

REPORT DOCUMENTATION PAGE

1. REPORT DATE Dec. 9, 2023		2. REPORT TYPE Final		3. DATES COVERED	
				START DATE Feb. 1, 2021	END DATE Sep. 10, 2023
4. TITLE AND SUBTITLE Relationship between Brain Measurement of Pilot Mental Workload during Rotorcraft Dynamic Interface Tests and Subjective Rating Scales (DIPES)					
5a. CONTRACT NUMBER		5b. GRANT NUMBER N004212110003		5c. PROGRAM ELEMENT NUMBER	
5d. PROJECT NUMBER		5e. TASK NUMBER		5f. WORK UNIT NUMBER	
6. AUTHOR(S) Bradley Hatfield, Roberto Celi, Hyuk Oh, Rodolphe Gentili, Kyle Jaquess, Kyle Pietro					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Maryland, College Park Office of Research Administration 3112 Lee Building, 7809 Regents Drive College Park, MD 20742-5141					8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Naval Air Warfare Center Aircraft Division (NAWCAD), Procurement Group, 21983 Bundy Road, Bldg. 441, Patuxent River, MD 20670				10. SPONSOR/MONITOR'S ACRONYM(S) NAWCAD	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION A. Approved for public release: distribution unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The proposed study addresses the area of Handling Qualities (HQ) and focuses on the problem of assessing HQ levels for rotorcraft pilots performing maritime-specific tasks such as shipboard-related tasks or ADS-33 Mission-Task-Elements (MTEs). We aim to 1) explain the variance of Cooper-Harper Rating (CHR) estimates by regressing them on objective, quantitative metrics obtained using multiple psychophysiological biomarkers, and 2) examine the relationship between CHR and psychophysiological and behavioral performance-related measures of workload, and explore differences between test pilots and fleet pilots. This investigation will improve our fundamental understanding of perceived handling qualities through an examination of the dynamic objective processes (i.e., physiological) that influence the final outcome (i.e., CHR).					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:				17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 96
a. REPORT	b. ABSTRACT	c. THIS PAGE			

RELATIONSHIP BETWEEN BRAIN MEASUREMENT OF
PILOT MENTAL WORKLOAD DURING ROTORCRAFT DYNAMIC
INTERFACE TESTS AND SUBJECTIVE RATING SCALES (DIPES)

DR. BRADLEY HATFIELD, PRINCIPAL INVESTIGATOR

DR. ROBERTO CELI, CO-PRINCIPAL INVESTIGATOR

DR. RODOLPHE GENTILI, CO-INVESTIGATOR

DR. KYLE JAQUESS, CO-INVESTIGATOR

DR. HYUK OH, CO-INVESTIGATOR

MR. KYLE PIETRO, GRADUATE STUDENT INVESTIGATOR

The University of Maryland

College Park, Maryland

Department of Kinesiology

Department of Aerospace Engineering

2023

EXECUTIVE REPORT

A: A comparison of actual accomplishments with the goals and objectives established for the award, the findings of the investigator, or both;

B: Reasons why established goals were not met, if appropriate; and

C: Other pertinent information.

A:

- Test pilots were not available for testing during the Period of Performance; our work was conducted on Naval Aviators and NROTC MIDN
- This project was executed at the University of Maryland and the United States Naval Academy

B:

- Test pilots were not available for testing during the Period of Performance
- Research team communicated with NAWCAD POC Matt Rhinehart and he approved our executed project:
- Our current study includes the usage of the Bowtie display (as opposed to Deck Interface Pilot Effort Scale); also approved by NAWCAD POC (Klyde et al., [2021](#))
- NAWCAD POC approved usage of Naval Aviators for study population
- Our team made several visits to USNTPS as an investment of effort to establish possible testing at this location
- Period of Performance was altered due to COVID-19 human-subjects testing restrictions

C:

- Our team is still collecting data on NROTC, AFROTC, and Air Force Aviators
- We will publish these results and additional measurements in the future as part of our ongoing program of research
- Doctoral candidate Kyle Pietro is executing his doctoral dissertation related to this project

- Established collaboration with United States Naval Academy
- No-cost extension approved in March 2023
- Our team communicated findings at:
 - Pietro, K. F., Oh, H., Shaver, A. A., Hatfield, B. D. (2023) Psychophysiological Assessment of Workload Under Varying Degrees of Demand and Controllability: A Validation Study. [Poster presentation]. 2023 Mid-Atlantic American College of Sports Medicine Annual Meeting, Lancaster, PA, USA.
 - Pietro, K. F., Celi, R., Gentili, R. J., Hatfield, B. D., Oh, H. (2022). Relationship Between Brain Measurement of Pilot Mental Workload During Rotorcraft Dynamic Interface [Conference Presentation]. Naval Air Warfare Center Aircraft Division (NAWCAD) - UMD Seed Grant Review, California, MD. Center Aircraft Division (NAWCAD) - UMD Seed Grant Review, California, MD.
- This study was approved by University of Maryland IRB, Army HRPO, and United States Naval Academy IRB for human-subjects testing

ABSTRACT

This study addresses the area of Handling Qualities (HQ) and focuses on the problem of assessing HQ levels for pilots performing maritime-specific tasks such as shipboard-related tasks or ADS-33 MTEs. We aim to 1) explain the variance of CHR estimates by regressing them on objective, quantitative metrics obtained using multiple psychophysiological biomarkers, and 2) examine the relationship between CHR and psychophysiological and behavioral performance-related measures of workload between in Naval Aviators. This investigation will improve our fundamental understanding of perceived handling qualities through an examination of the dynamic objective processes (i.e., physiological) that influence the final outcome (i.e., CHR).

Table of Contents

Table of Contents	ii
List of Tables	iv
List of Figures	v
List of Abbreviations	viii
Section 1: Introduction	1
1.1 Mental Workload vs. Attentional Reserve	1
1.2 Mental Workload and its Measurement	4
1.3 Human-Vehicle Interaction	9
1.4 Efficiency of Psychomotor Behavior	12
1.4.1 Refinement of Neural Networks	14
1.5 Expenditure of Mental Effort	15
1.5.1 Arousal, Activation, and Effectiveness versus Efficiency . . .	17
1.6 Handling Qualities and Attention	21
1.7 Brain Dynamics and the Psychophysiological Assessment of Work- load	23
Section 2: Methods	29
2.1 Approach	29
2.2 Institutional Review Board Process	30
2.3 Participant Recruitment	30
2.4 Experimental Design	31
2.4.1 Procedure	31
2.4.2 Description of Task	34
2.5 Signal Processing	38
2.5.1 Wavelet-based Regression Analyses	40
2.6 Performance Scoring Criteria	42
2.7 Statistical Analysis	43
2.7.1 Regression Approach	44
Section 3: Results	46
3.1 Subjective Ratings	46
3.1.1 Cooper-Harper Rating Scale	46
3.1.2 Bedford Workload Scale	47

3.1.3	NASA-TLX Overall Workload	48
3.1.4	NASA-TLX Rating of Performance	49
3.2	Power Spectral Density Results	50
3.2.1	Delta Power Spectral Density	50
3.2.2	Theta Power Spectral Density	51
3.2.3	Alpha Power Spectral Density	52
3.2.4	Beta Power Spectral Density	52
3.2.5	$\frac{\theta}{\alpha}$ Ratio	53
3.3	Performance	54
3.3.1	Single-axis	55
3.3.2	Multi-axis	56
3.4	Regressions	58
3.4.1	CHR-Brain Dynamics	58
3.4.2	CHR-Behavioral Performance	62
3.4.3	Brain Dynamics-Behavioral Performance	63
3.4.4	Wavelet-based Regression	67
Section 4: Summary of Findings		68
Appendix A:Supplemental Figures		72
A.1	Questionnaires and Survey Tools	72
References		82

List of Tables

1.1	<i>Frequency ranges of decomposed EEG signal, (Andreassi, 2006)</i> . . .	23
1.2	<i>Psychophysiological measures in response to increased mental work-load</i>	26
2.1	Demographics	30
2.2	Performance standards for task objectives	36
2.3	EEG channels selected for wavelet analysis based on ROI	40
2.4	Timing of CHR rating	40
3.1		57

List of Figures

1.1	Relationship between mental workload and attentional reserve . . .	2
1.2	A hypothetical interpretation of the supply vs. demand relationship associated with mental workload and performance (de Waard, 1996; Young et al., 2015)	4
1.3	Key components involved in the manual control mode that are necessary to model manual control dynamics (Lone & Cooke, 2014). .	9
1.4	Handling qualities impacted by three modes of operator behavior (Klyde et al., 2021; McRuer & Krendel, 1974)	10
1.5	Variation of crossover model incremental latency with forcing function bandwidth (McRuer & Krendel, 1974)	12
1.6	Illustration of the relationship between the neurobiological response to challenge, the emergent psychological state, and the resultant psychological process (Hatfield & Hillman, 2001)	15
1.7	Attentional Resources effect on P300 component of the event-related potential. Arousal level determines processing capacity available for attention allocation to concurrent tasks. More difficult or multiple task demands reduce P300 amplitude and lengthen peak latency (Kahneman, 1973; Polich, 2007)	17
1.8	Inverted-U model of the effect of stress level on behavioral and physiological adaptability. The primary source of stress is task demand.(P. A. Hancock & Szalma, 2006; P. A. Hancock & Warm, 1989; Yerkes & Dodson, 1908)	19
1.9	CHR ratings as a function of spare attentional capacity (Hess, 1977; McDonnell, 1968)	21
2.1	Experimental setup	31
2.2	Single-loop primary task with secondary cross-coupled loading task (McRuer et al., 1990)	32
2.3	Visual representation for SoS timeline	33
2.4	Bowtie heads-up display (magenta polygon) which participants controlled during the compensatory SoS tracking tasks (green dot and line) (Klyde et al., 2020)	34
2.5	Requirements for small-amplitude pitch (roll) attitude changes - hover and low speed (Baskett, 2000)	35
2.6	Study protocol	37

3.1	Mean Cooper-Harper Rating by Condition and Level of Handling Quality.	46
3.2	Mean Bedford Workload Rating by Condition and Level of Handling Quality.	47
3.3	Mean weighted NASA-TLX Overall workload rating by Condition and Level of Handling Quality.	48
3.4	Mean NASA-TLX Performance rating by Condition and Level of Handling Quality. Higher ratings are inversely related to performance.	49
3.5	Mean $\frac{\theta}{\alpha}$ Ratio by Condition and Level of Handling Quality	53
3.6	Scalp topographical maps of $\frac{\theta}{\alpha}$ Ratio. The top row represents the Single-axis Condition, Levels 1, 2, 3, left to right, respectively. The bottom row represents the Multi-axis Condition, Levels 1, 2, 3, left to right, respectively.	54
3.7	Mean Single-axis Performance by Condition and Level of Handling Quality. The Y-axis, Percentage, represents the percent time that participants met the desired and adequate performance standard.	55
3.8	Mean Multi-axis Performance by Condition and Level of Handling Quality. The Y-axis, Percentage, represents the percent time that participants met the desired and adequate performance standard.	56
3.9	Actual vs. Expected Cooper-Harper Ratings. Y-axis represents the participants' Cooper-Harper ratings. X-axis represents the expected Cooper-Harper level rating.	58
3.10	Actual vs. Expected Bedford Workload Ratings. Y-axis represents the participants' Bedford Workload ratings. X-axis represents the expected Cooper-Harper level rating.	59
3.11A:	Normalized Cooper-Harper ratings regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. B: Normalized Cooper-Harper ratings regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. Both panels represent the Single-axis Condition. Scatterplots include five brain regions per participant.	60
3.12A:	Normalized Cooper-Harper ratings regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. B: Normalized Cooper-Harper ratings regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. Both panels represent the Multi-axis Condition. Scatterplots include five brain regions per participant.	61
3.13	Normalized Cooper-Harper ratings regressed on desired and adequate performance for Levels 1, 2, and 3. Both panels represent the Single-axis Condition.	62
3.14	Top row represents normalized Cooper-Harper ratings regressed on desired and adequate pitch performance for Levels 1, 2, and 3. Bottom row represents normalized Cooper-Harper ratings regressed on desired and adequate roll performance for HQR Levels 1, 2, and 3. All panels represent the Multi-axis Condition.	63

3.15A: Percentage desired performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. B: Percentage adequate performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. C: Percentage desired performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. D: Percentage adequate performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. All panels represent the Single-axis Condition. Scatterplots include five brain regions per participant.	64
3.16A: Percentage desired performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. B: Percentage adequate performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. C: Percentage desired performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. D: Percentage adequate performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. All panels represent the pitch component of the the Multi-axis Condition. Scatterplots include five brain regions per participant.	65
3.17A: Percentage desired performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. B: Percentage adequate performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. C: Percentage desired performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. D: Percentage adequate performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. All panels represent the Roll component of the the Multi-axis Condition. Scatterplots include five brain regions per participant.	66
A.1 NROTC Mean Bedford Workload rating by Condition and Level of Handling Quality.	73
A.2 NROTC Mean weighted NASA-TLX Overall workload rating by Condition and Level of Handling Quality.	74
A.3 Scalp topographical maps of NROTC participants for $\frac{\theta}{\alpha}$ Ratio. The top row represents the Single-axis Condition, Levels 1, 2, 3, left to right, respectively. The bottom row represents the Multi-axis Condition, Levels 1, 2, 3, left to right, respectively.	75
A.4 NROTC Mean $\frac{\theta}{\alpha}$ Ratio by Condition and Level of Handling Quality.	76
A.5 NROTC Single-axis mean desired and adequate performance by Condition and Level of Handling Quality.	77
A.6 NROTC Multi-axis mean desired and adequate performance by Condition and Level of Handling Quality.	78
A.7 The Cooper-Harper Rating Scale (Cooper & Harper, 1969)	79
A.8 The 'Bedford' Workload Scale (Roscoe & Ellis, 1990)	80
A.9 The National Aeronautics and Space Administration Task Load Index (Hart & Staveland, 1988)	81

List of Abbreviations

ADS-33E	Aeronautical Design Standard Performance Specification
	Handling Qualities Requirements for Military Rotorcraft
ANOVA	Analysis of Variance
BWS	Bedford Workload Rating Scale
CHR	Cooper-Harper Handling Qualities Rating Scale
CSD	Current Source Density
EEG	Electroencephalogram
EKG	Electrocardiogram
EMG	Electromyogram
FDR	False Discovery Rate
IIR	Infinite Impulse Response
HQ	Handling Qualities
IC	Independent Components
ICA	Independent Component Analysis
IRB	Institutional Review Board
MWL	Mental Workload
MTE	Mission Task Element
NASA-TLX	National Aeronautics and Space Administration Task Load Index
SoS	Sum of Sines

Section 1: Introduction

Mental effort is defined as “the aspect of cognitive load that refers to the cognitive capacity that is actually allocated to accommodate the demands imposed by the task; thus, it can be considered to reflect the actual cognitive load” (Paas et al., [2003](#), p.64). The term ‘handling qualities’ refers to the characteristics which determine the control nature and behavior of a system, which includes any property of the operator/system interaction in addition to the type and amount of control effort required within each control loop (McDonnell, [1968](#)).

1.1 Mental Workload vs. Attentional Reserve

Mental and physical workload are often referred to as analogous to one another in that they each are comprised of two components: stress (task demand) and strain (related impact on the individual) (Young et al., [2015](#)). While this comparison is extremely simplistic, it offers an attractive approach to empirically explain mental workload, which is otherwise a complex multidimensional construct (P. A. Hancock & Warm, [1989](#); Young et al., [2015](#)). As such, the joint organization of mental processes functioning in concert with coordi-

nated motor processes is characterized as psychomotor behavior. Psychomotor behavior is indicated by the level of work output relative to the mental and physical effort required to execute a task. Essentially, psychomotor behavior is broken into two separate branches: mental workload and attentional reserve, where mental workload represents the amount of mental and physical effort required to complete a task, and attentional reserve is simply “what is left over” (Miller et al., 2013; Miller et al., 2011). There is empirical evidence that the two constructs are inversely related (Jaquess et al., 2017). Historically, mental workload has been assessed within three main classifications: physiological, subjective, and task-performance based measures. Physiological measures are reliant on evidence that increased mental demand and arousal lead to an objectively measurable response within the body (Moray, 1979). Subjective measures (both unidimensional and multidimensional) consider the opinion and perception of mental workload within the individual based on their experience. Task-performance based measurements rely on the capacity of the participant

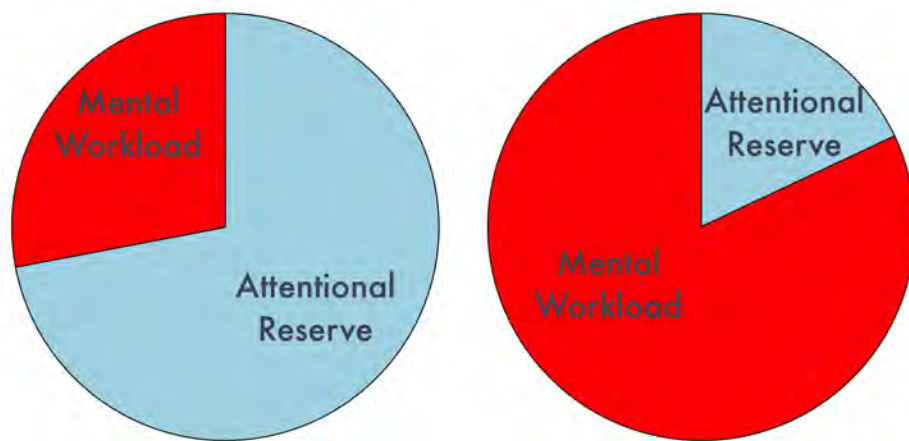


Figure 1.1: Relationship between mental workload and attentional reserve

to perform a task behaviorally, and they are scored by how well or poorly they perform at a given task load based on some criteria.

While each of these classifications have individual merits, their focused use over time has created a somewhat nebulous literature, which has made finding an accurate estimate of this multidimensional and abstract construct of mental workload increasingly difficult. Unfortunately, in many cases the researcher's definition of mental workload is so closely related to its method of measurement such that the results and conclusions that are drawn are simply not generalizable to other areas of research nor able to be applied in the field. While physiological measures have the benefit of being objective, they do not always occur in the same time scale, nor do they always occur as what might be predicted (i.e., increased stress and/or demand can lead to reduced mental workload (DiDomenico & Nussbaum, 2008; King & Schaefer, 2011)). On the other hand, subjective measures, while considering the individual's perception, can often be biased by the experience of the individual on the task as well as on the measure, itself. Task performance-based measurements can offer some insight into understanding an individual's workload, but failing in a task can increase perceptions of workload (P. A. Hancock, 1989), and also suffers from many of the same limitations as subjective measures. While each measurement modality has its own strengths and weaknesses, ultimately the sensitivity of each measurement type can vary based on the workload experienced by the operator (de Waard, 1996). **Thus, the optimal method for measuring mental workload is to employ a combination of techniques while considering the**

characteristics of the task, participant, and context of the measurement environment.

1.2 Mental Workload and its Measurement

The long lineage of the study of mental workload incorporates a considerable number of differing perspectives, some of which predate the formalization of the construct alone.

While the assessment of human capacities extends back into antiquity, research and interest in human mental workload assessment is much more recent. Interest in mental workload grew increasingly around World War II, when an urgent need sur-

rounding combat raised the issue of pre-eminent performance capacity (G. M. Hancock et al., 2021). During this time, the need to defeat the enemy drove engineers to make aircraft faster, deadlier, more fuel efficient, and perhaps most importantly, optimized for the flight crew with reduced errors. Historically, it has been challenging in the design, testing, and evaluation of new aircraft to

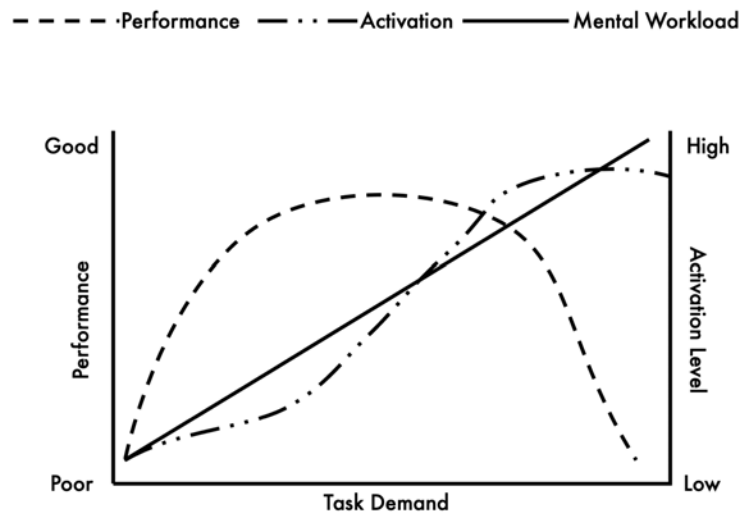


Figure 1.2: A hypothetical interpretation of the supply vs. demand relationship associated with mental workload and performance (de Waard, 1996; Young et al., 2015)

provide engineers with necessary information to improve and alter these unique flight “vessels”. However, It was not until the late 1960s, when this need grew to a point where two NASA researchers, Cooper and Harper ([1969](#)) conceived the numerical decision tree which allowed the experimental pilots and engineers to know how the operator felt about the aircraft they had flown.

