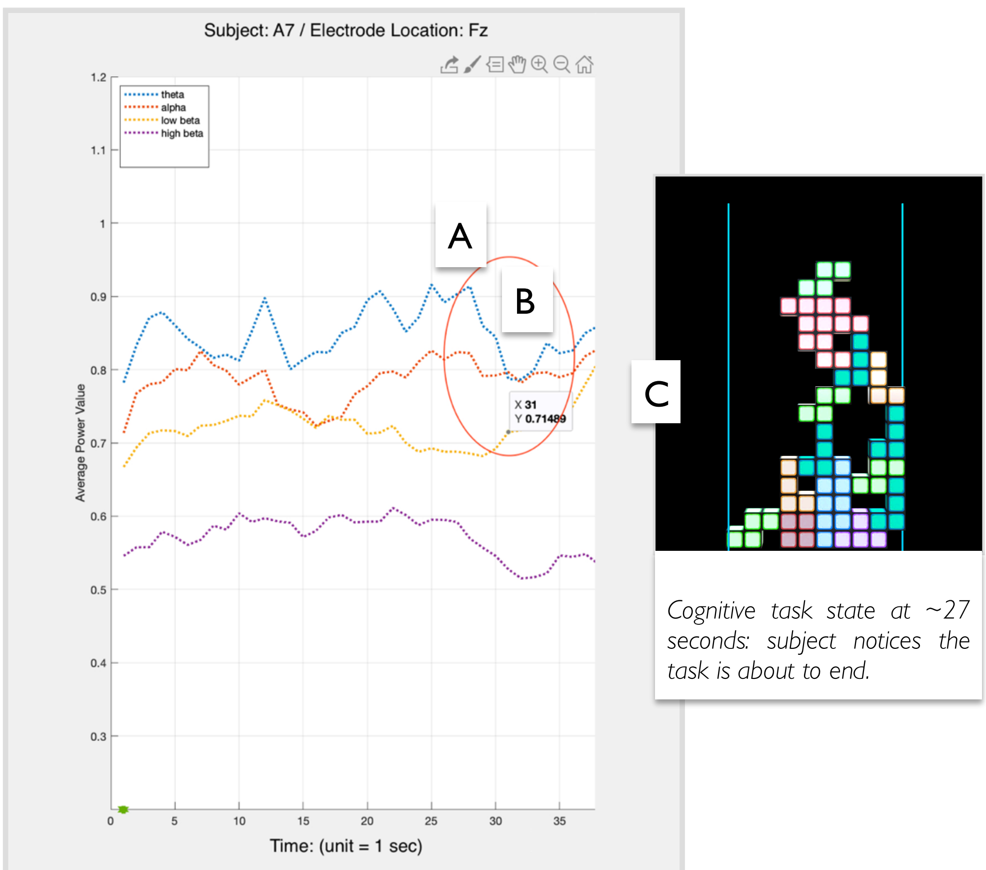


***Figure 1**: EEG data plot highlighting spectral power over time. Overall cognitive activity separated into discreet frequency ranges using fast Fourier transforms.*

A.3 Comparing EEG Data Impact of Various Task Input Methods

Since muscle movements have the potential to introduce noise into the EEG data, a trial study (n=4) was conducted to determine the potential impact of various user input modalities. The data indicated that none of the methods tested had a statistically meaningful effect (Fig. 1).