One of the most widely used subjective ratings to measure pilot workload is the Cooper-Harper Rating Scale (CHR; Cooper & Harper, [1969](#); Harper Jr. & Cooper, [1986](#)). When employed by an expert pilot, the CHR yields an efficient and useful subjective rating for the assessment of operator workload and handling qualities (HQ). The tool provides a unidimensional scaling of operator workload based on a simple-to-follow flowchart and decision tree. The operator provides a categorical integer rating from 1-10, where a rating of 1 indicates minimum operator workload and effort in relation to the aircraft handling qualities (see Figure [A.7](#)). Thus, the CHR was designed to evaluate perceptual, cognitive, and physical effort resulting in a parsimonious, integrated assessment of operator workload and handling quality. Relatedly, these handling qualities ratings (HQR) ascertained through the employment of the CHR in test pilots became a critical component of the U.S. Army’s Handling Qualities standard, ADS-33E-PRF (Baskett, [2000](#); Blanken et al., [2008](#)); the subjective element for the assigned HQ of the HQ assessment process (Memon et al., [2022](#)). Accordingly, Cooper and Harper ([1969](#)) took great consideration in the efforts to design the CHR when they describe both ‘performance’ and ‘workload’:

When applied to handling qualities, the term “performance” alone is intended to mean the precision of aircraft control attained by the pilot, that is, the pilot-vehicle performance. The term “workload” is intended to convey the amount of effort and attention, both physical and mental, that the pilot must provide to attain a given level of performance. (p. 12)

Additionally, a critical component of the HQR process is the subjective interpretation of operator control compensation. Accordingly, Cooper and Harper (1969) note this in their description of compensation such that:

Pilot compensation as used in the [CHR] scale is intended to indicate that the pilot must increase his workload to improve aircraft performance. It relates to the pilot’s difficulty in completing a task with the precision required for that task. (p.13)

Therefore, Cooper and Harper (1969, p.13) define operator compensation as “the measure of additional pilot effort and attention required to maintain a given level of performance in the face of less favorable or deficient vehicle characteristics”. Here, the understanding that there is always some amount of workload given the necessary completion or participation in the task that is separate from compensation, is fundamental to awarding an HQR (Memon et al., 2022). It should be noted however, that while the operator’s workload is reflected in the HQR within the CHR, this HQR cannot be strictly interpreted as a measure of workload, but rather a reflection of the control compensation element (e.g.,

additional effort above the task load) to achieve a given level of performance (Memon et al., 2022). As the need for more generalized and diverse measures grew over time, the CHR has encouraged the further developments of similar unidimensional assessments like the ‘Bedford’ Workload Scale (BWS; Roscoe & Ellis, 1990), and the well known multidimensional assessment, the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988).

However, as previously mentioned, the usage of subjective ratings of workload as a primary outcome metric for workload has several underpinnings which limit the extrapolation of the results. One such issue is that it is nearly impossible to collect subjective ratings of workload in real-time. Typically, the participant will complete the subjective rating(s) following a trial and/or condition, depending on the experiment. In this way, the temporal resolution of the tool, or the capacity to capture the moment to moment changes in workload, becomes severely reduced such that it is possible the participant is only recalling their perceived workload during the final moments of the previous trial, rather than their perceived workload throughout the *entire* trial. Recently, Memon et al. (2022) have noted that there exists a debate as to the understanding of whether the HQR is, or should be, awarded for a particularly difficult local phase within the task or is an average of compensation/performance achieved throughout the task. Paul and Rhinehart (2020) investigated this issue based on a discussion with staff from the U.S. Naval Test Pilot School (USNTPS) and highlighted that “the pilots are not explicitly instructed to consider the peak or average workload over the total task, leaving an open question which could

contribute to both inter- and intra-pilot variability” (Memon et al., [2022](#); Paul & Rhinehart, [2020](#); Paul et al., [2023](#)). For example, during a steady hovering task, it would appear to a researcher that overall physical and mental workload for an operator would be low given that the pilot is only required to maintain a steady altitude and heading, while not traversing any distance. However, this has been shown to be one of the more challenging and difficult tasks for an operator to complete given the high demand on the pilot and the expected performance of the aircraft. Paul et al. ([2023](#)) have noted that the precision hover task was the most challenging of the five tasks in their experiment, so much so that two participants were unable to complete the trial. This is one such case where metrics such as workload of the inceptor or electromyogram (EMG) might reflect a lower workload, but subjective reports vary greatly. This problem ultimately clouds the sensitivity and precision of the tools being used and makes drawing conclusions extremely difficult for the researcher since the sensitivity of each type of measurement can vary depending on the perceived level of workload experienced by the operator (de Waard, [1996](#)). Alternatively, it can be constructive to employ a combination of subjective and objective measures when assessing mental workload, since no single measure provides a universal solution in examining operator workload (Charles & Nixon, [2019](#)). Importantly, the majority of the published studies report the results of a single physiological measure and its relationship with mental workload.

1.3 Human-Vehicle Interaction

Manual control dynamics, related to the control of a vehicle, or an aircraft in this case, requires several key components. Figure 1.3 illustrates the necessary components that are required when modeling human manual control dynamics. Generally, for the system to function or progress, an objective is required which is ultimately processed in the brain. The brain can then generate a control action in order to progress toward the objective. If the operator experiences a period of increased urgency or emergency (e.g., reduced handling qualities), it is generally thought that they will react based on recall of learned procedure/protocol or react based on instinct. As such, any resulting operator control action determines cues and gains to establish feedback control for the operator (Lone & Cooke, 2014). The operator control action is executed via the

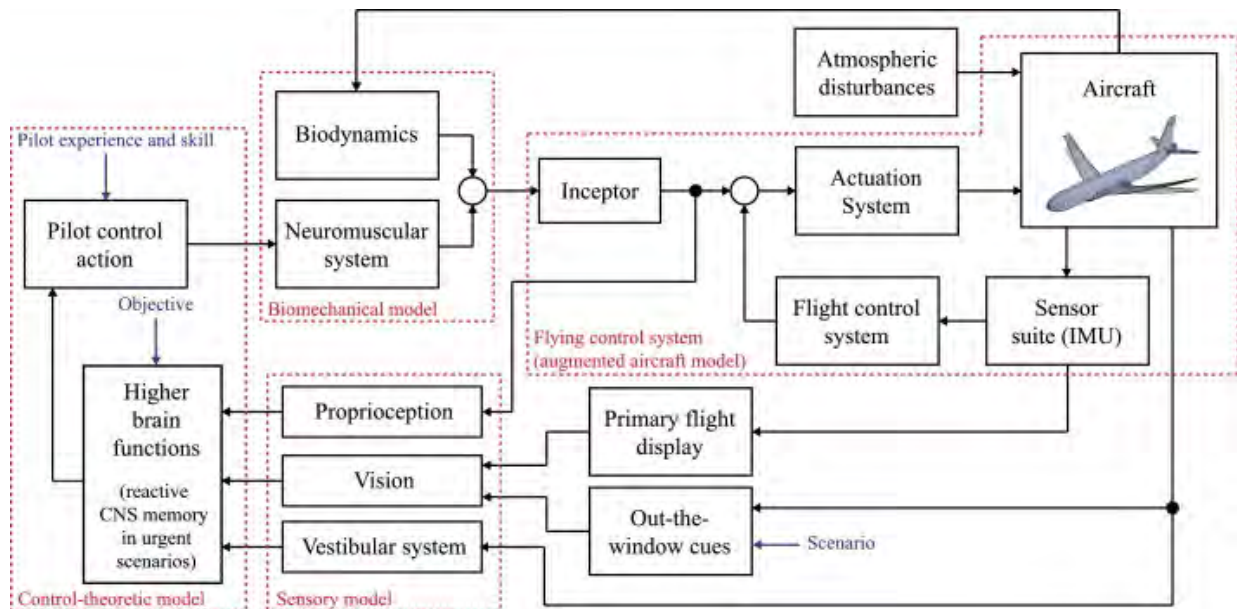


Figure 1.3: Key components involved in the manual control mode that are necessary to model manual control dynamics (Lone & Cooke, 2014).

neuromuscular system and ultimately is acted upon at the inceptor. Therefore, it is reasonable to assume that the operator control action occurs as a direct link in response to increased mental workload and is observable in the brain, muscle, and inceptor workload.

McRuer and Krendel (1974) defined three fundamental modes of dynamic behavior as shown in Figure 1.4 which essentially comprises the range of human operator control possibilities in an operator-vehicle system. As such, the operator compensation that is associated with controlling these systems will influence the handling qualities assessment. The metric which measures the frequency at which the operator closes the loop in compensatory tracking is defined as the crossover frequency (McRuer & Krendel, 1974). In such cases, a Sum of Sines (SoS) command signal may act as the forcing function and allows for the crossover frequency to be measured directly. The crossover frequency

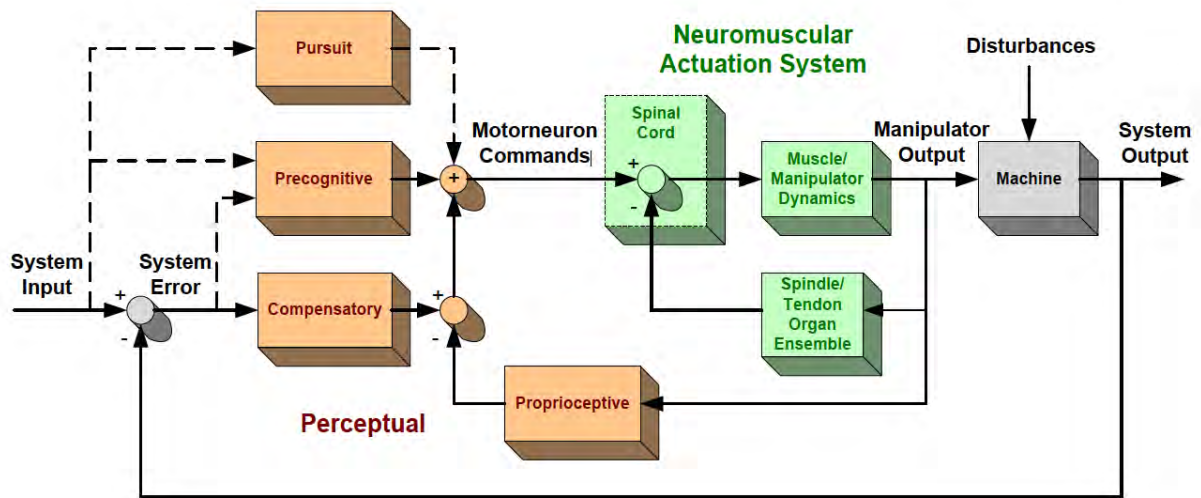


Figure 1.4: Handling qualities impacted by three modes of operator behavior (Klyde et al., 2021; McRuer & Krendel, 1974)

ω_c is utilized in the crossover model as seen in Figure 1.4 where at frequencies near the crossover frequency ω_c , the operator will adjust their characteristics such that the operator-vehicle system is represented as:

$$Y_{OL}(j\omega) = Y_p Y_c \doteq \frac{\omega_c e^{-j\omega\tau_e}}{j\omega} = \frac{m}{e} \quad (1.1)$$

where Y_p is the operator describing function, Y_c is the controlled element or transfer function, ω_c is the crossover frequency or the frequency at which $|Y_p Y_c| = 1$, and τ_e is the effective time delay within the operator. In this case, ω_c is equal to the loop gain and accounts for the operator's compensation for the controlled element gain (McRuer & Krendel, 1974). Here, both ω_c and τ_e are functions of the task variables. In the proposed experiment, the task variables and objectives will be defined in section 2.4. In order to manipulate level of handling qualities, the proposed experiment will adjust the forcing function bandwidth of the operator's input, such that while the SoS command signal will remain constant across all trials, the operator will perceive varying levels of handling qualities due to the level of responsiveness of the system. This is conceptually illustrated in Figure 1.5 where less experienced participants may perceive the manipulation as a delay in responsiveness.

According to McRuer and Jex (1967), once the crossover frequency is fixed, the primary adjustment within the operator to an increase in forcing function bandwidth is to increase their closed-loop damping ratio to avoid contributing appreciable tracking errors by tightening up their neuromuscular loop

(e.g., tighter grip on the control stick).

Thus, the delay perceived by the (inexperienced) operator in Figure 1.5 will effectively be reduced primarily by an increase in steady-state neuromuscular tension (McRuer & Krendel, 1974), which ultimately will be reflected as increased workload thus likely causing the operator greater strain over longer periods of time.

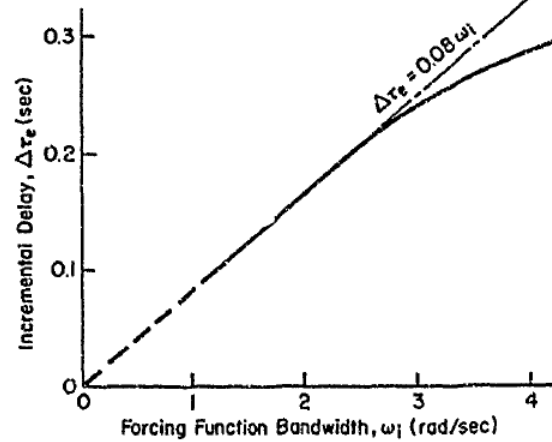


Figure 1.5: Variation of crossover model incremental latency with forcing function bandwidth (McRuer & Krendel, 1974)

This notion is similar to the results from Weinberg and Hunt (1976) where heightened levels of anxiety increased muscular activation beyond which was necessary for the task, resulting in an increase in errant throws. More specifically, they found that high-anxious individuals contracted their biceps and triceps for a longer period during a throwing task in a high-pressure condition compared to low-anxious individuals.

1.4 Efficiency of Psychomotor Behavior

Hatfield and Hillman (2001) have coined the term “Psychomotor Efficiency”, which postulates that an individual can refine their neural network to become more efficient such that a hallmark of expert level performance is a refinement of neural networks and psychomotor efficiency. In this state, neural communi-

cation to the muscle is refined so that only essential motor units required for a task are activated. More broadly, when operators accurately and successfully complete a task in a timely fashion with minimal mental and/or physical effort relative to the required task demand, they typify a state of performance efficiency (Hatfield et al., 2020; Hatfield & Kerick, 2007). The psychomotor efficiency hypothesis states that optimal performance is a result of economy of neural activation between **motor** and **non-motor** regions of the brain (Hatfield & Hillman, 2001; Rietschel, 2011). Higher efficiency of psychomotor behavior is indicative of an adaptive state, which suggests that the magnitude of effort required for a task is inversely related to attentional reserve; that is, decreased effort is associated with elevated reserve (Jaquess et al., 2017). Conversely, suboptimal performance is characterized by **inefficient** and **non-essential** neural activation of both motor and non-motor regions (Rietschel et al., 2012). Thus, non-essential neural activation creates inefficient attentional resource allocation (Hatfield & Hillman, 2001; Hatfield et al., 2020). Notably, increased availability of attentional reserve, resulting from mental and physical work reduction, allows the operator to respond to sudden increases in task demand, in addition to unexpected or surprise stimuli and events. However, a suboptimal response, resulting from scarcity or absence of attentional reserve, could result in operator error and even catastrophic consequences. Therefore, optimized efficiency of operator behavior is desirable as it contributes to safe and reliable manipulation of the system within a broad range of conditions.

1.4.1 Refinement of Neural Networks

In a classic study, Busk and Galbraith (1975) demonstrated that individuals are able to decrease overall levels of coherence through visuomotor practice. This result implies a refinement, or a honing of neural connectivity related to developing expertise in a task. Alternatively, some studies have utilized expert/novice comparisons to further explore this paradigm. One such example examined experts versus novices during a complex shooting task and showed that experts exhibited lower levels of functional connectivity compared to novices during the preparatory period before the shot (Deeny et al., 2009; Deeny et al., 2003). In other words, experts displayed less coherence, or cross-talk within the cortex, during aiming which suggests a refinement of their neural network architecture. These attributes illustrate that experts exhibit reduced involvement between cognitive and motor regions of the brain (Hatfield & Hillman, 2001). For less experienced individuals, higher levels of cognitive demand is hypothesized to consume the vast majority of neural resources, and attentional reserve will be much lower than that seen in experts for a given task (Gentili et al., 2018; Shaw et al., 2018). While an expert should be able to overcome increased levels of stress and challenge (e.g., reduced handling qualities) by relying on well-learned skills, there is evidence that increasing levels of task load and stress have the ability to regress an individual away from automaticity. These results symbolize decreased neuromuscular efficiency stemming from higher levels of mental workload and/or stress. An attenua-

tion in performance is dependent on whether the operator is able to regulate their state such that adverse effects of increased arousal do not interfere with task execution (Gaillard, 1993). **However, in such cases, the maintenance of performance comes at a cost (e.g., increased cognitive effort resulting in decreased efficiency).**

1.5 Expenditure of Mental Effort

The experience of the individual performing a task, in addition to their innate psychophysiological processes hold great consequence for the resultant quality of movement behavior. However, some individuals are able to maintain primary task execution in a stable fashion under stress and high demand (Hatfield et al., 2013), albeit their effective performance under in-

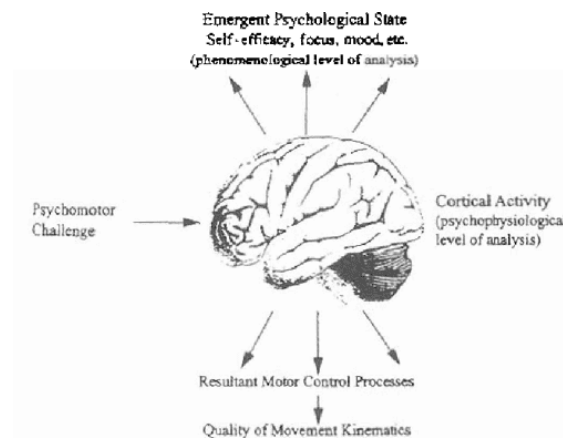


Figure 1.6: Illustration of the relationship between the neurobiological response to challenge, the emergent psychological state, and the resultant psychological process (Hatfield & Hillman, 2001)

creased task difficulty and stress is often accompanied by high levels of psychophysiological activation and subjective strain. Figure 1.6 illustrates conceptually the psychological state elicited by increased challenge and/or task demand. Indeed, mental and physical workload are directly related to the handling qualities of an aircraft such as in the control systems or usability and de-

sign. Increased effort and compensatory work that is required to achieve a desired flight operation, as related to the aircraft controls and handling qualities, will ultimately result in fatigue and reduced attentional reserve. Consequently, the Cognitive-Energetical Framework (Hockey, [1997](#)) argues that performance stability under demanding conditions is an active process that is likely under the control of the individual which requires the management of cognitive resources through the recruitment of cognitive effort. Essentially, when activation levels are too high such as in cases of threat, overload, and anxiety, the operator becomes distracted by task-irrelevant signals both from the environment as well as internally and the processing capacity available for the task is reduced (Gaillard, [1993](#)). Therefore, it is a reasonable endeavor to attempt to reduce an operator's workload while increasing attentional reserve so as to increase the likelihood of reacting to unexpected, stressful stimuli.

1.5.1 Arousal, Activation, and Effectiveness versus Efficiency

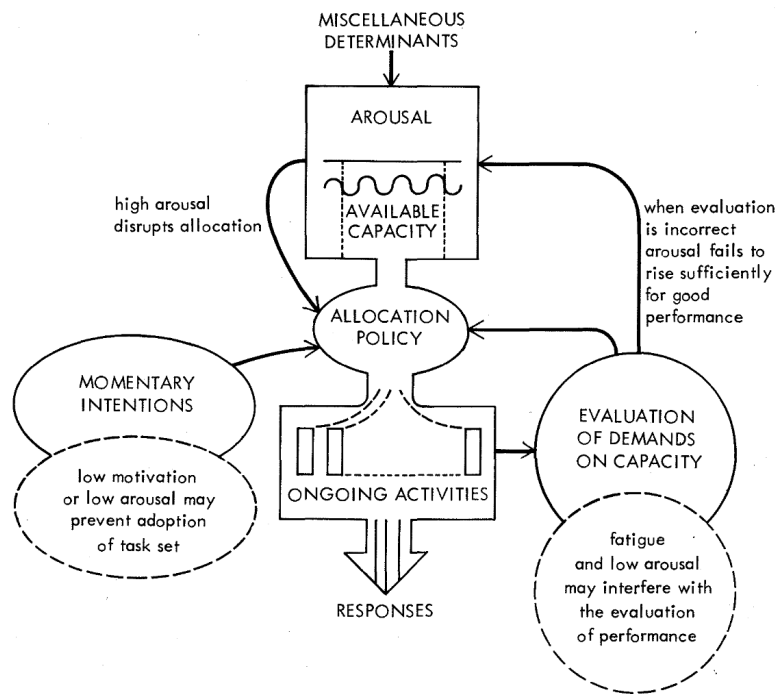


Figure 1.7: Attentional Resources effect on P300 component of the event-related potential. Arousal level determines processing capacity available for attention allocation to concurrent tasks. More difficult or multiple task demands reduce P300 amplitude and lengthen peak latency (Kahneman, 1973; Polich, 2007)

Increased task demand or decreased controllability of a system have the capacity to create anxiety and increased workload within the operator. Anxiety increases worry; worry consumes attentional capacity, impairs inhibitory control, and increases overall effort. As a consequence, the aggregate of this cascade causes attention to be reallocated to the detection of the source of the threat and deciding how to respond to that stimulus, which ultimately impairs attentional control. An operator's successful performance can conceptually be considered as a competition between **effectiveness** and **efficiency**. Effectiveness is essentially the quality of performance, whereas efficiency refers to the

relationship between the quality of performance divided by the investment of effort (Brown, 1964). Hence, improvement on a primary task with reduced effort or improvement on a secondary task with constant effectiveness on the primary task indicates enhanced efficiency (Mulder, 1986). Therefore, even when no performance decrement is observed, individuals may employ less efficient strategies as well as increased psychophysiological activation, effort, and fatigue (Hockey, 1997). Figures 1.7 and 1.8 illustrate a general framework regarding the relationship between arousal and attentional resource allocation. According to Kahneman (1973), the information processing system is modulated by arousal level, which governs the amount of attention that can be allocated for a given task. As a consequence, if arousal increases, the available capacity of attention which can be allocated to additional tasks and/or responding to changes in the environment becomes smaller. In his book, Kahneman (1973) illustrates this notion with a relatable example:

Imagine that you are conducting a conversation while driving an automobile through city traffic. As you prepare to turn into the traffic, you normally interrupt the conversation. Physiological measures would certainly indicate a surge of arousal at the same time, corresponding to the increased demands of the driving task. (p.4)

In such cases, the operator's effectiveness may be preserved due to heightened activation, but efficiency must be sacrificed. In the short term, this may appear as acceptable given that the primary objective is accomplished, but over longer

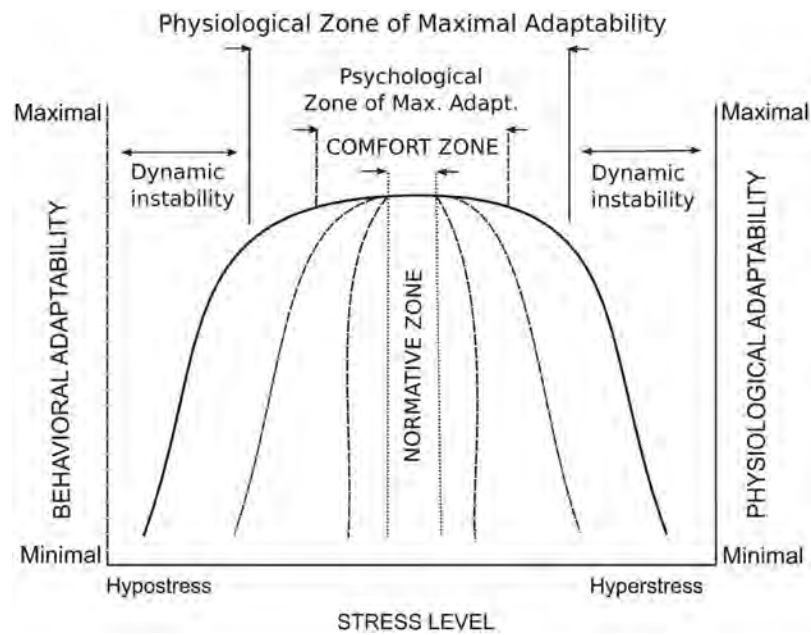


Figure 1.8: Inverted-U model of the effect of stress level on behavioral and physiological adaptability. The primary source of stress is task demand.(P. A. Hancock & Szalma, 2006; P. A. Hancock & Warm, 1989; Yerkes & Dodson, 1908)

periods of time, the reduced efficiency may appear as increased stress, fatigue, overload, or perhaps a catastrophic consequence.

If performance is to be maintained under conditions of high load, emotional and physiological sub-systems will attract costs to preserve regulatory activity (Hockey, 1997). As a result, these costs can be understood as an expenditure of mental resources. This expenditure can be interpreted as mental effort, which is identified with the action of maintaining a task activity in focal attention and, that effort provides an increase in overall resources to meet the prevailing demands (Kahneman, 1973). Task demands attract costs by way of increased cognitive activation (Dyke et al., 2014; Gentili et al., 2018; Hatfield et al., 2020) and sympathetic nervous system activation (Beatty, 1982).

Some results indicate that increased arousal narrows the range of cues that are processed by systematically reducing responsiveness to aspects of the environment which attract less attentional focus (Bacon, 1974). During periods of high demand, resources must constantly be divided between various components of the task, and in many cases, increased mental effort must be allocated when faced with suboptimal conditions (Gaillard, 1993), such as in the case of reduced handling qualities. In many cases, while an operator's performance may come without a visible decrement in performance, the measured increase in arousal and activation would indicate a reduction in cognitive efficiency; in which case, **the preservation of performance will come as a cost of compensatory cognitive effort, which cannot be maintained for longer periods of time.** For example, Padfield et al. (1994) formulated the control attack metric, A_η , to serve as an indirect measurement of operator workload. Memon et al. (2022) further explored this metric which is based on the rate and magnitude of the operator's control input along the lines of the attitude quickness parameter in order to quantify agility (Baskett, 2000). Originally utilized as a metric to reflect differences in operator control compensation between flight and simulation tests in test pilots, the attack parameter (A_η) is described as:

$$A_\eta = \frac{\dot{\eta}_{pk}}{\Delta\eta} \quad (1.2)$$

where (A_η) characterizes each discrete control input and is the ratio of the peak rate of control displacement ($\dot{\eta}_{pk}$) to the magnitude of change in the control dis-

placement ($\Delta\eta$) (Memon et al., 2022). More generally, larger attack values would indicate rapid-small control deflections and smaller attack values would indicate slower-large control deflections. While Padfield et al. (1994) and Memon et al. (2022) did not examine psychophysiological measurements, this metric, employed in addition to other objective and behavioral biomarkers could prove promising in the assessment of operator mental workload.

1.6 Handling Qualities and Attention

When workload becomes high, it is likely that operators will pay less attention to the task(s) at hand due to a narrowing of attention (Broadbent, 1957; Easterbrook, 1959). Conversely, workload that is sufficiently low creates boredom and allows the operator to disengage from the task. Either of these instances can ultimately end with catastrophic consequences. As such, it is necessary to design systems and develop training regimens and flight procedures that will reduce cognitive

demands to not exceed the capacities of the operator (Wilson, 2002) but that also maintain a level of mental workload that is not disengaging (Hebbar et al., 2021). Some authors relate subjective mental workload to the magni-

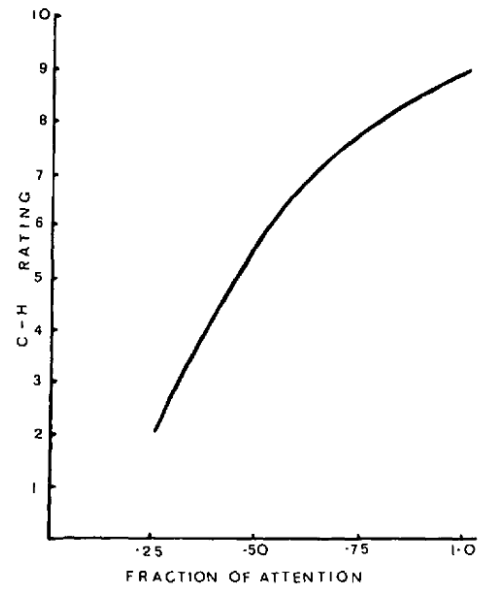


Figure 1.9: CHR ratings as a function of spare attentional capacity (Hess, 1977; McDonnell, 1968)

tude and/or frequency at which the operator perceives error. Taken with this notion, if the counterpart to mental workload is attentional reserve, such as in Figure 1.1, then the division of the attention devoted to a primary task and an increasingly unstable secondary task would be equal to one's spare capacity. McDonnell (1968) examined this within a population of pilots under the assumption that the maximum level of instability that the pilot could manage within the secondary axis while also maintaining performance within the primary axis should indicate the pilot's spare capacity. Although these results may be classified within the 'dual' task applications, their results lend credence to the study of increasing levels of mental workload and its relation to ratings of handling qualities. As an extension of this work, Hess (1977) modeled pilots during simulated flight tasks and found that as handling qualities degrade, workload increases (see Figure 1.9).

One factor that has the potential to reduce operator workload is improvement of the handling qualities of a system. As Klyde et al. (2021) have stated, "The overarching goal of the handling qualities discipline is to provide capability to allow the pilot to successfully execute a given mission set". In other words, the study of handling qualities allows for increased **ability** for operators to execute a task or mission. While the improvement of handling qualities is an expectation for military establishments, research in this area outside of defense entities is also an important endeavor since the equipment, machinery, and systems controlled by the operator must maintain a keen design, as well as the ability to sustain operation in austere, and all-weather conditions. However,

as the research and development of such systems progresses in addition to the refinement of operator capability, the number of “knobs and buttons” available to the operator can become overwhelming and thus less efficient (Klyde et al., 2021). **Furthermore, while the traditional method of subjective ratings of handling qualities have merit, they tend to incur increased costs paired with sub-optimal results. As such, one possible mode to satisfy the growing need for increased capability and reduced developmental costs and complexity is to pair objective biomarkers with subjective ratings.**

1.7 Brain Dynamics and the Psychophysiological Assessment of Workload

The electroencephalogram (EEG) provides an objective biomarker characterized by excellent temporal resolution to gain insight into the electrical activity in

Table 1.1: *Frequency ranges of decomposed EEG signal, (Andreassi, 2006)*

Frequency Band	Bandwidth (Hz)
Delta (δ)	0.5-3.5 Hz
Theta (θ)	4-7 Hz
Alpha (α)	8-13 Hz
Beta (β)	14-30 Hz
Gamma (γ)	30-100 Hz

the cerebral cortex which is an index of arousal and workload. Specifically, broadband alpha power indicates active inhibition of nonessential brain activity in regard to the task being performed (Jensen & Mazaheri, 2010). Moreover, desynchronized alpha activity reflects a state of high excitability, whereas synchronized alpha activity reflects a state of inhibition or comparatively low ex-

citability (Klimesch et al., 2007). As such, the central role of EEG alpha power is to selectively ignore irrelevant stimuli (Foxe & Snyder, 2011). Event-related desynchronization (ERD) of alpha reflects large populations of neurons are no longer oscillating in synchrony in response to the active processing of information (Klimesch et al., 2007) and with reference to a 'resting' condition, can be utilized as a metric for which synchrony is lost (Pfurtscheller & Aranibar, 1977). Therefore, the measurement of regional alpha activity can provide a metric for brain regions that are relatively inhibited during the processing of a task. Conversely, theta band activity, first described by Saul and Davis (1933) as "action currents" particularly over the frontal midline, plays an important integrational role in the organizing of task-related brain activity, working memory, and emotional arousal (Knyazev, 2007; Sauseng et al., 2010). Theta band power increases when experiencing difficult, attention-demanding tasks that require sustained concentration (Gevins, 1997; Gevins & Smith, 2003). When a challenging task-demand condition is compared to a resting condition, EEG alpha desynchronizes and EEG theta synchronizes (Klimesch et al., 2007). Additionally, Gevins and Smith (2003) have proposed utilizing the ratio of frontal theta power to parietal alpha power as a general task load index such that an increase in workload will be reflected by an increase in anterior frontal and frontal midline theta power and a reduction of parietal alpha power. In support of this, Hockey et al. (2009) identified that this ratio was associated with the "strain" that participants experienced while attempting to compensate for potential overload. Thus, it would be expected that the ratio of ratio of frontal

theta power to parietal alpha power in the TLI would increase in response to increased mental workload.

In addition to the theta and alpha bands of the EEG, the beta band is of particular interest given its relationship to sensorimotor function, decision making (Chen & Kwak, 2022; Kuc et al., 2021) and pronounced activity during steady muscular contraction (Engel & Fries, 2010). However, there exists no unifying hypothesis for beta activity in humans. Yet some results suggest that beta oscillatory activity may be the signature of an active process which promotes the current motor set while inhibiting processing for new movements (Engel & Fries, 2010; Pogosyan et al., 2009). As such, it appears that beta oscillatory activity may allow for more efficient processing of sensorimotor feedback required for monitoring the current state and recalibration of the sensorimotor system (Baker, 2007). It is reasonable to hypothesize that beta activity would be highly related to system handling qualities, which would signal their increased proprioceptive processing needed for the “recalibration” of the sensorimotor system. Baker (2007) have suggested that the interpretation of the proprioceptive state will likely be rendered more effectively if beta oscillatory activity is high in order to hold motor output in a more constant state. However, an increase in beta power, while strengthening the motor effects of sensory feedback, appears to come with reductions in the velocity of movements (Pogosyan et al., 2009). Additionally, the observation of beta band activity using time-frequency analysis may provide information to a critical moment when an operator makes their decision related to the handling qualities of a system. Recently, Kuc et

al. (2021) have suggested that high frontal beta band power may reflect increasing reliance on top-down mechanisms which facilitate the accumulation of decision-relevant sensory features. In support of this, Chen and Kwak (2022) also propose that beta oscillatory activity in the sensorimotor cortex indicates incentive motivational drive toward a choice action prior to a response, in which the action is computed from decision-related variables and stems from cortico-basal ganglia-thalamic circuitry. Thus, beta activity may serve as a valuable indicator for the decision-making process related to handling qualities as well as a derivative metric to indicate increased mental workload.

Additionally, the delta band of the EEG is even lesser understood as very few studies have investigated this frequency range and its relation to cognitive-motor output. Knyazev (2012) note that functional delta oscillations are implicated in brain activity related to autonomic function, motivational processes for both reward and atavistic defense, emotion, and cognitive processes related to attention and the detection of motivationally salient information (Harmony, 2013; Knyazev, 2007, 2012). In an early study, Vogel et al. (1968) suggested that delta activity is positively associated with efficient cognitive performance under conditions of mental effort. More recently, in support

Table 1.2: *Psychophysiological measures in response to increased mental workload*

Psychophysiological Measures	Indices of Increased Mental Workload
EEG - Alpha	Decrease
EEG - Theta	Increase
EEG - Beta	Increase
EEG - Theta/Alpha ratio	Increase
EEG - Coherence	Increase
Inceptor Workload	Increase
EMG	Increase

tions are implicated in brain activity related to autonomic function, motivational processes for both reward and atavistic defense, emotion, and cognitive processes related to attention and the detection of motivationally salient information (Harmony, 2013; Knyazev, 2007, 2012). In an early study, Vogel et al. (1968) suggested that delta activity is positively associated with efficient cognitive performance under conditions of mental effort. More recently, in support

of this notion, Harmony (2013) have proposed that sustained delta oscillations inhibit interference that may affect the performance of mental tasks by modulating the activity of networks that are not critical to the accomplishment of the task. In addition to its possible inhibitory function, delta oscillatory activity has been shown to make up a large portion of the P300, the late positive component of event-related potentials, implicated in attentional processing, signal detection, and decision making (Başar-Eroglu et al., 1992; Knyazev, 2012). Furthermore, Zarjam et al. (2011) have demonstrated the potential for the utilization of the delta frequency band as an indicator of increased mental workload and working memory load. These authors note that increased delta frequency activity contains significant electrophysiological correlates of cognitive processing and increased internal concentration (Zarjam et al., 2011). This group have also demonstrated beyond just delta spectral power that wavelet entropy measures are capable of discriminating up to seven levels of mental workload, particularly in the delta frequency range (Zarjam et al., 2013). For example, their results suggest that as mental workload increases, neural response increases in power and synchronization (as indicated by stronger peaks in coefficient distributions) in addition to increased reaction time, and decreased response accuracy (Zarjam et al., 2013). Moreover, delta oscillatory activity has been suggested to be a key mechanism of selective attention to visual stimulus streams (Lakatos et al., 2008). While much of the previous research on delta activity has been conducted using mental concentration and semantic tasks, the apparent lack of research related to delta activity and cognitive-motor processes make it a valid

metric to assess the motivation and concentration during a cognitive-motor task.

Accordingly, the current experiment aims to address a gap in the cognitive-motor neuroscience and aerospace engineering literatures regarding the combined assessment of dynamic objective psychophysiological variables explicitly related to mental workload, and subjective behavioral ratings of handling qualities, workload and effort. Consequently, the proposed experiment is designed to examine how objective brain dynamics and subjective behavioral ratings will respond to compensatory/multi-axis tracking tasks when handling qualities and task demand are manipulated. More specifically, the proposed study will assess multiple biomarkers explicitly related to mental workload as reported during flight-related compensatory tracking tasks of varying complexity while also challenged with progressively increasing levels of controllability. Thus, the contribution of the proposed experiment is the combination of objective brain dynamics with subjective ratings to estimate operator brain state during varying levels of task demand and controllability. Additionally, the results of the proposed work will address gaps in the literature related to the delta and beta bands of the EEG and their relation to workload and decision making. In the future, the expected results may be used in the design and testing of not only new aircraft, but in any system where handling qualities and/or controllability may be altered. Ultimately the expected results will provide confidence in the objective assessment of operator workload and thus the decision making process when criticizing the controllability of a system.

Section 2: Methods

2.1 Approach

This study was conducted to examine how psychophysiological and behavioral measures will respond to specific compensatory tracking tasks when handling qualities are manipulated. For each condition, there were three different levels of handling qualities. As such, the independent variables include Condition and Level of Handling Quality. This study followed a 2 (Condition: Single-axis vs. Multi-axis) x 3 (Level of Handling Quality: Level 1, Level 2, Level 3) repeated-measures design. Order of condition and levels of handling qualities were counterbalanced across all participants. The primary dependent variables of interest include subjective ratings (CHR, BWS, NASA-TLX), EEG power spectral density (delta, theta, alpha, beta), EEG $\frac{\theta}{\alpha}$ Ratio, and behavioral performance (desired, adequate). Other variables mentioned herein will be included as a programmatic effort in future scholarly publications.

2.2 Institutional Review Board Process

This project has undergone the respective Institutional Review Board (IRB) review processes at the United States Naval Academy, University of Maryland, College Park, and the Naval Air Warfare Center Aircraft Division Patuxent River (NAWCAD) and has secured approval at all three locations.

2.3 Participant Recruitment

A total of eight participants were recruited and consented according to best clinical research practices under approved United States Naval Academy and University of Maryland, College Park IRB consent processes (see Table 2.1). Eligible participants (healthy, right-handed, 19-55 years of age) free of mental disorders and medication for mental and attentional disorders were recruited via email to participate in this experiment. Hand dominance was evaluated according to the Edinburgh Handedness Inventory (Oldfield, 1971). Participants selected for the experiment were provided with a thorough explanation of the study design, purpose, and risks of participation. Participants were given the opportunity to ask questions and were reminded that they are

Table 2.1: Demographics

Participant Characteristics (N = 8)	
Variable	Overall ¹
Age (years)	37.1 (3.6)
Gender	
F	3
M	5
Sim Type	
Fixed Wing	10
Rotorcraft	3
Simulation Hours	627.4 (231.3)
Cockpit Hours	1,748.2 (355.0)
¹ mean (S.E.)	

under no obligation to participate and may withdraw from the study at any time without consequence or penalty. Fleet pilots employed at the United States Naval Academy were the primary participant pool for this study, however one United States Air Force Aviator also participated.

2.4 Experimental Design

2.4.1 Procedure

All data collection occurred in a laboratory space at the United States Naval Academy. Before testing began, participants were briefed on the protocol and objectives of the study, and informed consent was obtained. Participants were sized and fitted with a stretchable fabric cap in which electrodes recorded electrical brain activity via 64 electrodes housed within a stretchable lycra cap (actiCAP slim,

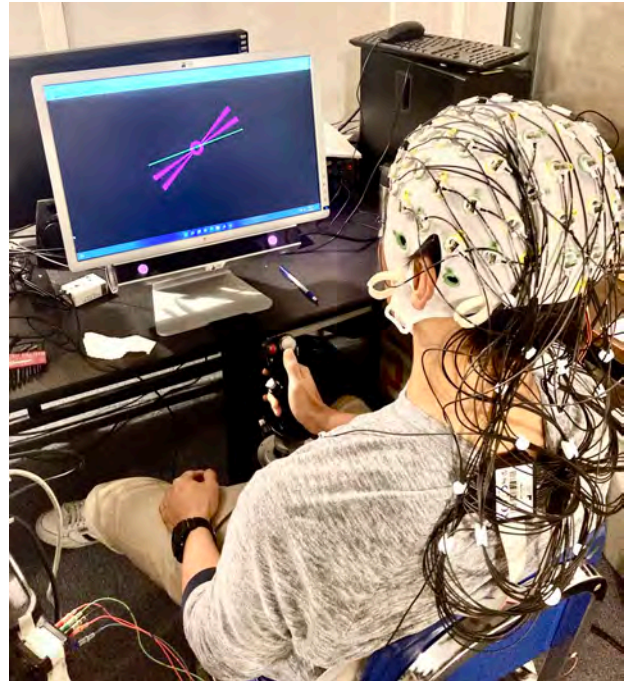


Figure 2.1: Experimental setup

actiCHamp, BrainProducts, GmbH) sampled at 2500 Hz and referenced to site FCz with a common ground. Electrode placement followed the international 10–10 system (Chatrian et al., 1985). To collect EMG, a total of eight sensors were placed on the skin, two each on the biceps, triceps, flexors and extensors

of right arm to record muscle activity of the upper limb during the task performance. One electrode was positioned on the left clavicle for EKG recording. Lastly, an eye-tracking camera (SMI RED250, Sensomotoric Instruments (SMI), Apple, Inc.) was placed in front of the participant. The distance from the screen to participants' eyes was held to 65cm to ensure proper calibration and data tracking from the eye-tracking system.

Participants were seated comfortably at a distance of 65 cm in front of a 47cm x 29cm computer monitor (Dell). Participants controlled all input to the system via a HOTAS WARTHOGTM (Thrustmaster®, France) joystick controller. To simulate a more realistic experience, the joystick was attached to an ad-

justable height extension apparatus, below the computer monitor. Following equipment set up, participants completed two, two-minute reference conditions, which consisted of sitting passively 1) with their eyes closed, and 2) while viewing the bowtie display. Finally, the participants were oriented to the subjective metrics of interest CHR, BWS, and NASA-TLX, and the objective of the task were described to the participant.