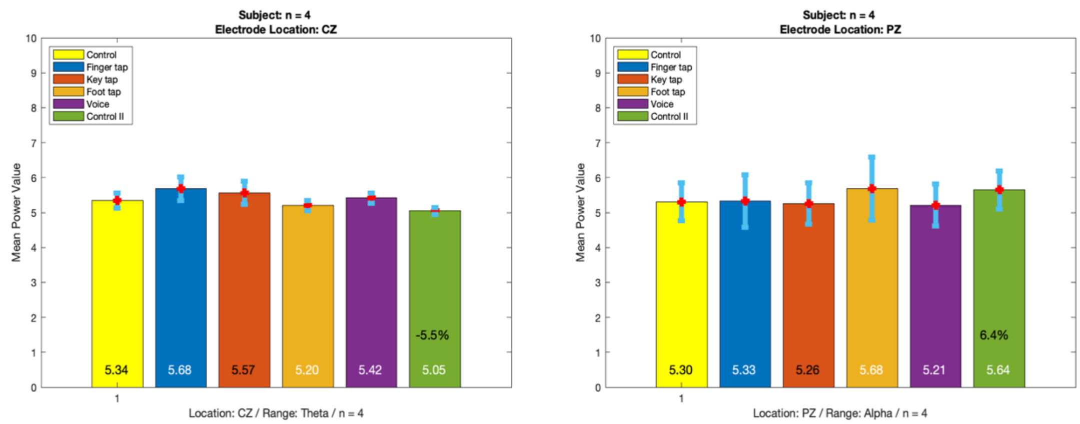


Figure 1: *Average theta and alpha range EEG power values for two active control conditions and the following experimental conditions: finger tap, keyboard tap, foot trigger tap, voice command. No statistically meaningful effect detected.*

The task employed for this experiment was a visual attention task. Against a black backdrop, arrows appeared on a computer screen in one of the four cardinal directions. When the arrow pointed down, the subject was instructed to respond with either a finger tap on a touchpad, a keypress, a foot tap, a voiced nonsense syllable "Ge-" or simply acknowledge the down arrow silently to themselves (active control condition). In each trial, the stimulus was presented for 2000 ms, and the experiment comprised 30 trials. The size of the stimulus on the screen corresponded to 2° of visual angle (Fig 2).

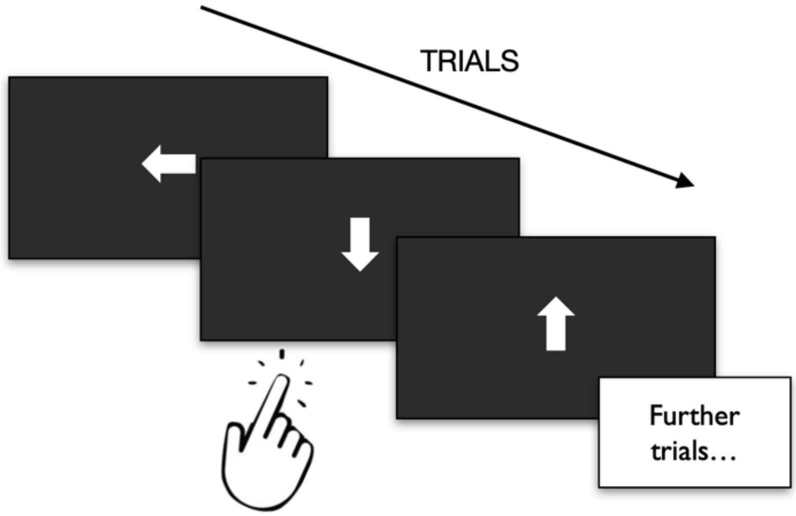


Figure 2: *Experimental task to assess the EEG impact of various subject response modalities. The appearance of a down arrow required the user to indicate acknowledgment in the requested modality.*

A.4 Identifying EOG Events in EEG Data-stream

While simultaneously recording EOG data, eye-blink events were initially identified using video documentation (Fig. 1).

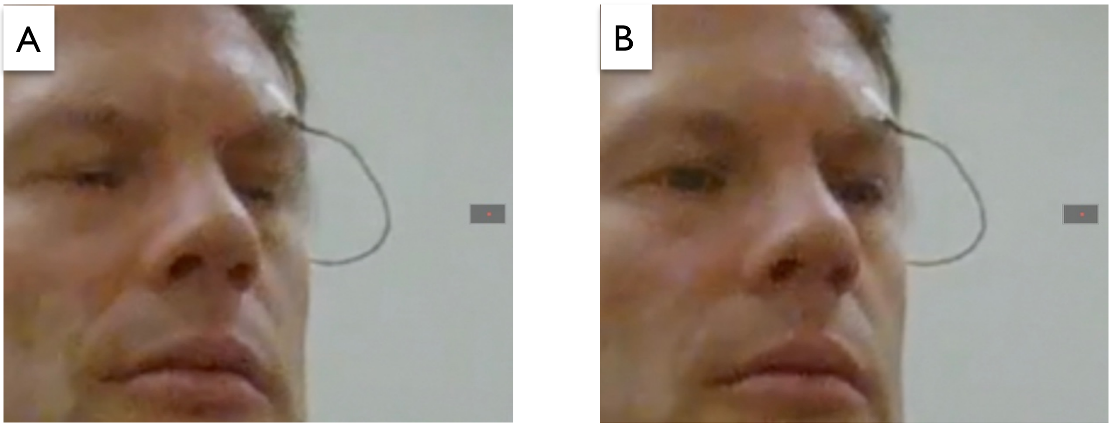


Figure 1*: Blink (A) and non-blink (B) events were first identified using video documentation and the corresponding time stamps noted.*

EEG data epochs corresponding to confirmed eye-blink events were then inspected, and an event "profile" was created. This enabled a MatLab algorithm to identify and classify subsequent EOG events in the data stream automatically (Fig. 2).