Participants completed a series of compensatory tracking tasks (see section Section 2.4.2 and Table 2.2 for task description and performance criteria,

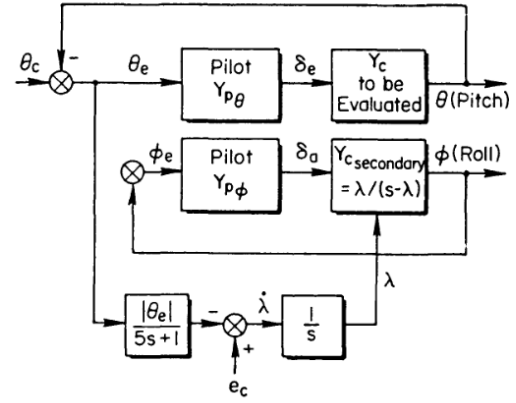


Figure 2.2: Single-loop primary task with secondary cross-coupled loading task (McRuer et al., 1990)

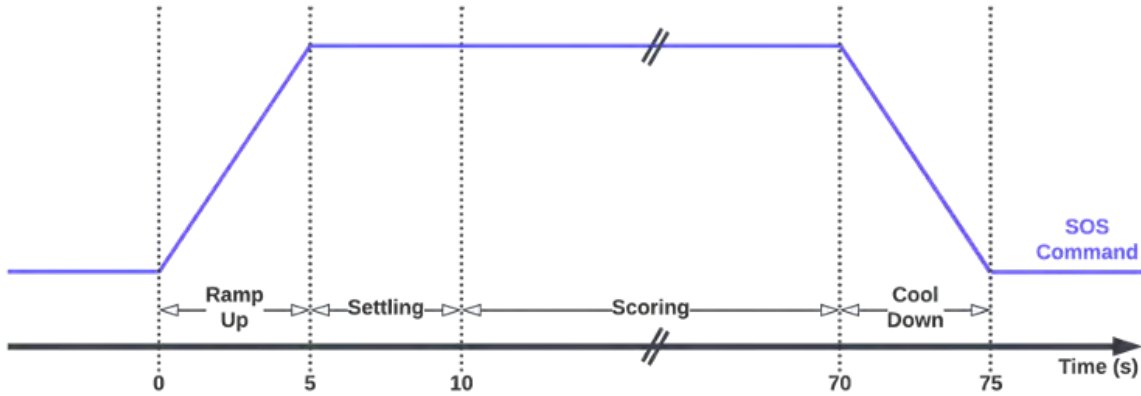


Figure 2.3: Visual representation for SoS timeline

respectively). The seminal multi-axis task (Figure 2.2) devised by McRuer et al. (1989) was executed to quantify the “attentional demand” and the “excess control capacity” that are associated with the pilot’s CHR, where the attentional demand and the excess control capacity are negatively related. The multi-axis task is composed of the primary task (single-axis; compensatory pitch task in the top portion of Figure 2.2), and the cross-coupled secondary task (multi-axis; compensatory roll task in the rest of the system).

The Sum of Sines in the task are driven by an automated command signal generated by a custom coded simulation environment (MATLAB r2022b, The MathWorks Inc.). The input begins with 10 s of non-scoring time (i.e., 0 to 10 s) that includes a 5 s initial linear ramp up in amplitude. This is followed by 60 s of scoring time (i.e., 10 to 70 s) and a 5-s cool down period (i.e., 70 to 75 s) where the amplitude is ramped back to zero (see Figure 2.3). Thus, each tracking task consists of 75 s test and post-scenario questionnaires. However, the window of interest for this experiment consisted of the 60 s scoring time

window.

2.4.2 Description of Task

Prior to beginning testing, participants completed three “practice” trials of the task related to the assigned counterbalanced order of flight task (i.e., single-axis or multi-axis). Participants viewed the display (Figure 2.4) and the details of its operation was described such that the pitch indication is represented by the green dot at the center of the image and the roll indication is represented by the green horizontal line. The objective of the flight task for pitch (desired performance; see Table 2.2) is to keep the green dot within the inner circle, while adequate performance is to keep it within the magenta circle. Desired performance for roll is to keep the green line within the inner most wedge shape, and adequate performance is to keep it within the magenta wedge shape. Participants were asked to aggressively track the displayed green line representing the SoS command signal and to keep errors within the specified tolerances in table 2.2.

Following the three practice trials, participants were provided an oppor-

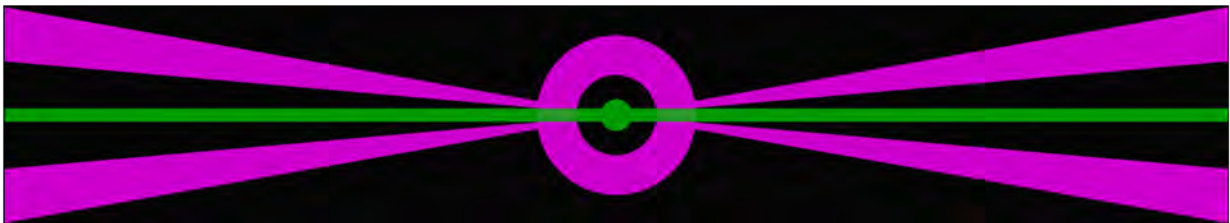


Figure 2.4: Bowtie heads-up display (magenta polygon) which participants controlled during the compensatory SoS tracking tasks (green dot and line) (Klyde et al., 2020)

tunity to ask any questions about the task and the objective of the was repeated. Following the protocol outlined in Figure 2.6 participants completed a total of six SoS tracking tasks following a counterbalanced combination of three levels of handling quality (Low (system parameters corresponding to the ADS-33E (Baskett, 2000)) Level 1 HQ), Medium (Level 2), and High (Level 3)) nested within two conditions ((i) Single-Axis (Pitch), (ii) Multi-Axis (Pitch and Roll)). Accordingly, the CHR decision tree HQR is designed to tightly couple:

1. the extent of the HQ deficiencies (e.g., Level 1: no worse than mildly unpleasant, Level 2: minor to very objectionable, Level 3: major, Level 10: so severe that there is a high risk of loss of control),
2. with the achieved performance (desired, adequate, inadequate),
3. and the required compensation (not a factor, minimal, moderate, considerable, extensive, maximum tolerable, controllability in question) (Cooper & Harper, 1969).

What is considered desired and adequate task performance for when operators are given a mission objective, or in the terminology of the ADS-33E, mission task elements (MTEs), is often governed by the bounds of the CHR. For example, for a given MTE, both desired and adequate levels of performance can be cal-

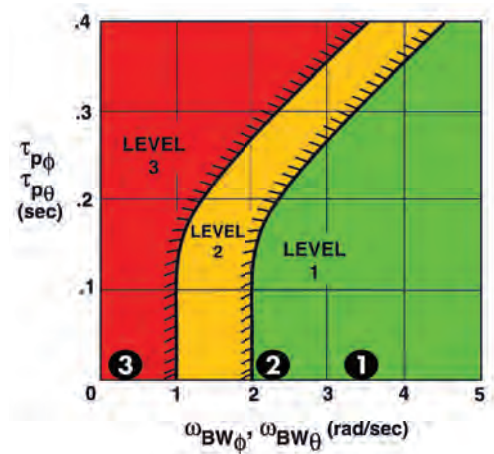


Figure 2.5: Requirements for small-amplitude pitch (roll) attitude changes - hover and low speed (Baskett, 2000)

Table 2.2: Performance standards for task objectives

	Desired	Adequate
Pitch: At least $x\%$ of the scoring time within pitch attitude error tolerance:	50% $\pm 1^\circ$	75% $\pm 2^\circ$
Roll: At least $x\%$ of the scoring time within roll attitude error tolerance:	50% $\pm 5^\circ$	75% $\pm 10^\circ$

culated after task completion based upon

the scoring period. The MTE dictates what is considered desired and adequate (see Table 2.2 for proposed experiment criteria), and an overall performance assessment can be provided. This result, paired with the combined usage of objective and subjective metrics, can provide not only measures related to handling qualities and increased mental workload, but also when exactly the operator may have felt their highest (or lowest) level of workload to make their decision for the subjective rating.

The task environment was manipulated via employing the single or multi-task environment and altering the bandwidth of the controller. The appropriate bandwidth and delay values are provided in the ADS-33E, which were implemented for the given task and level of handling quality (see Figures 2.5) where M_q and L_p represent pitch and roll, respectively.

Participants were introduced to and briefed on the objectives of the CHR, BWS, and NASA-TLX prior to beginning the experiment. Participants were allowed to ask questions for clarification on these questionnaires, and the descriptions of each were provided to the participant if they needed to refer back

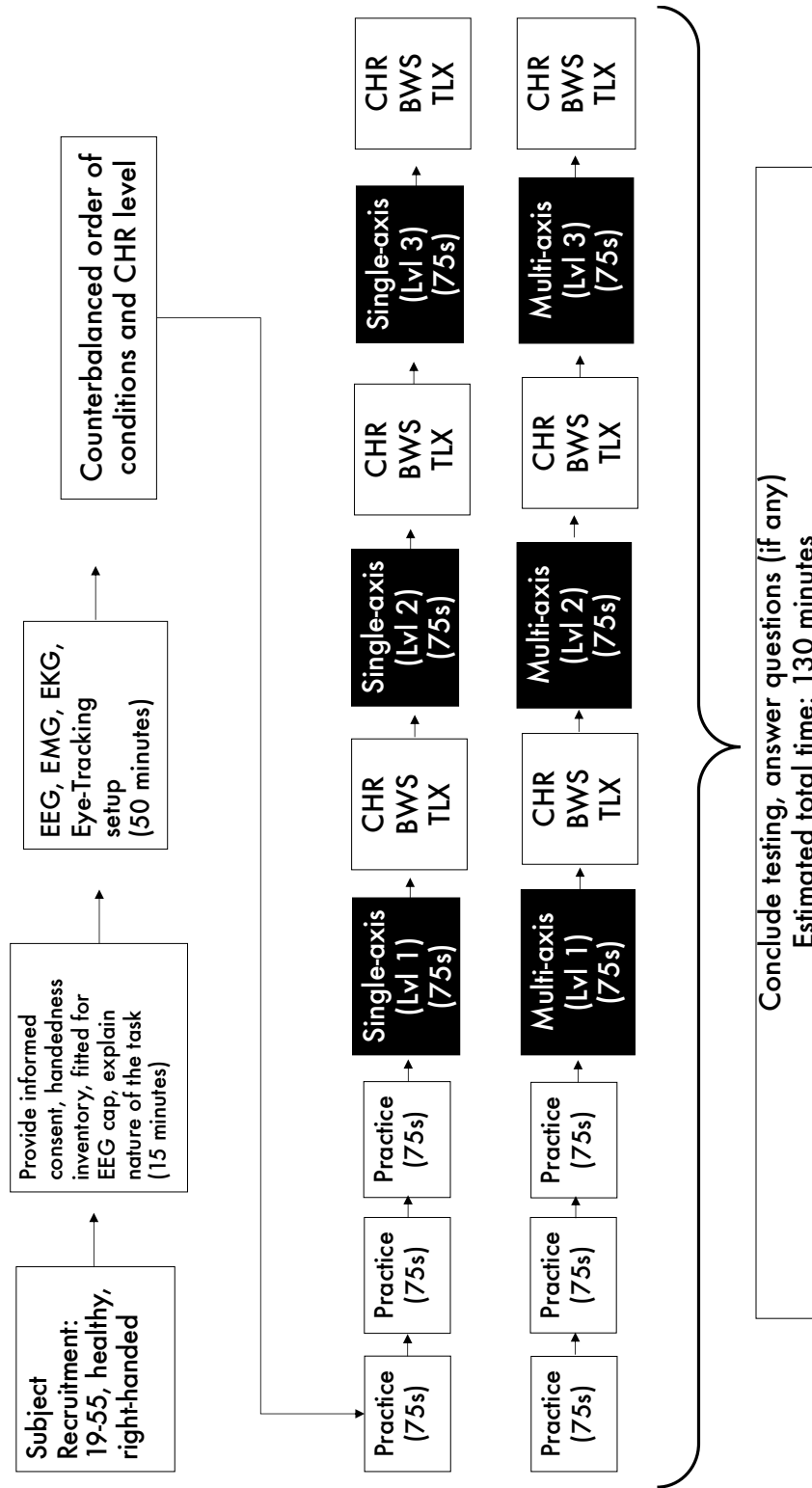


Figure 2.6: Study protocol

to this information while completing each questionnaire. The degree to which each of the six factors within the NASA-TLX contribute to the workload of the specific task to be evaluated is determined by the participants' responses to pair-wise comparisons among the six factors (Hart, 1986). There are 15 possible pair-wise comparisons of the six scales, where each comparison is presented to the participant on a flash card. Participants are asked to circle the scale title that represents the more important contributor to workload for the specific task being performed in the experiment (Hart, 1986). The ratings of factors deemed most important in creating the workload of a task are given more weight in computing the overall workload score. Participants completed the CHR, BWS, and NASA-TLX following each trial.

2.5 Signal Processing

EEG data pre-processing was conducted in BrainVision Analyzer (v. 2.1, BrainProducts, GmbH), MATLAB R2023b (The MathWorks Inc.), and EEGLAB v.2022.1 (Delorme & Makeig, 2004). EEG data was inspected for bad channels in BrainVision Analyzer and, when necessary, bad channels will be replaced with spherically interpolated channels. Raw EEG data was re-referenced to a common-average of all EEG channels, excluding all EOG, EMG, and EKG sensors, and was first high-pass filtered with a zero-phase shift Infinite Impulse Response Butterworth filter. The low-cutoff frequency was set to 0.2 Hz and the filter order was set to eight. The EEG data were then subjected to a biased

extended Infomax Independent Component Analysis (ICA) to correct for ocular artifact in BranVision Analyzer. The EEG data were then exported for further processing in MATLAB.

EEG data were subjected to a custom elliptic infinite impulse response (IIR) band-pass filter from 1 to 30 Hz. EEG data were then downsampled to 256 Hz and subjected to the extended Infomax Independent Component Analysis. Decomposition matrices will be applied to the non-downsampled EEG data, in order to improve the quality of ICA decomposition. Independent components (ICs) determined to be unrelated to the task-relevant EEG signal were removed using the Multiple Artifact Rejection Algorithm (MARA) EEGLAB plugin MARA (Winkler et al., 2011). A current source density (CSD) transformation was applied to these denoised EEG data using the CSD Toolbox (version 1.1; Kayser, 2009; Kayser & Tenke, 2006a, 2006b) with the spherical spline flexibility factor (m) of 4 and the smoothing factor (λ) of 1.0×10^5 , respectively. EEG data were baseline corrected using the mean of the one-second period prior to the begin of SoS scoring. Power spectral densities were computed for each 60-s trial using Welch's method (Welch, 1967) utilizing a periodic hamming window of 4 s with 50% overlap and the number of fast Fourier transform (FFT) points of 5000. Power spectra were approximated by applying the trapezoidal rule to each of seven EEG frequency bands, including delta (1-4 Hz), theta (4-8 Hz), broadband alpha (8-13 Hz), and broadband beta (13-30 Hz).

2.5.1 Wavelet-based Regression Analyses

Twelve regional average EEG activities were calculated over the baseline corrected EEG corresponding to the anatomical brain regions of interest (ROIs; see Table 2.3 based on Koessler (2009). The identified brain regions of interest were selected to capture the perception, action and decision-making related to the task.

Table 2.3: EEG channels selected for wavelet analysis based on ROI

Anatomical Regions of Interest (ROIs)	EEG Channels within ROI
Superior Frontal Gyrus (SFG)	Fp1, Fz, F3, FC1, FCz, AF3, AFz, F1, Fp2, F4, FC2, AF4, F2
Middle Frontal Gyrus (MFG)	Fp1, F3, AF7, AF3, F1, F5, FC3, Fp2, F4, AF8, AF4, F2, F6, FC4
Inferior Frontal Gyrus (IFG)	F7, FC5, F5, F8, FC6, F6
Precentral Gyrus (PCG)	FC5, C3, FC3, C1, C5, FC6, C4, FC4, C2, C6
Postcentral Gyrus (POG)	C3, CP1, C1, C5, CPz, C4, CP2, C2, C6, CPz
Superior Parietal Lobule (SPL)	CP1, Pz, P1, CP2, Pz, P2
Inferior Parietal Lobule (IPL)	CP5, P3, C5, CP3, CP6, P4, C6, CP4
Precuneus (PCN)	Pz, P3, P1, POz, P4, P2
Cuneus (CN)	O1, Oz, POz, O2
Superior Temporal Gyrus (STG)	T7, CP5, FT7, T8, CP6, FT8
Middle Temporal Gyrus (MTG)	T7, P7, TP7, P5, T8, P8, TP8, P6
Inferior Temporal Gyrus (ITG)	FT9, TP9, P7, TP7, FT10, TP10, P8, TP8

The regional EEG wavelet power spectra were computed per these ROIs for all six experiment conditions per participant. Specifically, a continuous wavelet transform filter bank was designed with the analytic Morse wavelet having a periodic boundary condition

Table 2.4: Timing of CHR rating

Subject ID	Decision-Making Period
USN0001	Average
USN0002	Middle
USN0003	Early
USN0004	Average
USN0005	Late
USN0006	Average
USAF0007	Average
USN0008	Early

on the reflected series and the following parameters: gamma of 3, the time-bandwidth product of 75, 48 voices per octave, 8 octaves, and the frequency limits between 1 and 30 Hz. The magnitude-squared scalogram (i.e., wavelet

power spectrum) was averaged over scale (i.e., frequency) and then normalized to equal the variance of the EEG time series (i.e., the total power of EEG time series), through which the regional EEG power was obtained for the four constituent spectral bands (i.e., delta, theta, alpha, and beta bands).

Next, in order to examine the association between the objective regional EEG power and the subjective rating scales, the linear mixed-effects models were fitted to the variables specified by the formula represented by the Wilkinson notation:

$$RatingScale \sim Biomarker + (1|EEG) + (Biomarker - 1|EEG) \quad (2.1)$$

where *RatingScale* is a response variable for the subjective rating scales (i.e., CHR), *EEG* is the grouping variable for all the combinations of twelve brain regions and four EEG bands, *Biomarker* is the fixed-effects predictor variables for the objective biomarkers (i.e., the regional EEG powers), and $(1|EEG)$ and $(Biomarker - 1|EEG)$ are the random-effects terms for the intercept and the objective biomarkers, respectively. Specifically, the random-effect terms for the intercept and the biomarkers may not be uncorrelated, so these terms are specified separately. Particularly, for the linear fixed-effects models, both the mean and maximum regional EEG power (a) over the whole 60-s scoring period, (b) over the three phases, including early (i.e., 0-20 s), middle (i.e., 20-40 s), and late (i.e., 40-60 s), and (c) the individual responses when the decision of the subjective ratings were made (see Table 2.4), again represented as early,

middle, and late phases, respectively.

As a metric of mental workload, the theta/alpha ratio was computed as the average theta power of all EEG channels comprising the superior frontal gyrus divided by the average alpha power of all EEG channels comprising the superior parietal lobule (i.e., $\frac{\bar{x}_\theta}{\bar{x}_\alpha}$).

2.6 Performance Scoring Criteria

Participants' performance scoring was computed for the middle 60 seconds of each trial according to their adherence to the performance standards criteria (see Figure 2.3, and Table 2.2). A custom scoring code was generated such that a constraint was placed on the performance data where pitch and/or roll are invalidated if either of the axes were inadequate (i.e., not meeting the performance standards listed in the table). Performance was scored for 60 seconds, using a sliding 5 second window, with 50% overlap. Pitch performance was scored as the combination of time and difference between the center of the pitch reference in the SoS and the center of the pitch attitude in the bowtie. Roll performance was calculated by first defining the centroid of the segments SoS and the centroids of each polygon of the bowtie. Additionally, angle operations were computed in order to determine roll angle and directionality during task performance. Roll performance was calculated similarly to pitch, such that performance consisted of the combination of time and angle between the bowtie and SoS. Finally, overall mean scores for desired, adequate,

and inadequate were generated for each participant and each trial. Moreover, a constraint was applied for when one dimension was deemed inadequate, the other dimension was not counted during this time.

2.7 Statistical Analysis

Statistics were calculated in RStudio (RStudio, Boston, MA). The independent variables include Condition and Level of Handling Quality. The primary dependent variables of interest include subjective ratings (CHR, BWS, NASA-TLX), EEG power spectral density (delta, theta, alpha, beta), EEG $\frac{\theta}{\alpha}$ Ratio, and behavioral performance (desired, adequate). Pearson correlations were tested at $\alpha = 0.05$. Simple effects of task condition (single-axis vs. multi-axis) as well as level of handling qualities on performance and mental workload were conducted via paired *t*-tests. Simple effects of condition (single-axis vs. multi-axis) on performance and EEG band power were conducted via paired *t*-tests. P-values were adjusted for multiple comparisons using the False Discovery Rate. Effect sizes will be reported where appropriate. Significant relationships are reported as : * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The dependent variables for subjective ratings (CHR, BWS, NASA-TLX) were subjected to separate 2 (Condition: Single-axis vs. Multi-axis) x 3 (Level of Handling Quality: Level 1, Level 2, Level 3) completely within-subjects repeated-measures ANOVAs to examine main and interactive effects for workload and level of handling quality.

The dependent variables for EEG power spectra were subjected to separate 2 (Condition: Single-axis vs. Multi-axis) x 3 (Level of Handling Quality: Level 1, Level 2, Level 3) x 5 (Region: Frontal, Central, Parietal, Temporal, Occipital) completely within-subjects repeated-measures ANOVAs to examine main and interactive effects for workload and level of handling quality.

The dependent variable for EEG $\frac{\theta}{\alpha}$ Ratio was subjected to a 2 (Condition: Single-axis vs. Multi-axis) x 3 (Level of Handling Quality: Level 1, Level 2, Level 3) completely within-subjects repeated-measures ANOVA to examine main and interactive effects for workload and level of handling quality.

The dependent variables for behavioral performance (Single-axis, Multi-axis) were subjected to separate 2 (Condition: Single-axis vs. Multi-axis) x 3 (Level of Handling Quality: Level 1, Level 2, Level 3) completely within-subjects repeated-measures ANOVAs to examine main and interactive effects for workload and level of handling quality.

2.7.1 Regression Approach

A linear regression approach was utilized in order to understand the relationship between Cooper-Harper Ratings and objective brain dynamics. Such an approach will allow for an examination of the ability of the objective psychophysiological features to predict subjective ratings of handling quality.

In addition to predicting brain dynamics, an additional linear regression approach was utilized in order to understand the relationship between behav-

ioral performance and Cooper-Harper Ratings. Such an approach will allow for an examination of the ability of the behavioral performance to predict subjective ratings of handling quality.

Furthermore, an additional linear regression approach was utilized in order to understand the relationship between behavioral performance and objective brain dynamics. Such an approach will allow for an examination of the ability of the objective psychophysiological features to predict behavioral performance.

A series of scatterplots were produced which allows for illustration of the relationship between these independent and dependent variables to quantify the slopes and intercepts of the relationship.

Section 3: Results

3.1 Subjective Ratings

3.1.1 Cooper-Harper Rating Scale

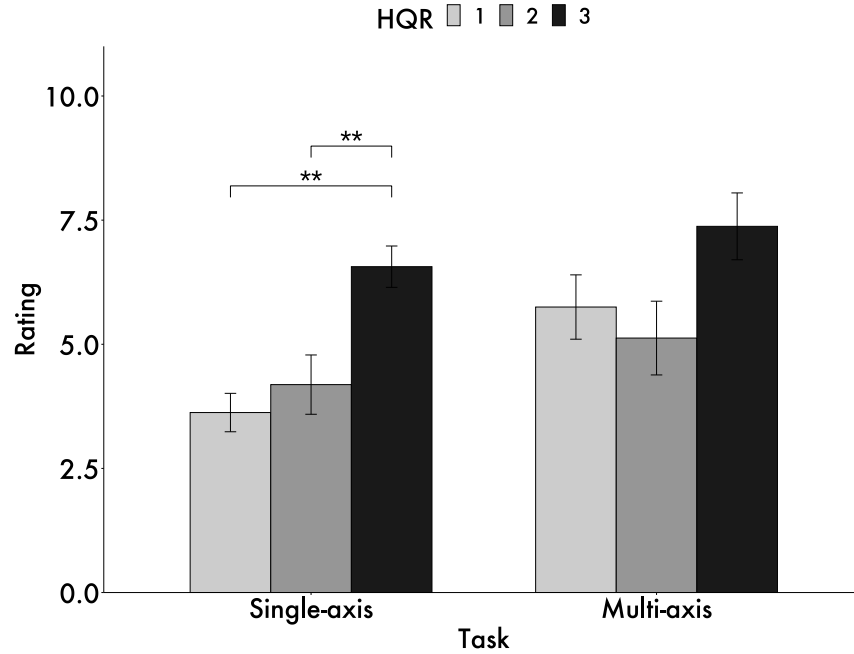


Figure 3.1: Mean Cooper-Harper Rating by Condition and Level of Handling Quality.

A repeated measures ANOVA was performed to compare effects of Single-axis and Multi-axis conditions as well as level of Handling Quality on Cooper-Harper ratings. There was no significant main effect for Condition ($F_{(1,7)} = 6.21, p = 0.06$). There was a significant main effect for HQR ($F_{(2,14)} = 17.28, p <$

0.001). Pairwise comparisons for the Single-axis Condition revealed that the mean Cooper-Harper rating was significantly different in Level 1 ($M = 3.63 \pm 0.4$) vs. Level 3 ($M = 6.56 \pm 0.4$) HQR ($t_{(7)} = -5.38, p < 0.01, d = -1.9$), and in Level 2 ($M = 4.19 \pm 0.6$) vs. Level 3 ($M = 6.56 \pm 0.4$) HQR ($t_{(7)} = -4.46, p < 0.01, d = -1.58$).

3.1.2 Bedford Workload Scale

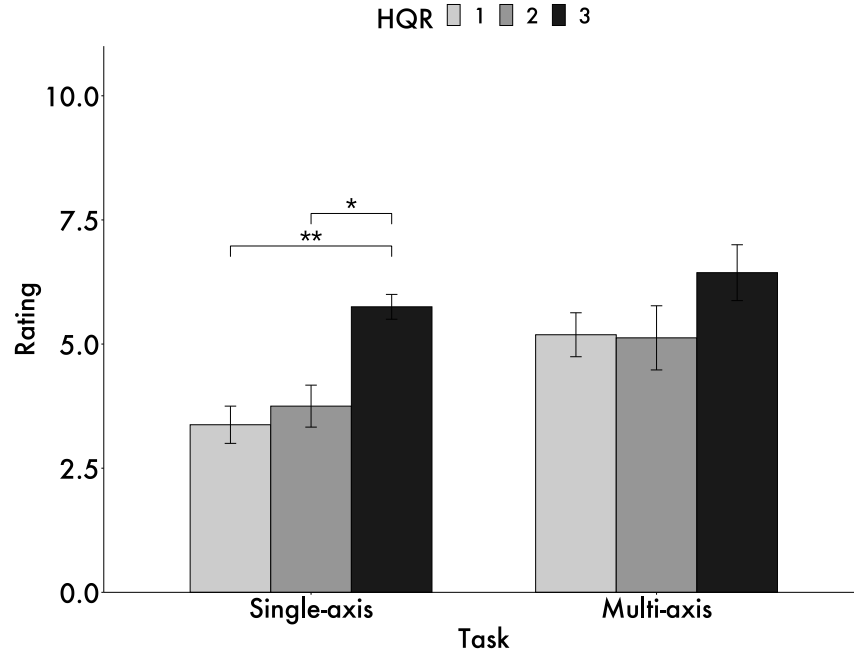


Figure 3.2: Mean Bedford Workload Rating by Condition and Level of Handling Quality.

A repeated measures ANOVA was performed to compare effects of Single-axis and Multi-axis conditions as well as level of Handling Quality on Bedford Workload ratings. There was no significant main effect for Condition ($F_{(1,7)} = 6.08, p = 0.06$). There was a significant main effect for HQR ($F_{(2,14)} = 13.86, p < 0.01$). Pairwise comparisons for the Single-axis Condition revealed that the mean Bedford Workload rating was significantly different in Level 1

($M = 3.38 \pm 0.38$) vs. **Level 3** ($M = 5.75 \pm 0.3$) **HQR** ($t_{(7)} = -5.65, p < 0.01, d = -2.0$), and in **Level 2** ($M = 3.75 \pm 0.42$) vs. **Level 3** ($M = 5.75 \pm 0.3$) **HQR** ($t_{(7)} = -3.39, p < 0.05, d = -1.20$).

3.1.3 NASA-TLX Overall Workload

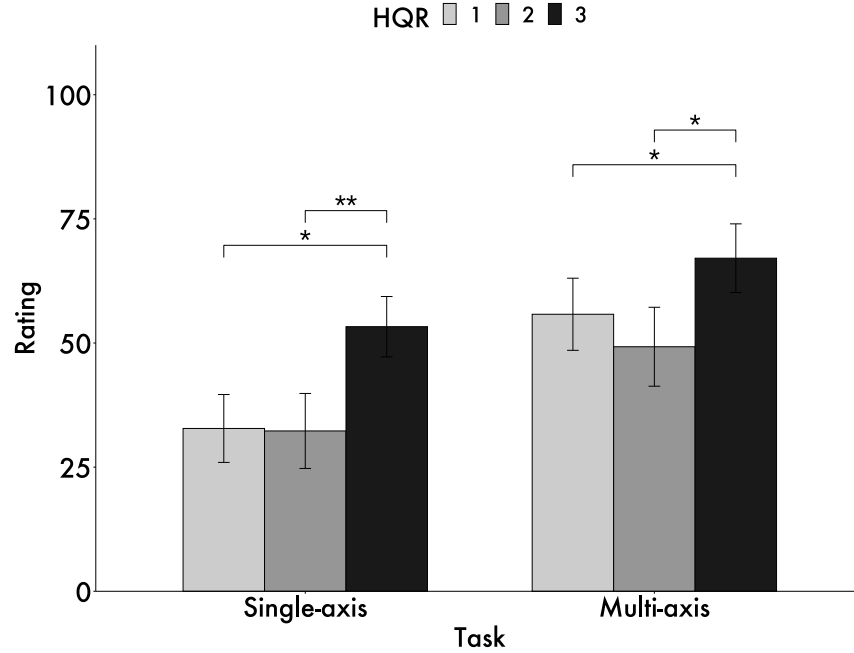


Figure 3.3: Mean weighted NASA-TLX Overall workload rating by Condition and Level of Handling Quality.

A repeated measures ANOVA was performed to compare effects of Single-axis and Multi-axis conditions as well as level of Handling Quality on NASA-TLX Overall workload ratings. There was a significant main effect for Condition ($F_{(1,7)} = 15.84, p < 0.01$). There was a significant main effect for HQR ($F_{(2,14)} = 23.83, p < 0.001$). Pairwise comparisons for the Single-axis Condition revealed that the mean NASA-TLX Overall workload rating was significantly different in **Level 1** ($M = 32.89 \pm 6.84$) vs. **Level 3** ($M = 53.3 \pm 6.1$) **HQR** ($t_{(7)} = -3.63, p <$

0.05, $d = -1.28$), and in Level 2 ($M = 32.3 \pm 7.6$) vs. Level 3 ($M = 53.3 \pm 6.1$) HQR ($t_{(7)} = -5.97, p < 0.01, d = -2.11$). Pairwise comparisons for the Multi-axis Condition revealed that the mean NASA-TLX Overall workload rating was significantly different in Level 1 ($M = 55.8 \pm 7.3$) vs. Level 3 ($M = 67.1 \pm 6.91$) HQR ($t_{(7)} = -4.0, p < 0.05, d = -1.41$), and in Level 2 ($M = 49.25 \pm 7.95$) vs. Level 3 ($M = 67.1 \pm 6.91$) HQR ($t_{(7)} = -3.70, p < 0.05, d = -1.31$).

3.1.4 NASA-TLX Rating of Performance

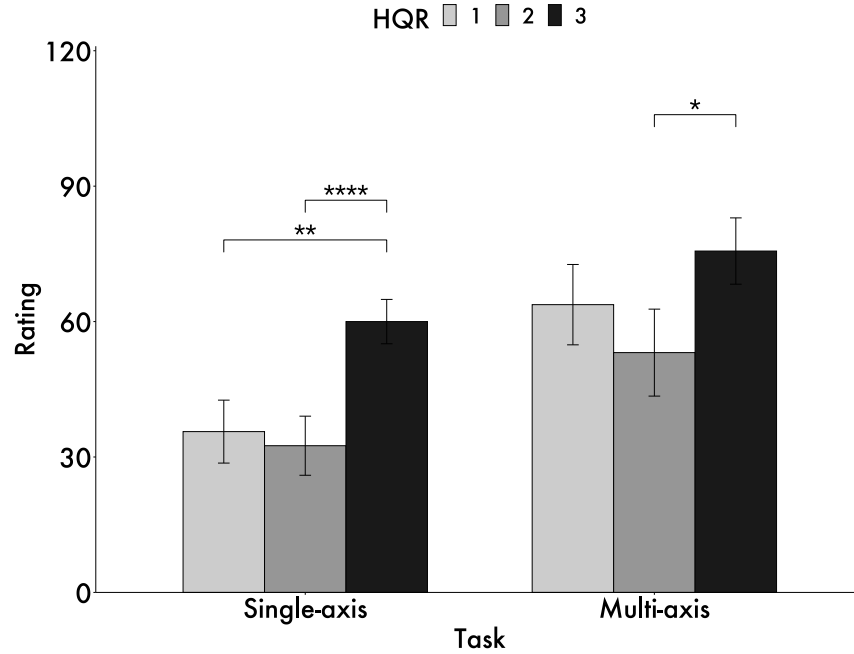


Figure 3.4: Mean NASA-TLX Performance rating by Condition and Level of Handling Quality. Higher ratings are inversely related to performance.

A repeated measures ANOVA was performed to compare effects of Single-axis and Multi-axis conditions as well as level of Handling Quality on NASA-TLX Performance ratings. There was a significant main effect for Condition ($F_{(1,7)} = 19.47, p < 0.001$). There was a significant main effect for HQR ($F_{(2,14)} =$

30.81, $p < 0.01$). Pairwise comparisons for the Single-axis Condition revealed that the mean NASA-TLX Performance rating was significantly different in Level 1 ($M = 35.63 \pm 6.97$) vs. Level 3 ($M = 60.0 \pm 4.91$) HQR ($t_{(7)} = -4.92, p < 0.01, d = -1.74$), and in Level 2 ($M = 32.5 \pm 6.55$) vs. Level 3 ($M = 60.0 \pm 4.91$) HQR ($t_{(7)} = -11.0, p < 0.001, d = -3.89$). Pairwise comparisons for the Multi-axis Condition revealed that the mean NASA-TLX Performance rating was significantly different in Level 2 ($M = 53.13 \pm 9.63$) vs. Level 3 ($M = 75.63 \pm 7.35$) HQR ($t_{(7)} = -3.97, p < 0.05, d = -1.40$).

3.2 Power Spectral Density Results

A series of repeated measures ANOVAs were performed to compare the effects of Single-axis and Multi-axis conditions as well as level of Handling Quality on EEG Power Spectral Density. This experiment utilized a 2 Condition (Single-axis vs. Multi-axis) x 3 HQR (Level 1, Level 2, Level 3) x 5 Region (Frontal, Central, Parietal, Temporal, Occipital) repeated measures ANOVA design. Alpha levels were set at 0.05, and p-values were adjusted for multiple comparisons using the False Discovery Rate. Effect sizes will be reported where appropriate.

3.2.1 Delta Power Spectral Density

A repeated measures ANOVA was performed to compare effects of Single-axis and Multi-axis conditions as well as level of Handling Quality on EEG

Delta power spectral density. In the frontal Region, there were no significant main effects for Condition ($F_{(1,7)} = 0.18, p = 0.84$), nor HQR ($F_{(2,14)} = 0.33, p = 0.84$). In the central Region, there were no significant main effects for Condition ($F_{(1,7)} = 2.83, p = 0.5$), nor HQR ($F_{(2,14)} = 2.74, p = 0.5$). In the parietal Region, there were no significant main effects for Condition ($F_{(1,7)} = 2.39, p = 0.5$), nor HQR ($F_{(2,14)} = 0.9, p = 0.5$). In the temporal Region, there were no significant main effects for Condition ($F_{(1,7)} = 0.94, p = 0.6$), nor HQR ($F_{(2,14)} = 1.62, p = 0.53$). In the occipital Region, there were no significant main effects for Condition ($F_{(1,7)} = 1.03, p = 0.6$), nor HQR ($F_{(2,14)} = 0.39, p = 0.8$).

3.2.2 Theta Power Spectral Density

A repeated measures ANOVA was performed to compare effects of Single-axis and Multi-axis conditions as well as level of Handling Quality on EEG Theta power spectral density. In the frontal Region, there were no significant main effects for Condition ($F_{(1,7)} = 1.4, p = 0.77$), nor HQR ($F_{(1.07,7.46)} = 0.12, p = 0.87$). In the central Region, there were no significant main effects for Condition ($F_{(1,7)} = 0.38, p = 0.84$), nor HQR ($F_{(1.07,7.49)} = 0.63, p = 0.77$). In the parietal Region, there were no significant main effects for Condition ($F_{(1,7)} = 1.73, p = 0.77$), nor HQR ($F_{(2,14)} = 2.32, p = 0.77$). In the temporal Region, there were no significant main effects for Condition ($F_{(1,7)} = 1.01, p = 0.77$), nor HQR ($F_{(2,14)} = 0.95, p = 0.77$). In the occipital Region, there were no significant main effects for Condition ($F_{(1,7)} = 0.24, p = 0.85$), nor HQR ($F_{(2,14)} = 1.0, p = 0.77$).

3.2.3 Alpha Power Spectral Density

A repeated measures ANOVA was performed to compare effects of Single-axis and Multi-axis conditions as well as level of Handling Quality on EEG Alpha power spectral density. In the frontal Region, there were no significant main effects for Condition ($F_{(1,7)} = 1.21, p = 0.76$), nor HQR ($F_{(1.19,8.3)} = 1.21, p = 0.8$). In the central Region, there were no significant main effects for Condition ($F_{(1,7)} = 0.23, p = 0.8$), nor HQR ($F_{(2,14)} = 1.81, p = 0.76$). In the parietal Region, there were no significant main effects for Condition ($F_{(1,7)} = 0.53, p = 0.8$), nor HQR ($F_{(1.1,7.69)} = 0.82, p = 0.76$). In the temporal Region, there were no significant main effects for Condition ($F_{(1,7)} = 1.86, p = 0.76$), nor HQR ($F_{(2,14)} = 1.08, p = 0.76$). In the occipital Region, there were no significant main effects for Condition ($F_{(1,7)} = 7.02, p = 0.44$), nor HQR ($F_{(2,14)} = 0.31, p = 0.79$).

3.2.4 Beta Power Spectral Density

A repeated measures ANOVA was performed to compare effects of Single-axis and Multi-axis conditions as well as level of Handling Quality on EEG Beta power spectral density. In the frontal Region, there were no significant main effects for Condition ($F_{(1,7)} = 2.14, p = 0.65$), nor HQR ($F_{(2,14)} = 2.1, p = 0.65$). In the central Region, there were no significant main effects for Condition ($F_{(1,7)} = 2.20, p = 0.65$), nor HQR ($F_{(1.02,7.17)} = 0.57, p = 0.65$). In the parietal Region, there were no significant main effects for Condition ($F_{(1,7)} = 1.13, p = 0.65$), nor HQR ($F_{(1.15,8.05)} = 0.65, p = 0.65$). In the temporal Region, there were no significant main

effects for Condition ($F_{(1,7)} = 0.97, p = 0.65$), nor HQR ($F_{(1.09,7.65)} = 0.091, p = 0.85$).

In the occipital Region, there were no significant main effects for Condition ($F_{(1,7)} = 0.13, p = 0.84$), nor HQR ($F_{(1.1,7.4)} = 0.03, p = 0.89$).

3.2.5 $\frac{\theta}{\alpha}$ Ratio

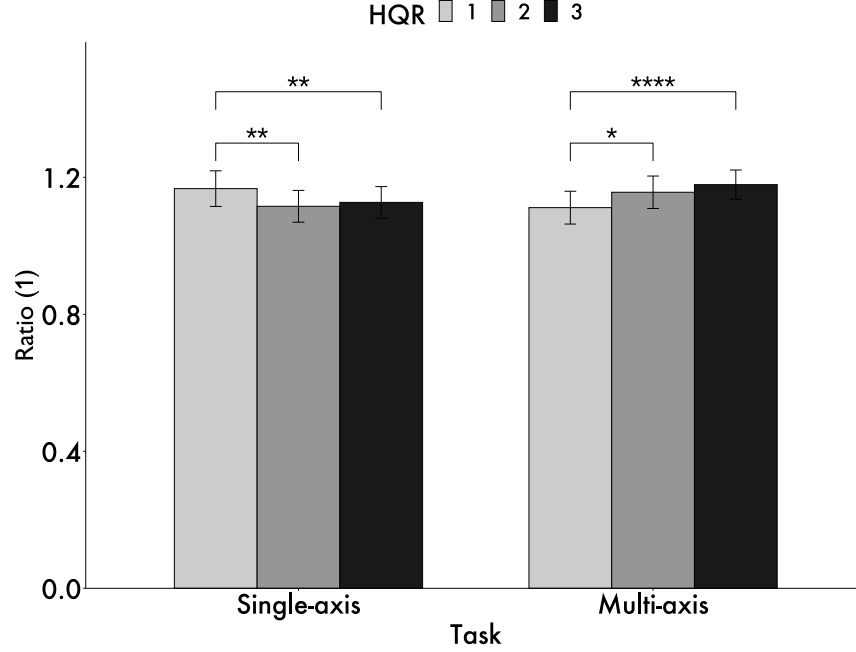


Figure 3.5: Mean $\frac{\theta}{\alpha}$ Ratio by Condition and Level of Handling Quality

A repeated measures ANOVA was performed to compare effects of Single-axis and Multi-axis conditions as well as level of Handling Quality on EEG $\frac{\theta}{\alpha}$ Ratio. There were no significant main effects for Condition ($F_{(1,58)} = 0.73, p = 0.34$), nor HQR ($F_{(2,116)} = 1.22, p = 0.34$). A significant Condition x HQR interaction was observed ($F_{(1.66,96.55)} = 13.95, p < 0.001$). Pairwise comparisons for the Single-axis Condition revealed that the mean $\frac{\theta}{\alpha}$ was significantly different in Level 1 ($M = 1.17 \pm 0.05$) vs. Level 2 ($M = 1.12 \pm 0.05$) HQR ($t_{(58)} = 2.91, p < 0.01, d = 0.38$), and in Level 1 ($M = 1.17 \pm 0.05$) vs. Level 3 ($M = 1.13 \pm 0.05$) HQR ($t_{(58)} =$

2.9, $p < 0.01$, $d = 0.38$). Pairwise comparisons for the Multi-axis Condition revealed that the mean $\frac{\theta}{\alpha}$ was significantly different in Level 1 ($M = 1.11 \pm 0.05$) vs. Level 2 ($M = 1.16 \pm 0.05$) HQR ($t_{(58)} = -2.43$, $p < 0.05$, $d = -0.32$), and in Level 1 ($M = 1.17 \pm 0.05$) vs. Level 3 ($M = 1.18 \pm 0.04$) HQR ($t_{(58)} = -5.2$, $p < 0.001$, $d = -0.68$). This metric was also alternatively analyzed by computing the mean theta power spectral density of all frontal electrodes divided by the mean alpha power spectral density of all parietal electrodes. While this approach yielded a similar pattern, it did not produce significant results.