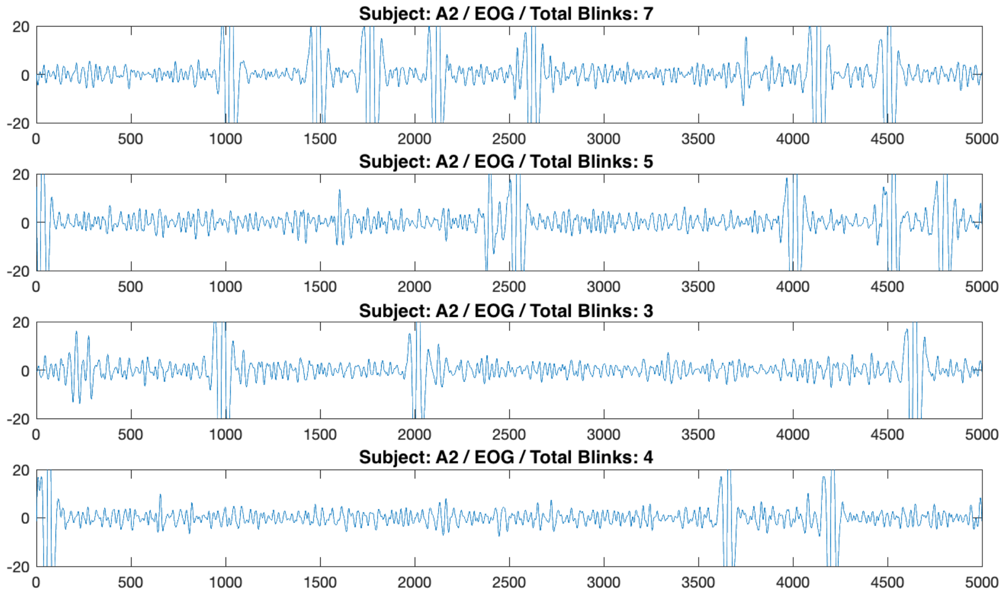


***Figure 2**: EEG data for a typical subject showing eye-blink artifacts being automatically classified and compiled.*

A.5 Ocular Anatomy

The highest density of cone cells, which are responsible for visual acuity, is found in the macular area of the retina (Fig. 1). Figures 2 and 3 show the average diameters and visual angles of the various parts of the macula.

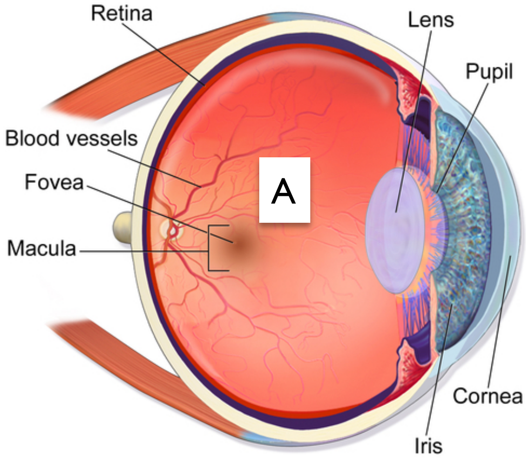


Figure 1: *The macula is the oval-shaped pigmented area near the center of the retina of the human eye (A).*

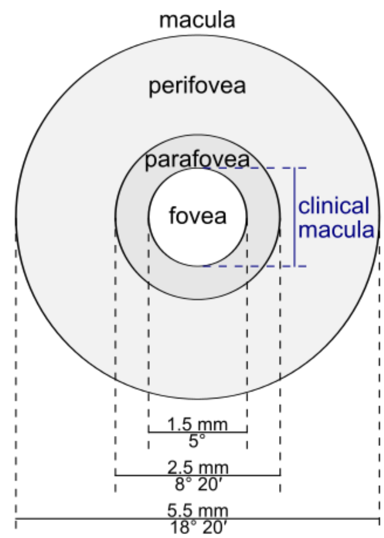


Figure 2: *Average diameters of various parts of the macular area. Total combined visual angle = 18° 20'*

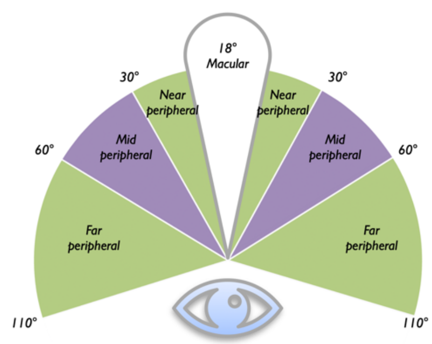


Figure 3: *Designations and approximate horizontal view angles from the central to the far peripheral areas.*

Data source: L.A. Remington, Clinical Anatomy of the Visual System - 2nd Edition, Elsevier Health Sciences, pp. 83-84, 2011.

Image sources: https://en.wikipedia.org/wiki/Macula_of_retina

A.6 Comparing Results from Three Protocol Iterations

Three protocol iterations were conducted for experiment one. Results from the initial protocol iteration (Fig. 1) and the second iteration (Fig. 2) revealed extremely similar increases in theta power for the VR condition at electrode Fz. All iterations: n=12.

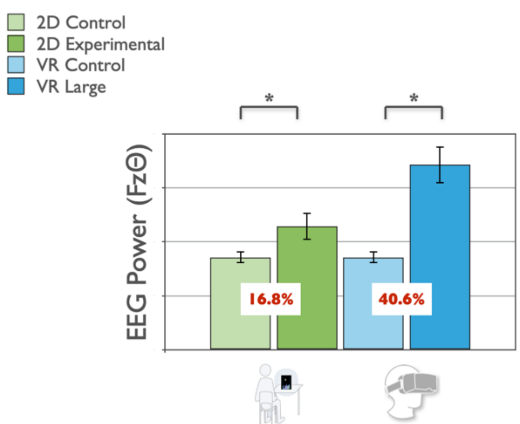


Figure 1: Initial protocol iteration; bright room during 2D experimental condition.

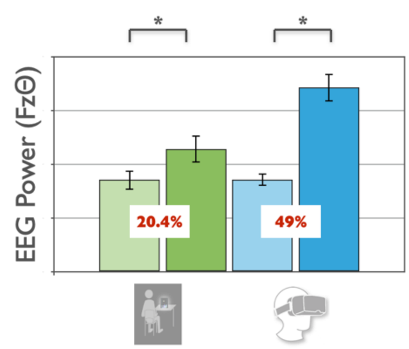


Figure 2: Protocol iteration two; all conditions conducted in a darkened space.

Protocol iteration three also showed similar trends between VR and non-VR conditions. In addition, the EEG power for the smaller stimulus condition was roughly 5% lower, a statistically significant difference (Fig. 3).

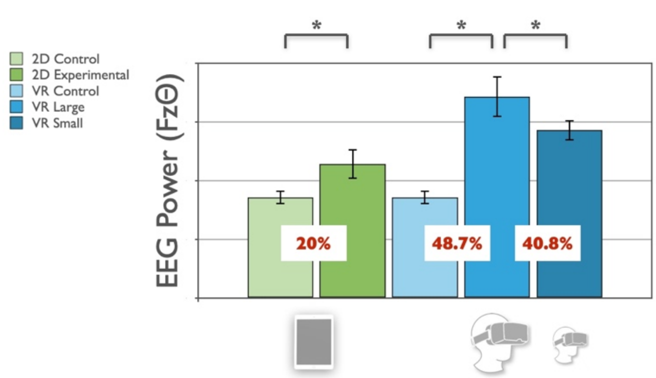


Figure 3: Protocol iteration three; smaller VR stimulus resulted in significantly less frontal theta power.

Additional electrode data for protocol 2 is presented below (Fig. 4).