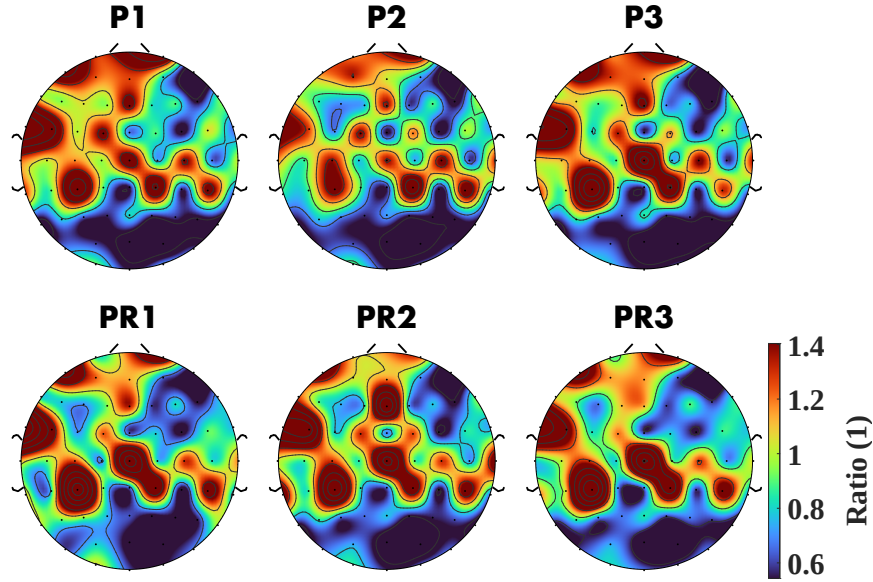


Figure 3.6: Scalp topographical maps of $\frac{\theta}{\alpha}$ Ratio. The top row represents the Single-axis Condition, Levels 1, 2, 3, left to right, respectively. The bottom row represents the Multi-axis Condition, Levels 1, 2, 3, left to right, respectively.

3.3 Performance

Separate repeated measures ANOVAs were conducted for the Single-axis and Multi-Axis conditions as the roll component of the Single-axis condition was held constant and not applicable.

3.3.1 Single-axis

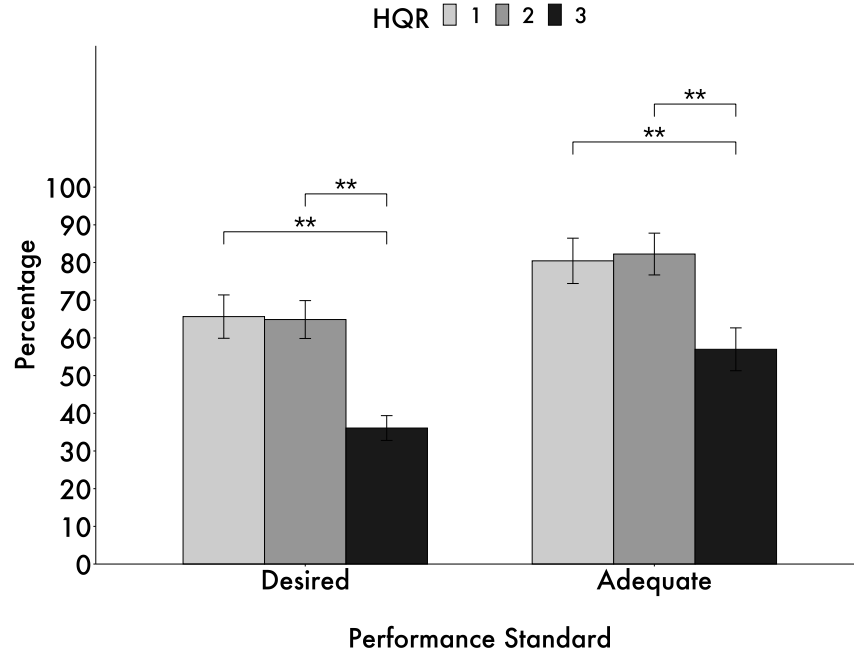


Figure 3.7: Mean Single-axis Performance by Condition and Level of Handling Quality. The Y-axis, Percentage, represents the percent time that participants met the desired and adequate performance standard.

A repeated measures ANOVA was performed to compare effects of the Single-axis condition as well as level of Handling Quality on participant performance. There were significant main effects for HQR for desired performance ($F_{(1.1,7.73)} = 27.83, p < 0.001$), and adequate performance ($F_{(2,14)} = 24.65, p < 0.001$). Pairwise comparisons for the Single-axis Condition revealed that the mean desired performance was significantly different in Level 1 ($M = 65.65 \pm 5.75$) vs. Level 3 ($M = 36.1 \pm 3.3$) HQR ($t_{(7)} = 5.12, p < 0.01, d = 1.81$), and in Level 2 ($M = 64.86 \pm 5.02$) vs. Level 3 ($M = 36.1 \pm 3.3$) HQR ($t_{(7)} = 5.74, p < 0.01, d = 2.03$). Pairwise comparisons for the Single-axis Condition revealed that the mean adequate performance was significantly different in Level 1 ($M = 89.7 \pm 5.25$) vs.

Level 3 ($M = 65.92 \pm 5.71$) HQR ($t_{(7)} = 4.9, p < 0.01, d = 1.73$), and in Level 2 ($M = 91.5 \pm 3.8$) vs. Level 3 ($M = 65.92 \pm 5.71$) HQR ($t_{(7)} = 5.5, p < 0.01, d = 1.93$).

3.3.2 Multi-axis

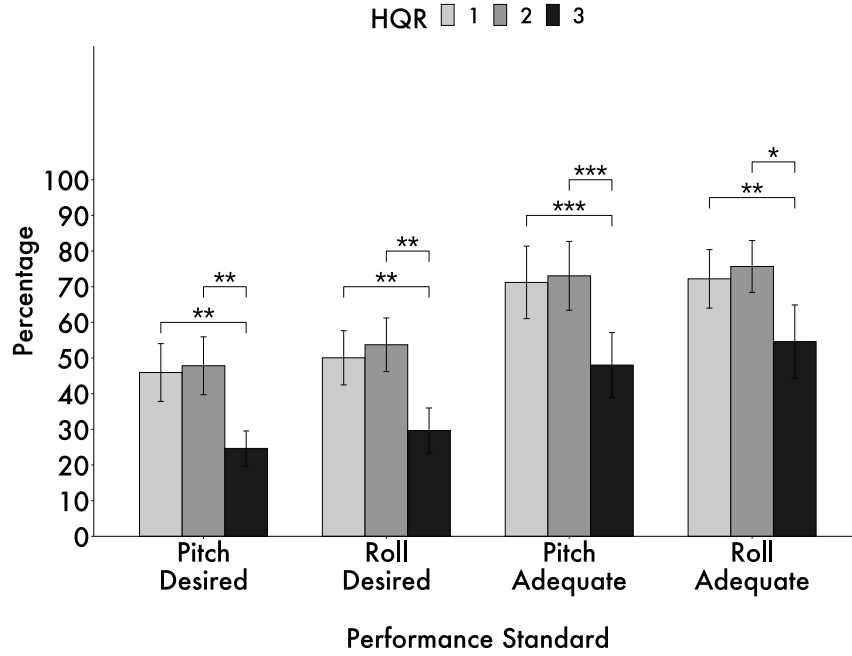


Figure 3.8: Mean Multi-axis Performance by Condition and Level of Handling Quality. The Y-axis, Percentage, represents the percent time that participants met the desired and adequate performance standard.

A repeated measures ANOVA was performed to compare effects of the Multi-axis condition as well as level of Handling Quality on participant performance. There were significant main effects for HQR for desired pitch performance ($F_{(1.1,7.8)} = 27.05, p < 0.01$), and desired roll performance ($F_{(2,14)} = 27.01, p < 0.001$). There were significant main effects for HQR for adequate pitch performance ($F_{(2,14)} = 24.12, p < 0.001$), and adequate roll performance ($F_{(1.18,8.3)} = 10.1, p < 0.01$).

Pairwise comparisons for the Multi-axis Condition revealed that the mean

desired pitch performance was significantly different in Level 1 ($M = 45.94 \pm 8.11$) vs. Level 3 ($M = 24.65 \pm 5$) HQR ($t_{(7)} = 4.91, p < 0.01, d = 1.73$), and in Level 2 ($M = 47.84 \pm 8.11$) vs. Level 3 ($M = 24.65 \pm 5$) HQR ($t_{(7)} = 5.7, p < 0.01, d = 2.02$). Pairwise comparisons for the Multi-axis Condition revealed that the mean desired roll performance was significantly different in Level 1 ($M = 50.1 \pm 7.6$) vs. Level 3 ($M = 29.67 \pm 6.32$) HQR ($t_{(7)} = 5.81, p < 0.001, d = 2.05$), and in Level 2 ($M = 53.72 \pm 7.52$) vs. Level 3 ($M = 29.67 \pm 6.32$) HQR ($t_{(7)} = 5.9, p < 0.001, d = 2.07$).

Pairwise comparisons for the Multi-axis Condition revealed that the mean adequate pitch performance was significantly different in Level 1 ($M = 71.21 \pm 10.16$) vs. Level 3 ($M = 48.03 \pm 9.12$) HQR ($t_{(7)} = 4.81, p < 0.01, d = 1.70$), and in Level 2 ($M = 73.1 \pm 9.7$) vs. Level 3 ($M = 48.03 \pm 9.12$) HQR ($t_{(7)} = 5.4, p < 0.01, d = 1.91$). Pairwise comparisons for the Multi-axis Condition revealed that the mean adequate roll performance was significantly different in Level 1 ($M = 72.21 \pm 8.21$) vs. Level 3 ($M = 54.62 \pm 10.3$) HQR ($t_{(7)} = 4.50, p < 0.01, d = 1.59$), and in Level 2 ($M = 75.67 \pm 7.3$) vs. Level 3 ($M = 54.62 \pm 10.3$) HQR ($t_{(7)} = 3.22, p < 0.05, d = 1.14$).

Table 3.1: Expected HQR vs. Subjective Rating Scales

	Dependent variable:					
	CHRexp		BWSexp		OVRexp	
	(1)	(2)	(3)	(4)	(5)	(6)
HQR1		-0.186*** (0.007)	-0.176*** (0.006)			0.028*** (0.009)
HQR2	0.186*** (0.007)			0.176*** (0.006)	-0.028*** (0.009)	
HQR3	0.851*** (0.007)	0.665*** (0.007)	0.384*** (0.006)	0.560*** (0.006)	0.776*** (0.009)	0.804*** (0.009)
Constant	1.526*** (0.005)	1.712*** (0.005)	1.684*** (0.004)	1.508*** (0.004)	1.470*** (0.006)	1.442*** (0.006)
Observations	34.016	34.016	34.016	34.016	32.472	32.472
R ²	0.318	0.318	0.236	0.236	0.242	0.242
Adjusted R ²	0.318	0.318	0.236	0.236	0.242	0.242
Residual Std. Error	0.535 (df = 34013)	0.535 (df = 34013)	0.420 (df = 34013)	0.420 (df = 34013)	0.654 (df = 32469)	0.654 (df = 32469)
F Statistic	7,929.080*** (df = 2; 34013)	7,929.080*** (df = 2; 34013)	5,258.181*** (df = 2; 34013)	5,258.181*** (df = 2; 34013)	5,182.144*** (df = 2; 32469)	5,182.144*** (df = 2; 32469)

Note:

*p<0.1; **p<0.05; ***p<0.01

3.4 Regressions

3.4.1 CHR-Brain Dynamics

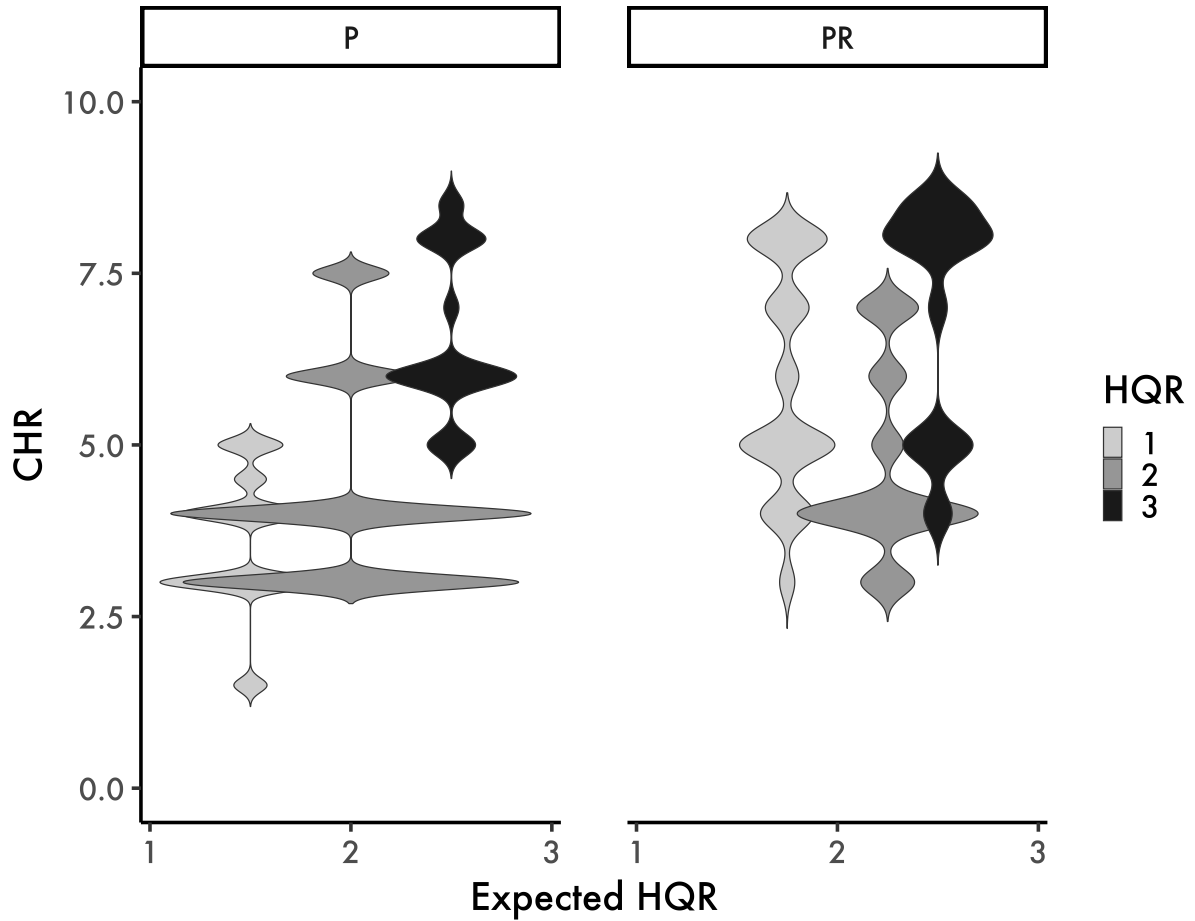


Figure 3.9: Actual vs. Expected Cooper-Harper Ratings. Y-axis represents the participants' Cooper-Harper ratings. X-axis represents the expected Cooper-Harper level rating.

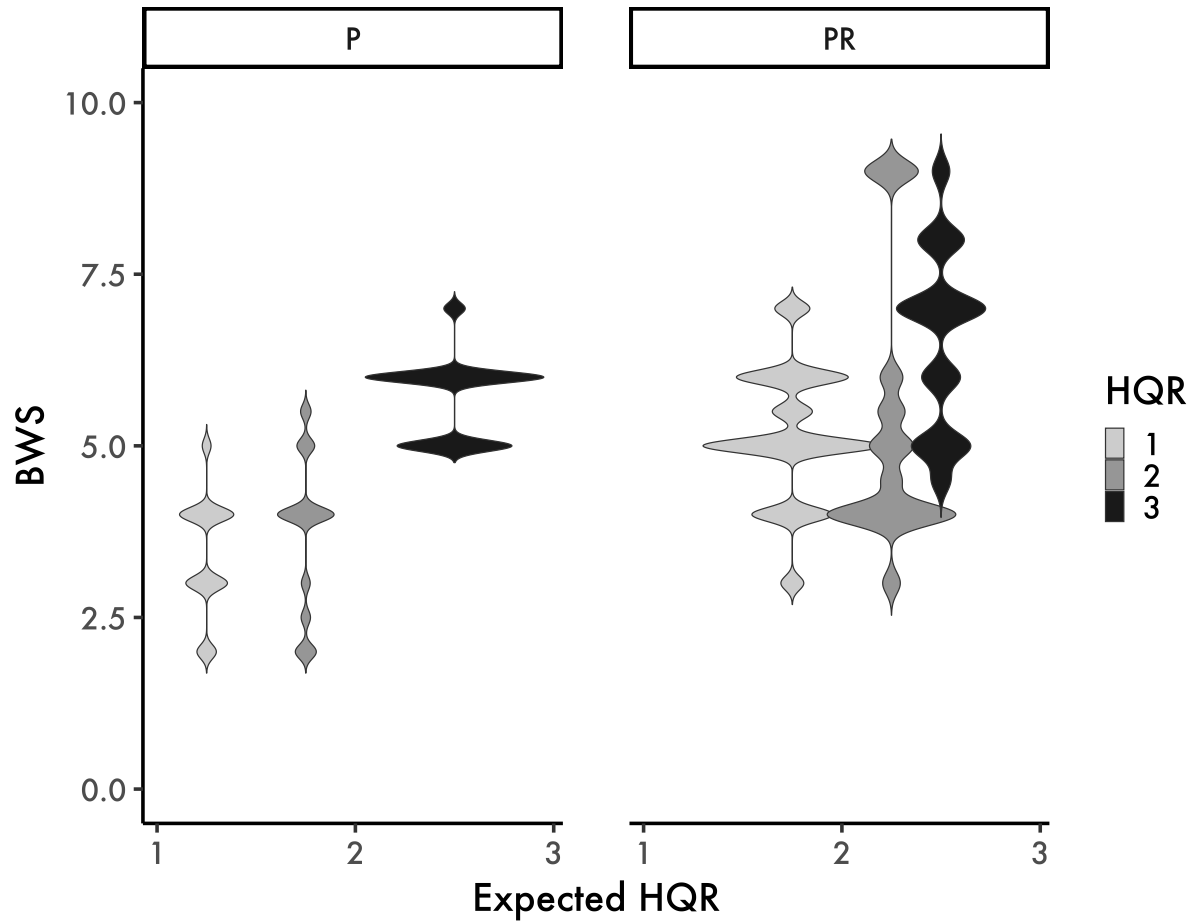


Figure 3.10: Actual vs. Expected Bedford Workload Ratings. Y-axis represents the participants' Bedford Workload ratings. X-axis represents the expected Cooper-Harper level rating.

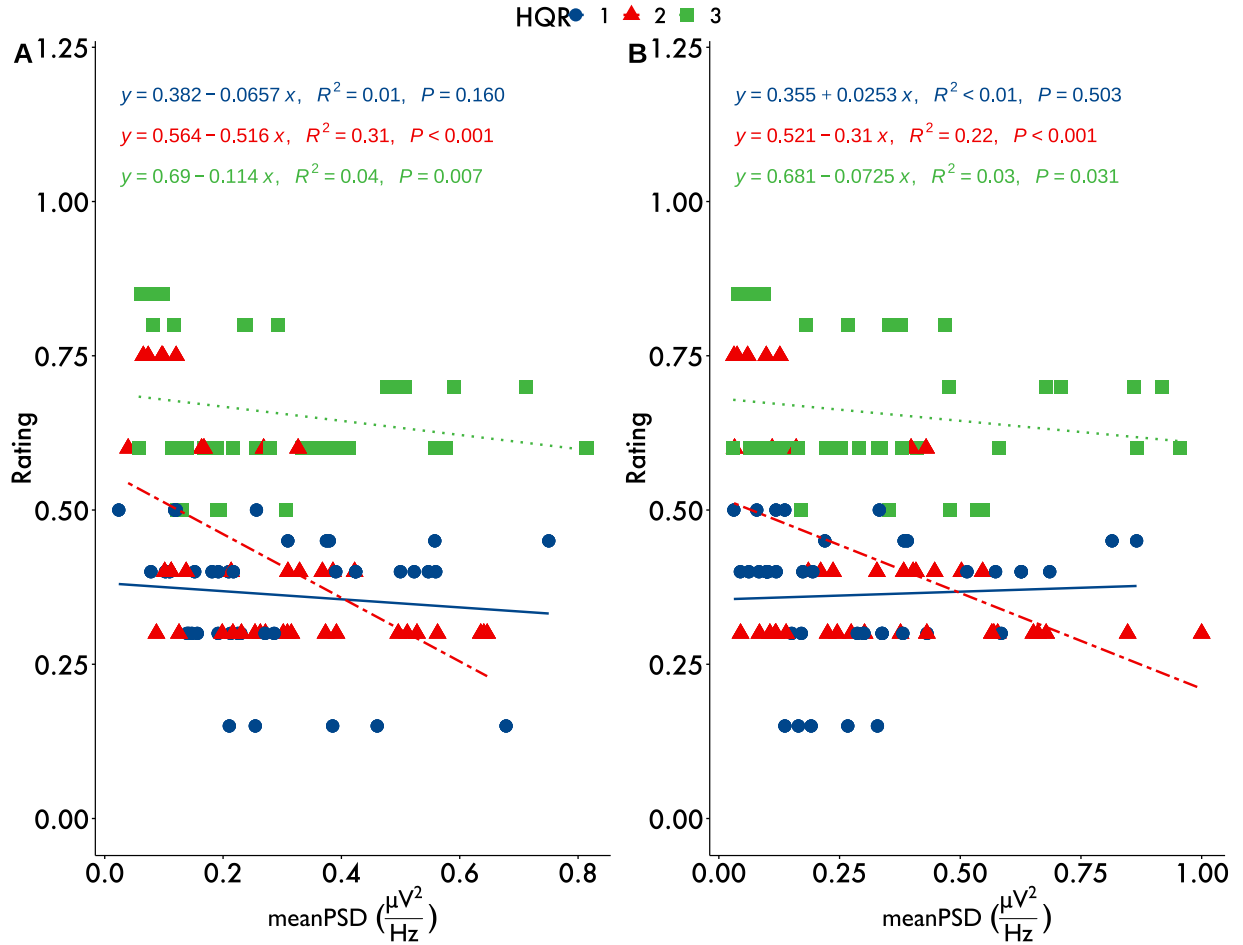


Figure 3.11: A: Normalized Cooper-Harper ratings regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. B: Normalized Cooper-Harper ratings regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. Both panels represent the Single-axis Condition. Scatterplots include five brain regions per participant.