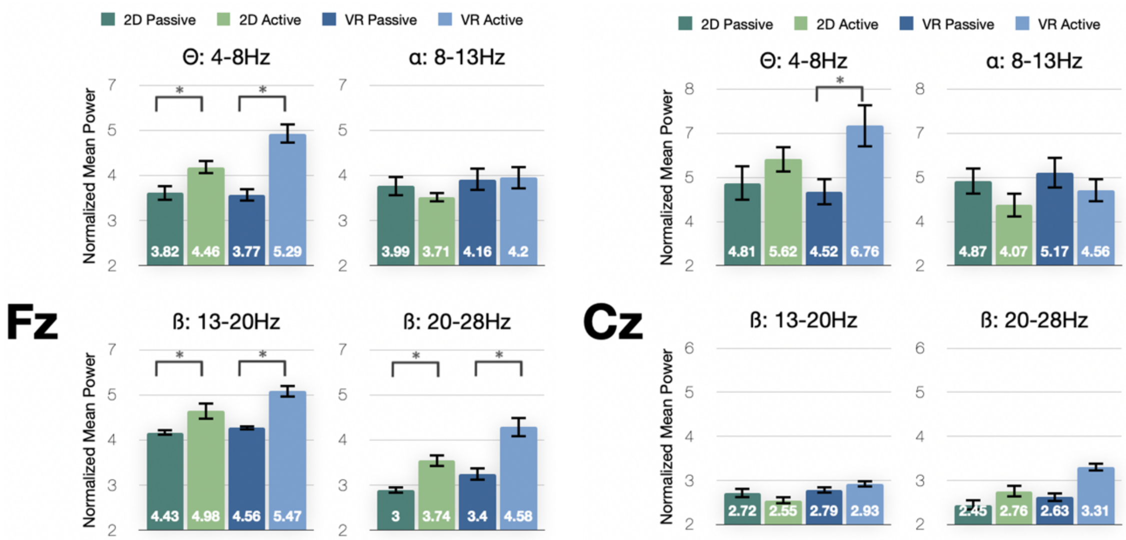


***Figure 4**: Spectral power by condition and frequency at EEG locations FZ and CZ; n=12; SE; statistical significance calculated with Wilcoxon signed-rank test (* = p < .05, ** = p < .01)*

A.7 Impact of Screen Size on Cognitive Training Task Performance / EEG Data

EEG power plots and post-hoc significance calculations for all electrodes and experimental conditions are presented in figure 1.

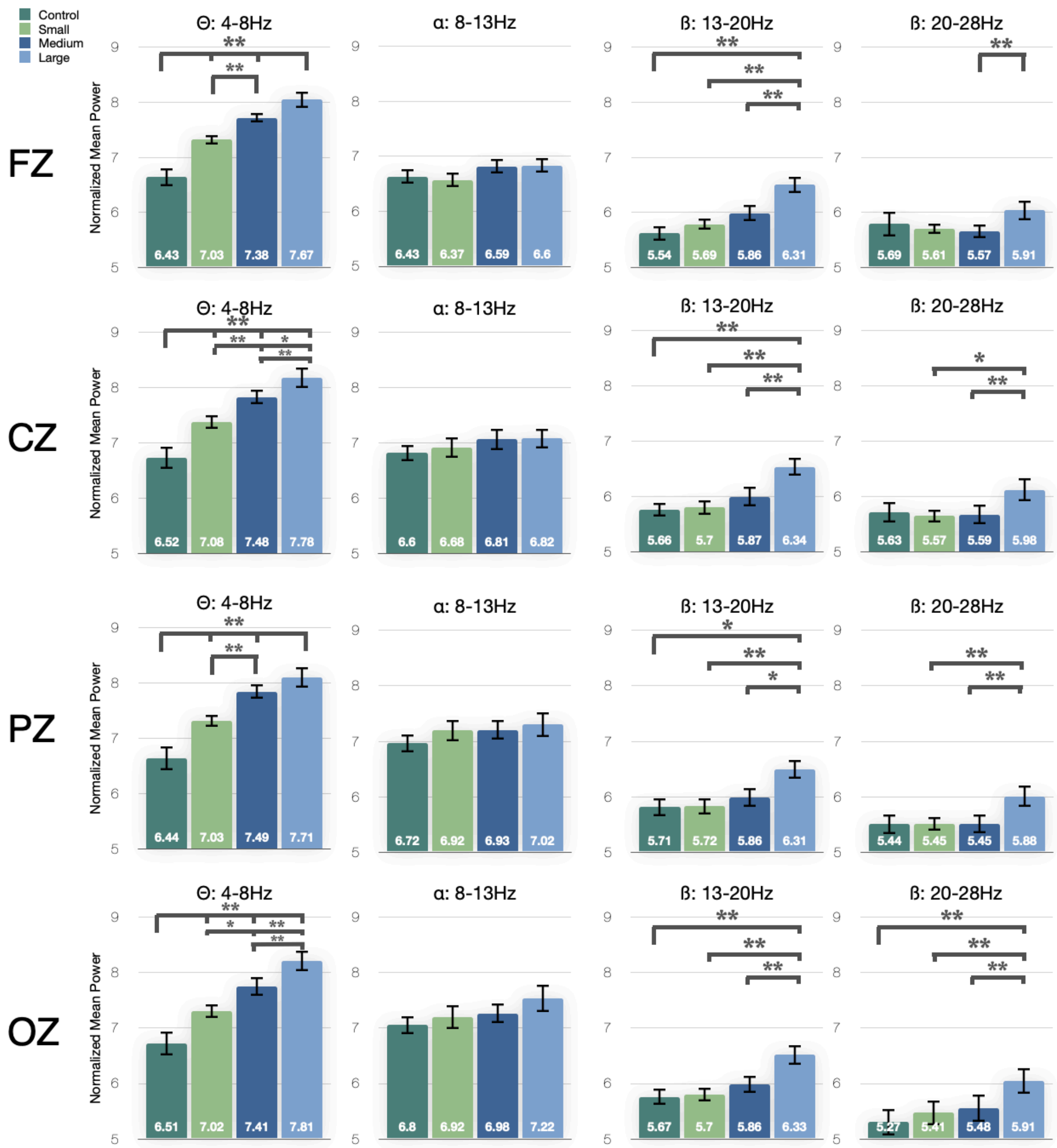


***Figure 1**: Spectral power by stimulus size (control, small, medium, large) and frequency at EEG locations FZ, CZ, PZ & OZ; n=20; SE; statistical significance calculated with Wilcoxon signed-rank test (* = p < .05, ** = p < .01)*

A.8 Impact of Screen Size on Cognitive Training Task Performance / Sample Correlation Data

Sample correlation data comparing theta power at Fz with *reaction time* and *accuracy*. Each point represents a single subject (Fig. 1). Spearman correlations (non-parametric) were not significant.

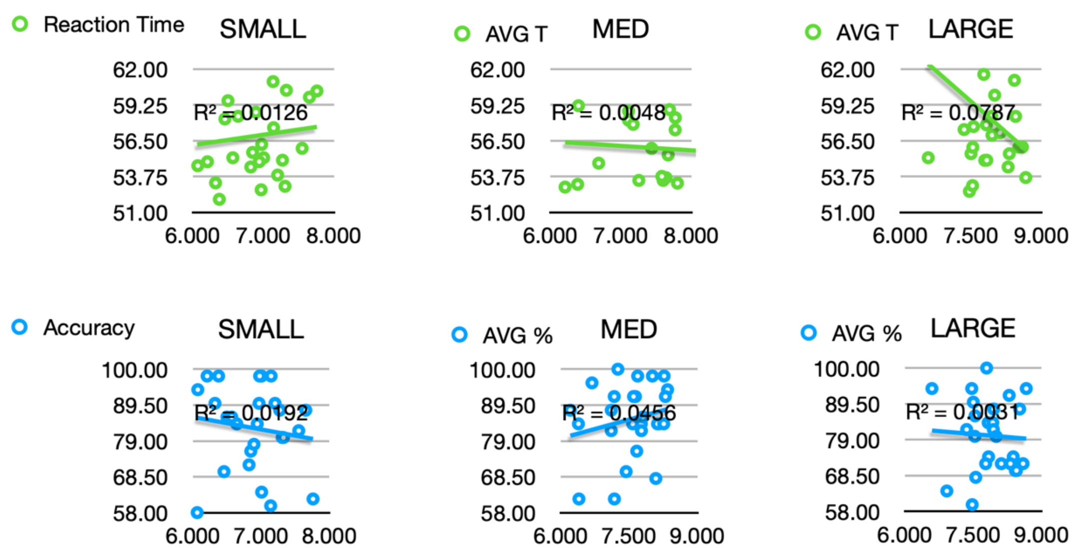


Figure 1: *Scatter-plot of all experimental conditions showing average theta (4-8Hz) power at EEG position Fz, correlated with reaction time and accuracy. X-axis is normalized EEG power and Y-axis is performance. Correlations are not significant.*

A.9 Impact of Game-like Features on Cognitive Performance in a Visual Memory Task / EEG Data

EEG power plots and post-hoc significance calculations for all electrodes and experimental conditions (Fig. 1).