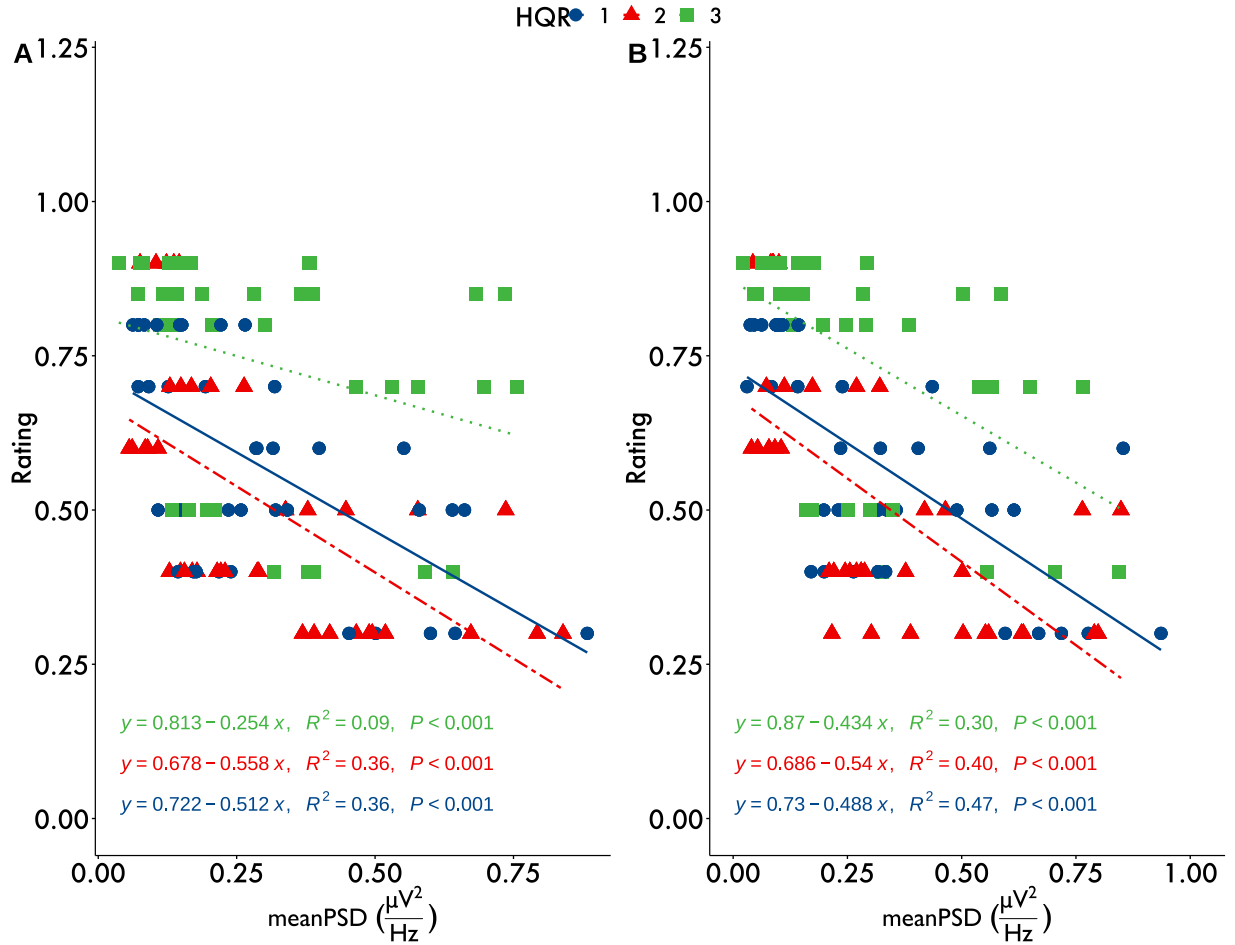


Figure 3.12: A: Normalized Cooper-Harper ratings regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. B: Normalized Cooper-Harper ratings regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. Both panels represent the Multi-axis Condition. Scatterplots include five brain regions per participant.

3.4.2 CHR-Behavioral Performance

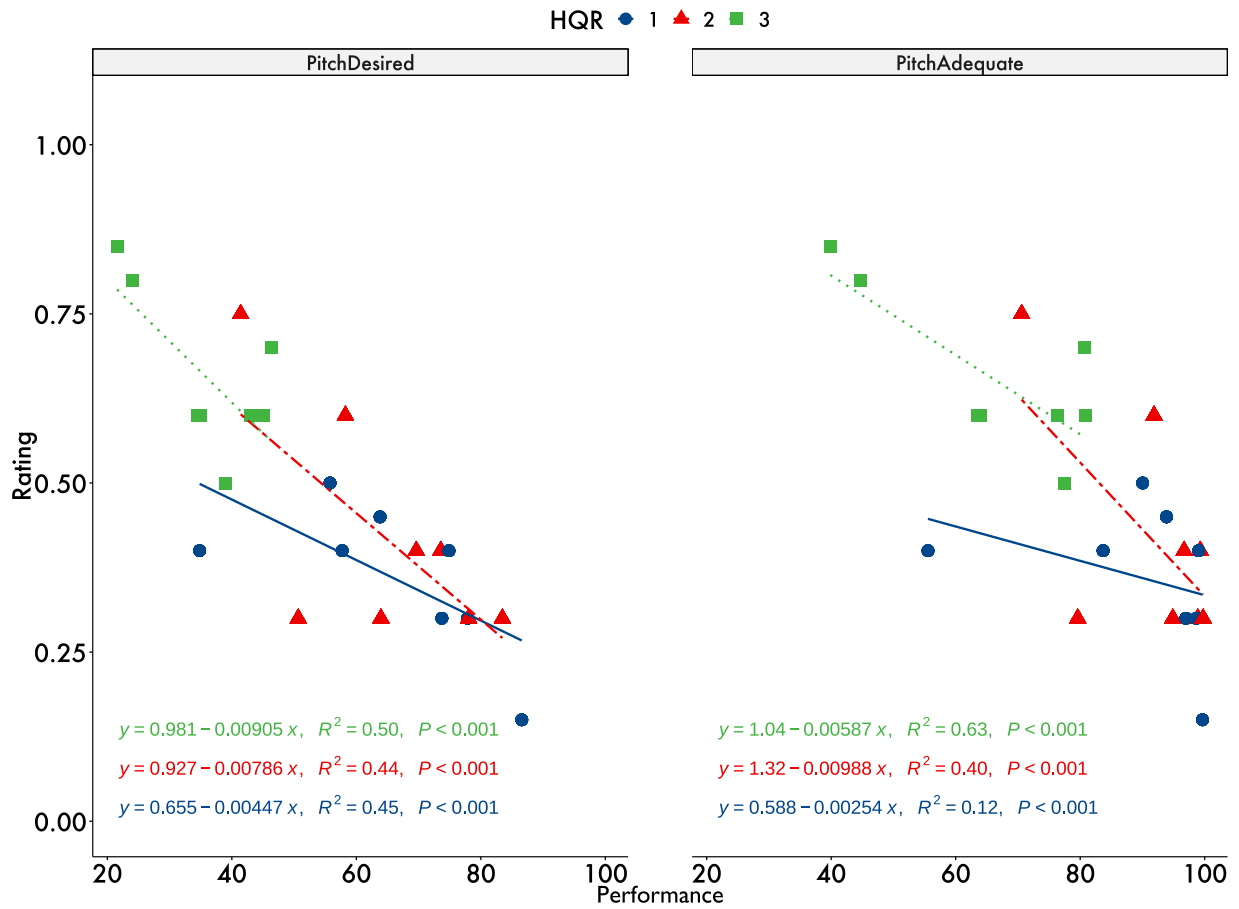


Figure 3.13: Normalized Cooper-Harper ratings regressed on desired and adequate performance for Levels 1, 2, and 3. Both panels represent the Single-axis Condition.

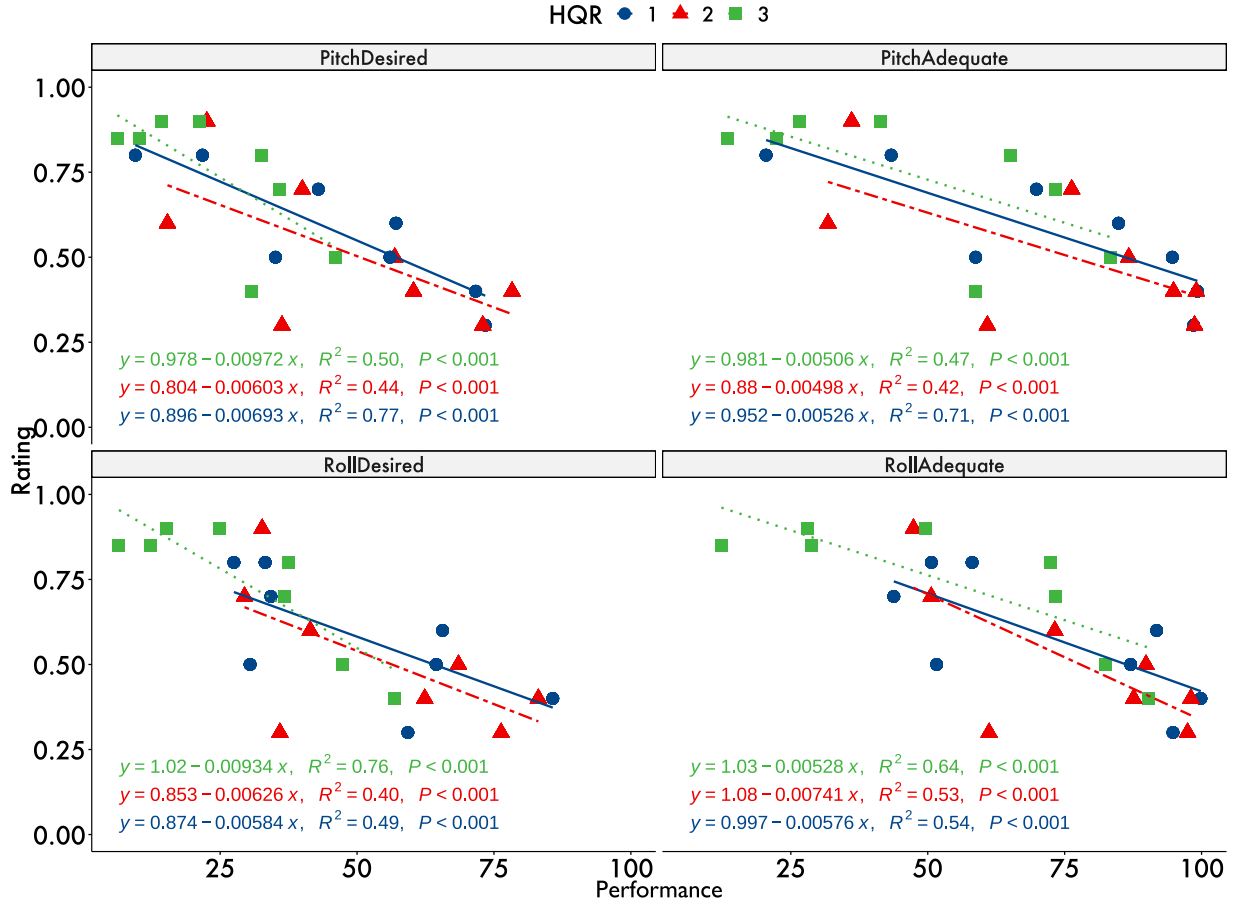


Figure 3.14: Top row represents normalized Cooper-Harper ratings regressed on desired and adequate pitch performance for Levels 1, 2, and 3. Bottom row represents normalized Cooper-Harper ratings regressed on desired and adequate roll performance for HQR Levels 1, 2, and 3. All panels represent the Multi-axis Condition.

3.4.3 Brain Dynamics-Behavioral Performance

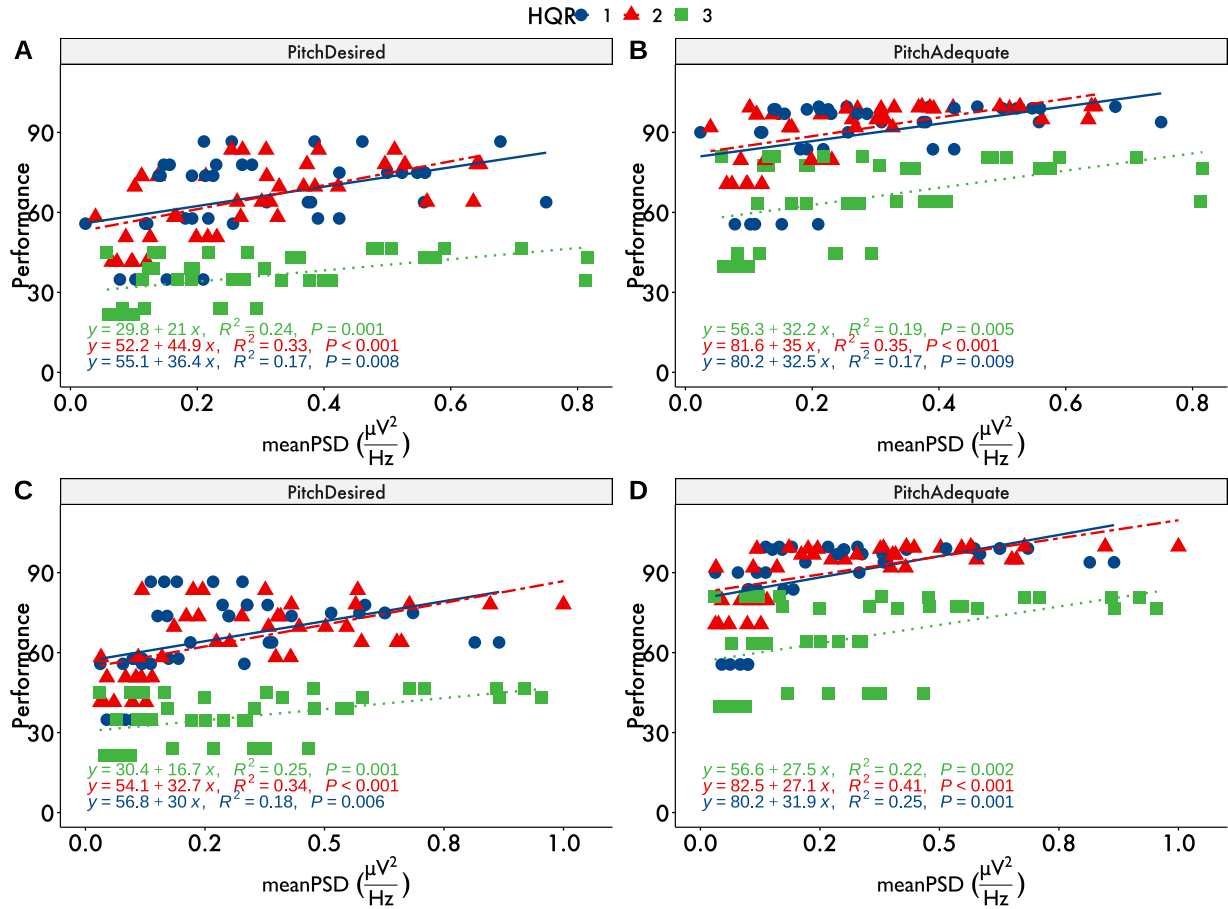


Figure 3.15: A: Percentage desired performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. B: Percentage adequate performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. C: Percentage desired performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. D: Percentage adequate performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. All panels represent the Single-axis Condition. Scatterplots include five brain regions per participant.

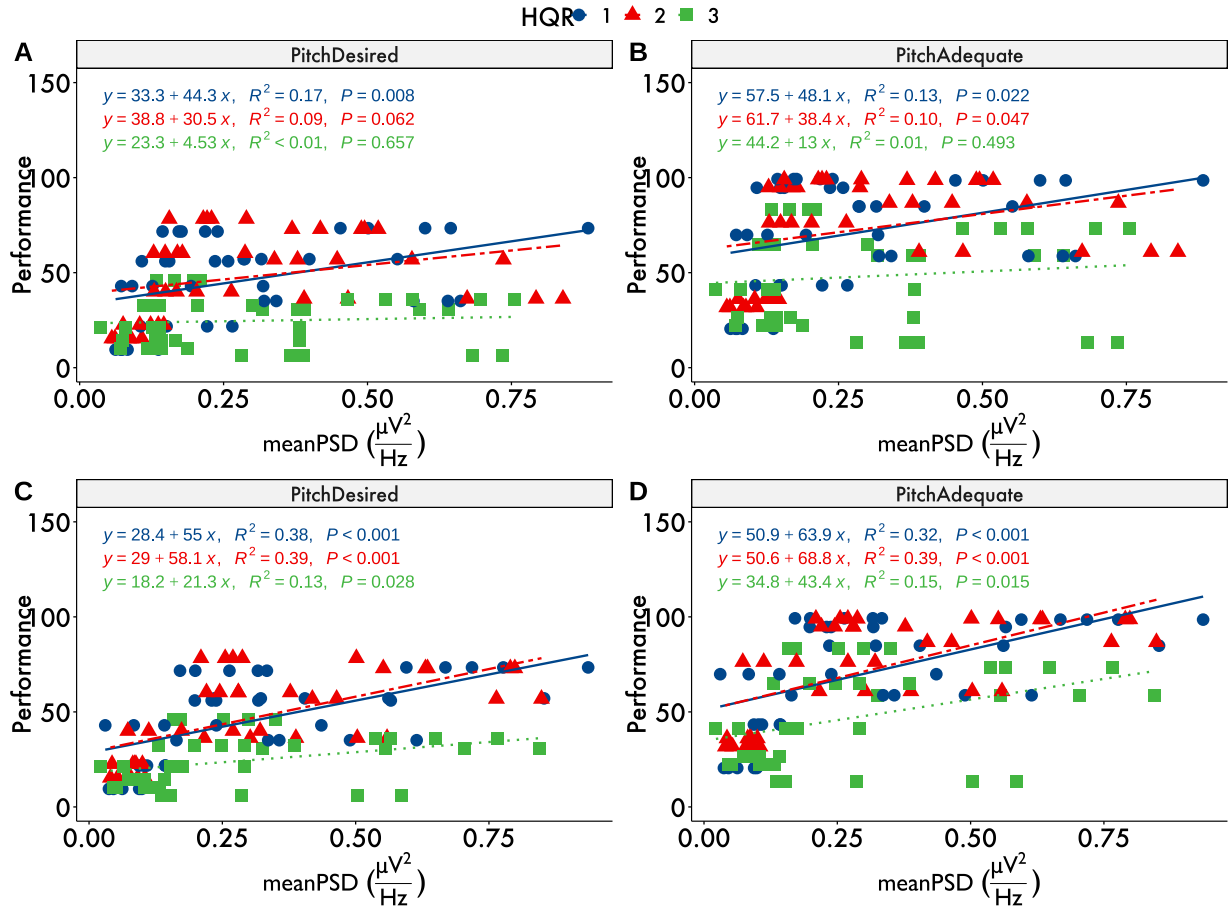


Figure 3.16: A: Percentage desired performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. B: Percentage adequate performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. C: Percentage desired performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. D: Percentage adequate performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. All panels represent the pitch component of the the Multi-axis Condition. Scatterplots include five brain regions per participant.