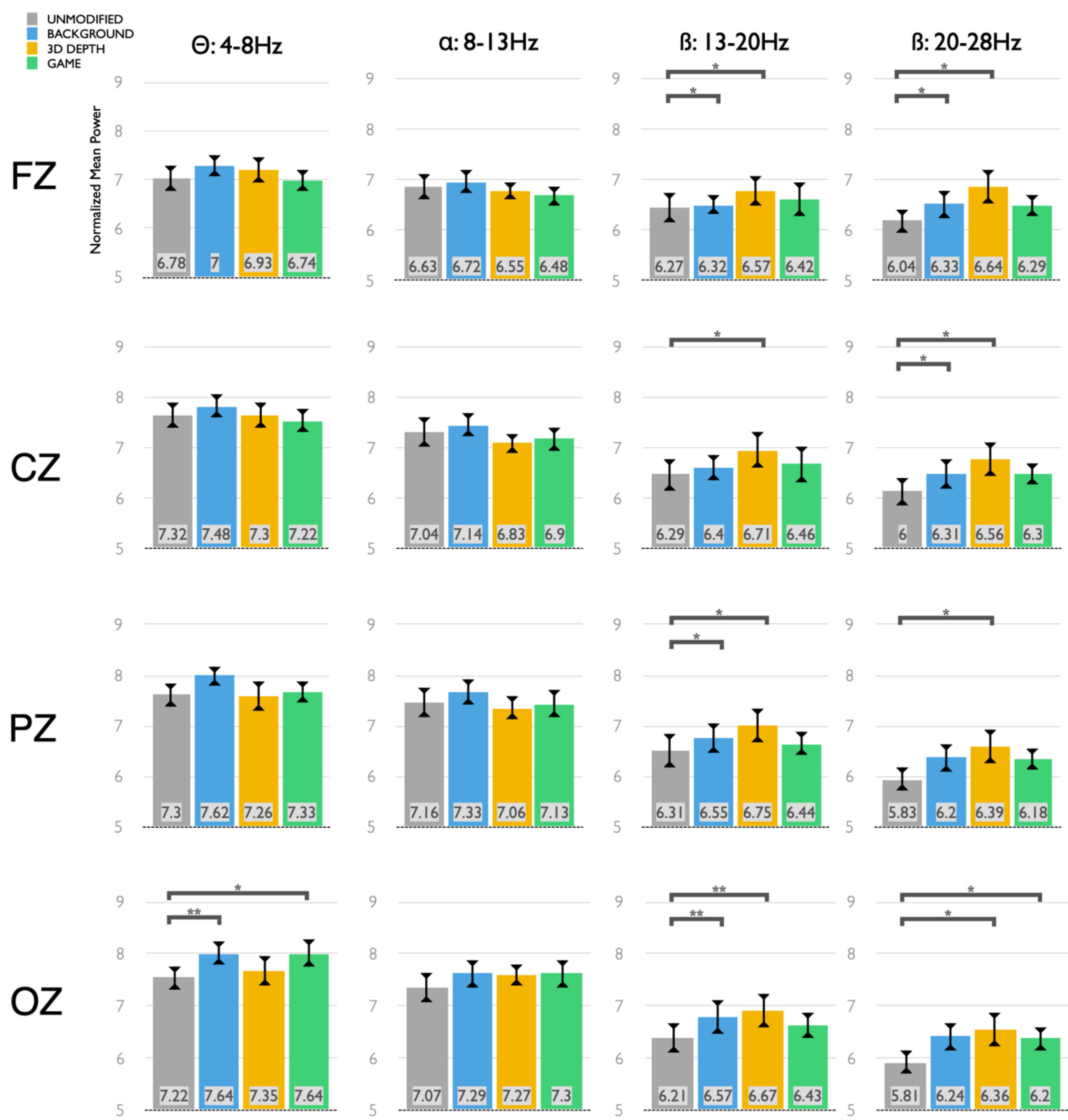


Figure 1*: Spectral power by condition and frequency at EEG locations FZ, CZ, PZ & OZ; n=20, SE; statistical significance calculated with Wilcoxon Signed-Rank test (* = p < .05, ** = p < .01)*