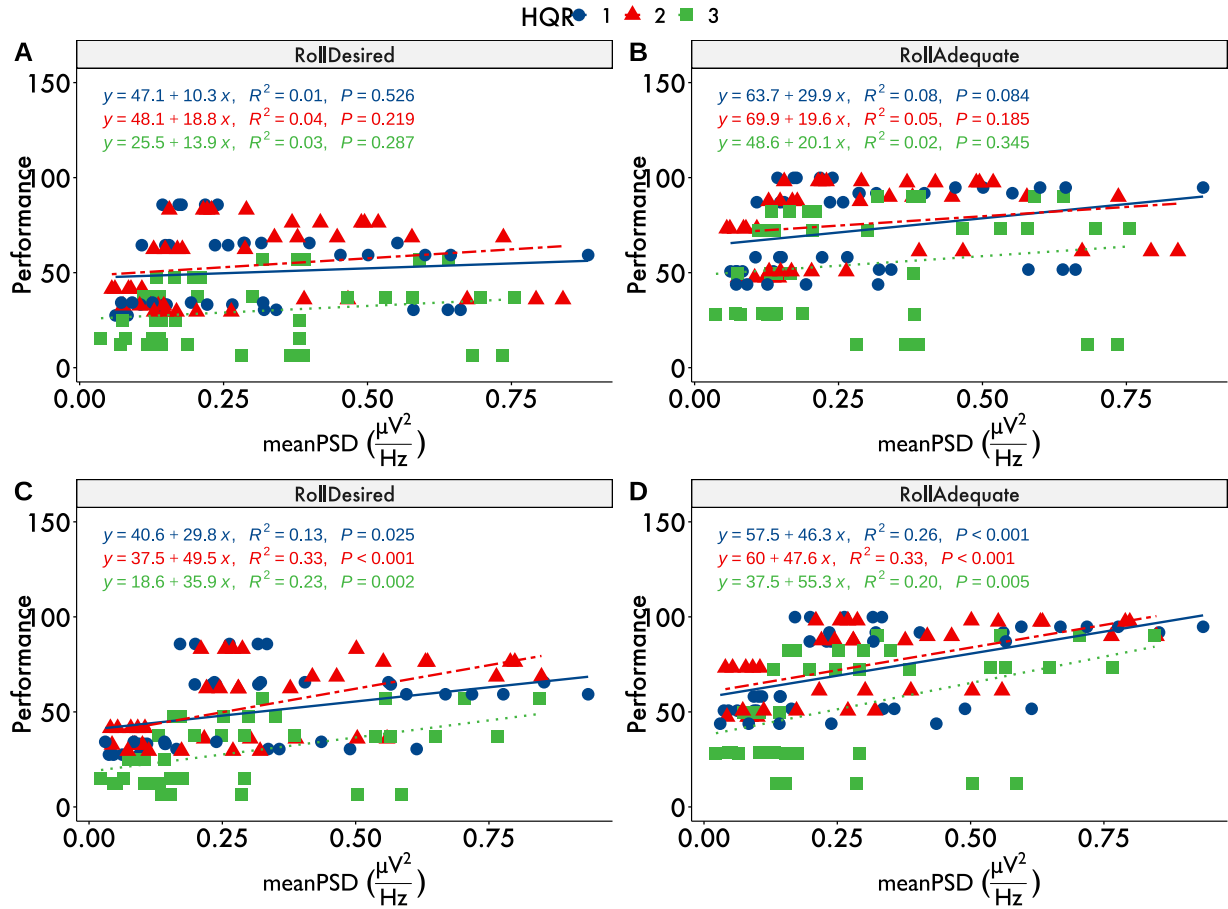


Figure 3.17: A: Percentage desired performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. B: Percentage adequate performance regressed on mean theta power spectral density for HQR Levels 1, 2, and 3. C: Percentage desired performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. D: Percentage adequate performance regressed on mean alpha power spectral density for HQR Levels 1, 2, and 3. All panels represent the Roll component of the the Multi-axis Condition. Scatterplots include five brain regions per participant.

3.4.4 Wavelet-based Regression

In all linear fixed-effects models, the inclusion of random-effects did not substantially lower the AIC and BIC. Therefore, the results included only the fixed-effects terms.

The maximum regional EEG activity within the decision-making period was statistically significant in estimating the CHR ($p - value = 0.0045954$, $AIC = -801.32$, $BIC = -772.61$, $LogLikelihood = 405.66$, $Deviance = -811.32$), while the mean regional EEG activity had a trend toward statistical significance ($p - value = 0.062518$, $AIC = -796.76$, $BIC = -768.05$, $LogLikelihood = 403.38$, $Deviance = -806.76$). Interestingly, the maximum regional EEG activity within the middle phase was also statistically significant in estimating the CHR with lower AIC and BIC ($p - value = 0.0061518$, $AIC = -800.8$, $BIC = -772.08$, $LogLikelihood = 405.4$, $Deviance = -810.8$), and the mean regional EEG activity had a trend toward statistical significance ($p - value = 0.071588$, $AIC = -796.54$, $BIC = -767.82$, $LogLikelihood = 403.27$, $Deviance = -806.54$). However, the other models were not statistically significant.

Section 4: Summary of Findings

Objective brain biomarkers are useful in the formation and prediction of perceived handling qualities ratings as a supplement to institutionalized subjective scales. Detailed findings listed below to support this conclusion.

Following the completion of all trials, participants were asked “Can you please comment on the general time or typical moment when you had realized what level of handling quality rating you would provide?” The responses below reveal the individual and dynamic nature related to the handling qualities rating process.

Selected responses:

- “I felt my performance influenced very heavily my handling quality choices. I used the middle portion of the trial for my decision.”
- “I decided early-on what level I thought it was. I would say I made my decision within the first five seconds.”
- “I reflected on the whole trial as an average for my rating.”
- “I used the latter half of the trials for my decisions. During the first half, I was checking how the system responded and then in the second half, I

made my decision.”

- Absence of main effects for Condition in CHR and BWS implies veridical ratings of the Naval Aviators. In other words, the participants’ perceptions of the handling qualities within each Level were consistent across Conditions.
- As expected, participants rated worse performance in in Level 3 for both Conditions.
- As expected, performance declined under the highest conditions of challenge, relative to that observed in Levels 1-2, which were undifferentiated. A similar pattern was observed for the Multi-axis Condition for both pitch and roll.
- A similar pattern of findings was revealed for the Bedford Workload ratings to that reported above.
- Although the $\frac{\theta}{\alpha}$ Ratio was able to detect differences in Level of handling quality, specific EEG band power spectral density was not (see Figures 3.5 and 3.6). Generally, averaging within a frequency band over the entire testing period is likely not sensitive to capture dynamic workload. Therefore, spectro-temporal analyses based on continuous wavelet transform were employed (refer to Section 2.5.1).
- Actual reported CHR compared to the expected were generally aligned, but the lack of clear discrimination between levels, particularly for the Multit-

axis Condition, is likely explained by participant experience (see Figure 3.9).

Regression Analyses:

- CHR regressed on Performance:
 - There was a negative directional relationship between CHR and performance. That is, elevated performance was associated with lower CHR handling ratings for both Single- and Multi-axis Conditions.
- Performance regressed on EEG band power:
 - There was a positive directional relationship between performance and both EEG theta and alpha power spectral density for both the Single- and Multi-axis Conditions. That is, elevated performance was associated with elevated EEG band power for both Single- and Multi-axis Conditions.
 - These findings are reasonable given that theta power is indicative of enhanced attentional engagement and alpha power is indicative of cortical relaxation/inhibition of non-essential neural processes.
- CHR regressed on EEG band power:
 - There was a negative directional relationship between CHR and both EEG theta and alpha power spectral density for both the Single- and Multi-axis Conditions. That is, lower CHR ratings were associated

with elevated EEG band power for both Single- and Multi-axis Conditions.

- These findings are reasonable given that theta power is indicative of enhanced attentional engagement and alpha power is indicative of cortical relaxation/inhibition of non-essential neural processes.
- Wavelet-based regression analysis:
 - During the early phase (i.e., 0-20s), the CHR is not related to participants' objective biomarkers of handling qualities ratings. This is likely due to the participants needing some additional time to evaluate the system.
 - The middle phase (i.e., 21-40s) is significantly associated with participants' objective biomarkers of handling qualities which implies that participants actively used the ongoing trial in their decision-making process.
 - The late phase (i.e., 41-60s) is not significantly related to participants' objective biomarkers of handling qualities. This is likely due to the participants having already evaluated the system and made their decision.
 - While these results reveal a general tendency, they also speak to the individual differences in the handling qualities rating process and support the utility of the objective biomarkers such as the EEG time-series.

Appendix A: Supplemental Figures

A.1 Questionnaires and Survey Tools

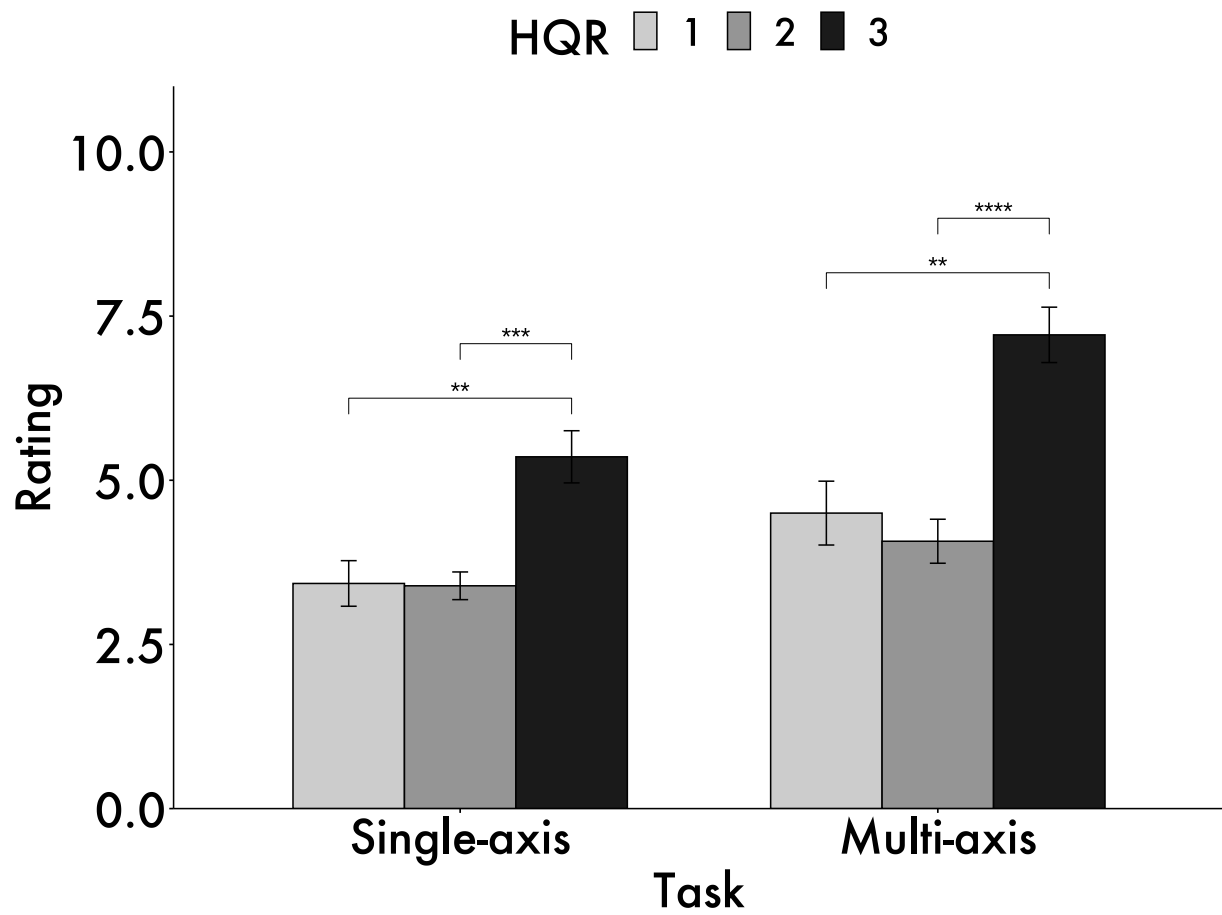


Figure A.1: NROTC Mean Bedford Workload rating by Condition and Level of Handling Quality.

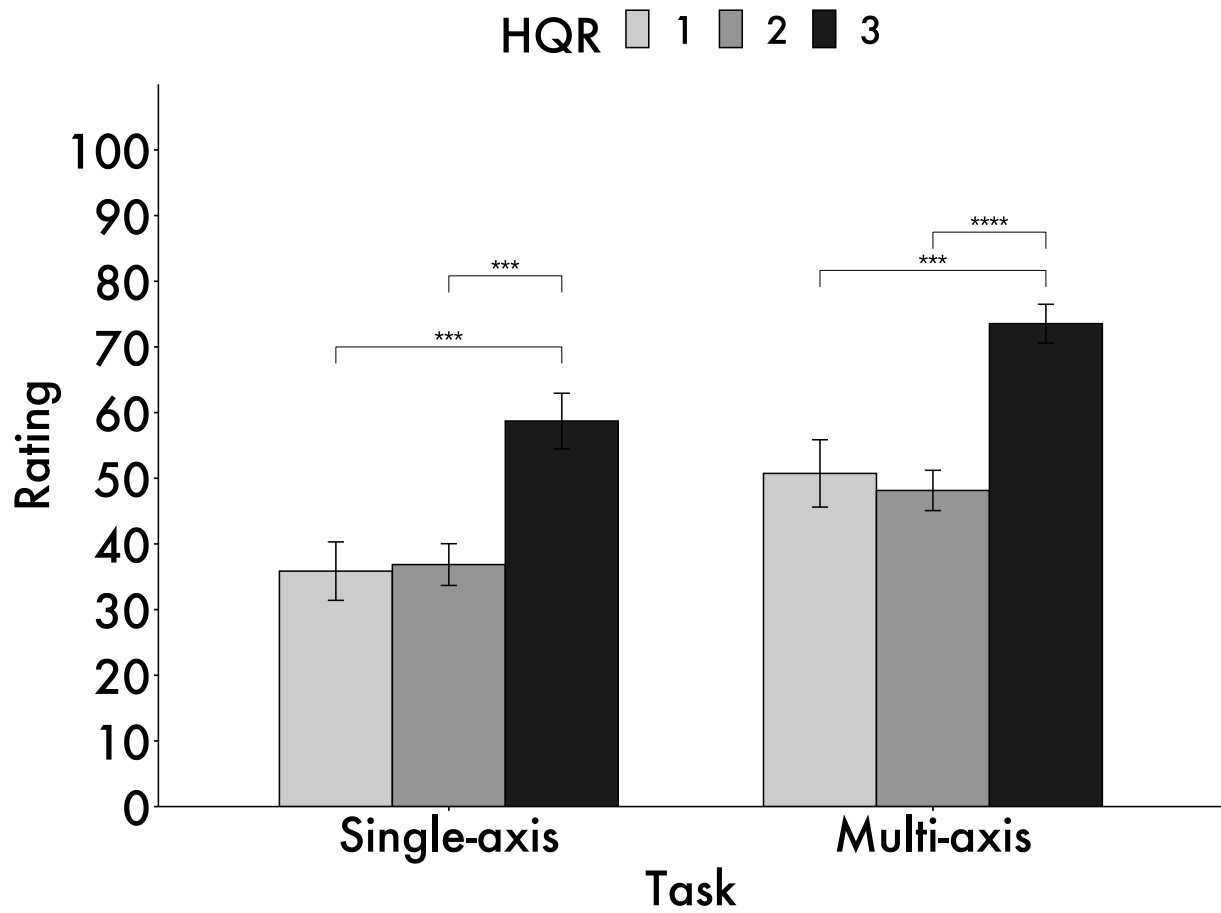


Figure A.2: NROTC Mean weighted NASA-TLX Overall workload rating by Condition and Level of Handling Quality.

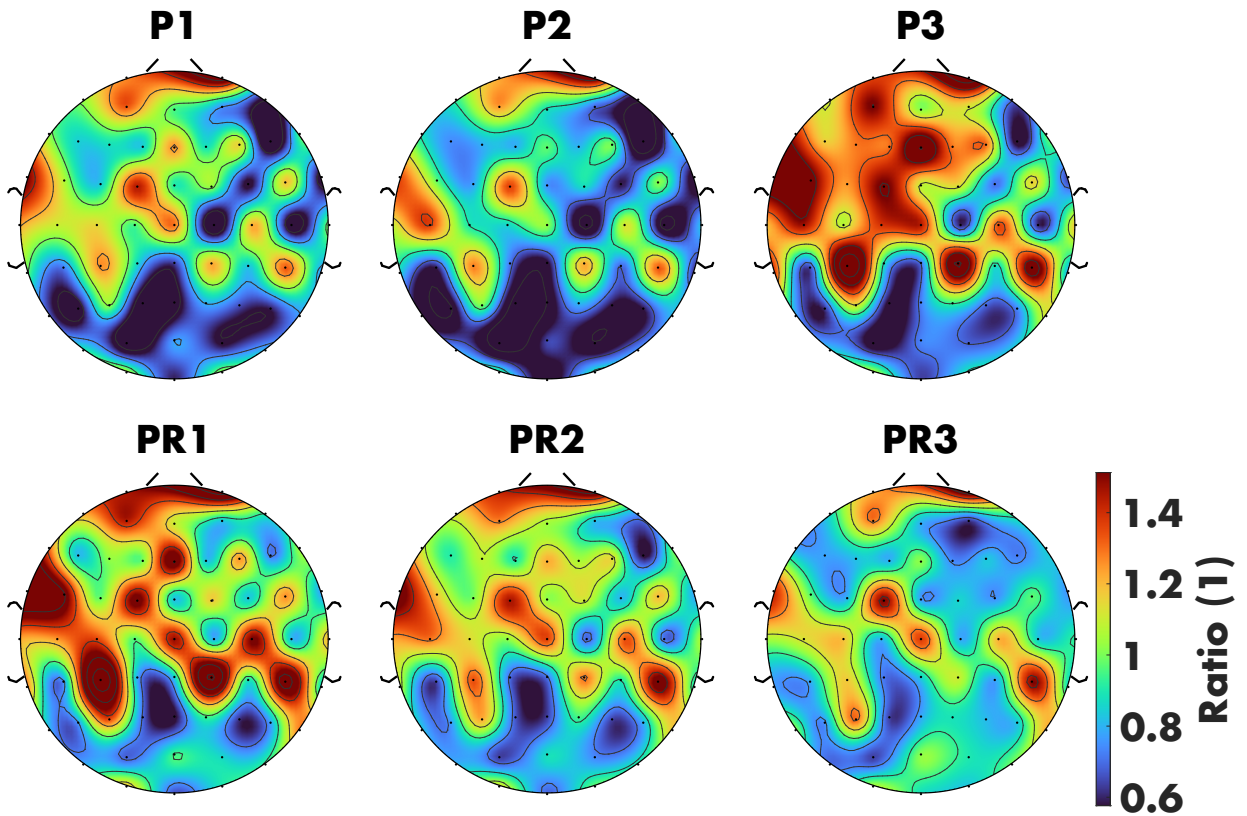


Figure A.3: Scalp topographical maps of NROTC participants for $\frac{\theta}{\alpha}$ Ratio. The top row represents the Single-axis Condition, Levels 1, 2, 3, left to right, respectively. The bottom row represents the Multi-axis Condition, Levels 1, 2, 3, left to right, respectively.

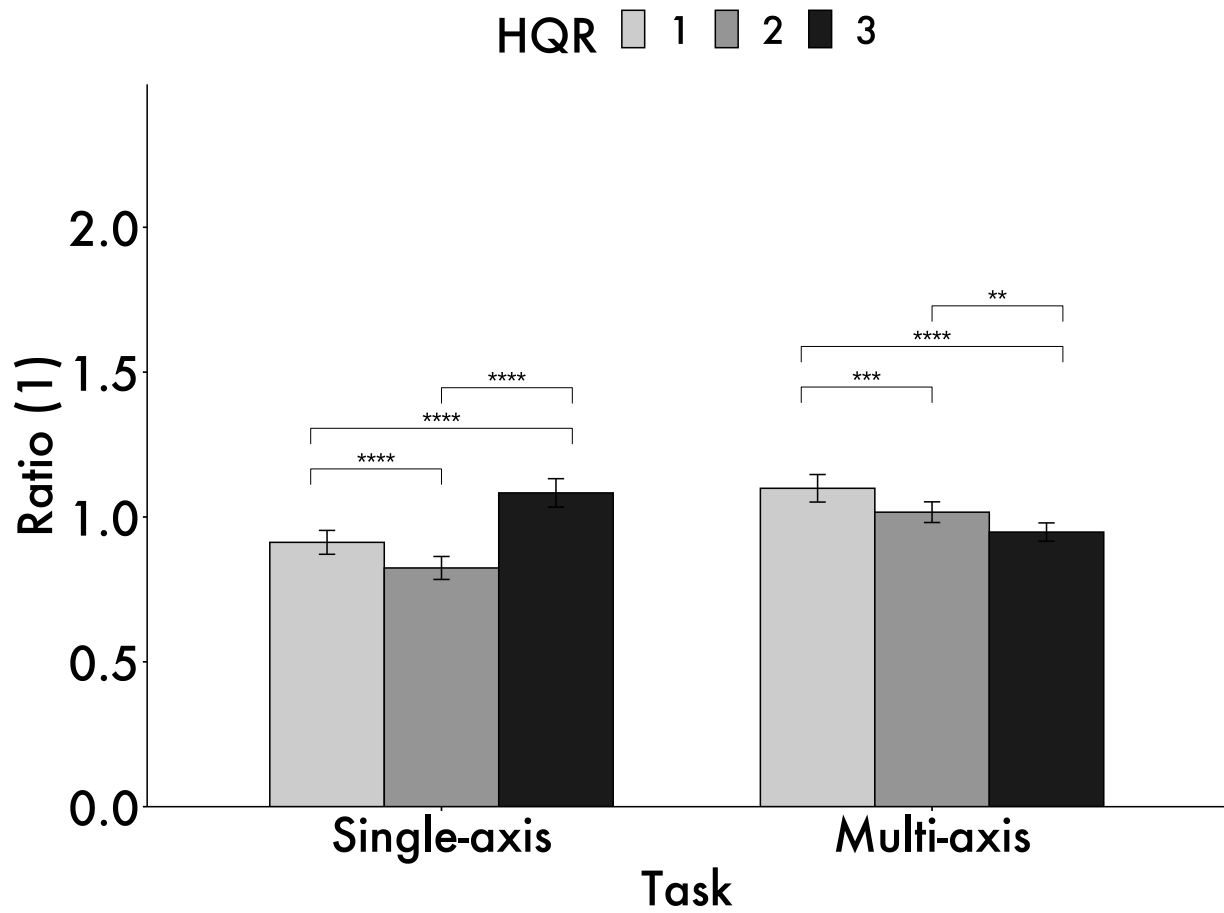


Figure A.4: NROTC Mean $\frac{\theta}{\alpha}$ Ratio by Condition and Level of Handling Quality.

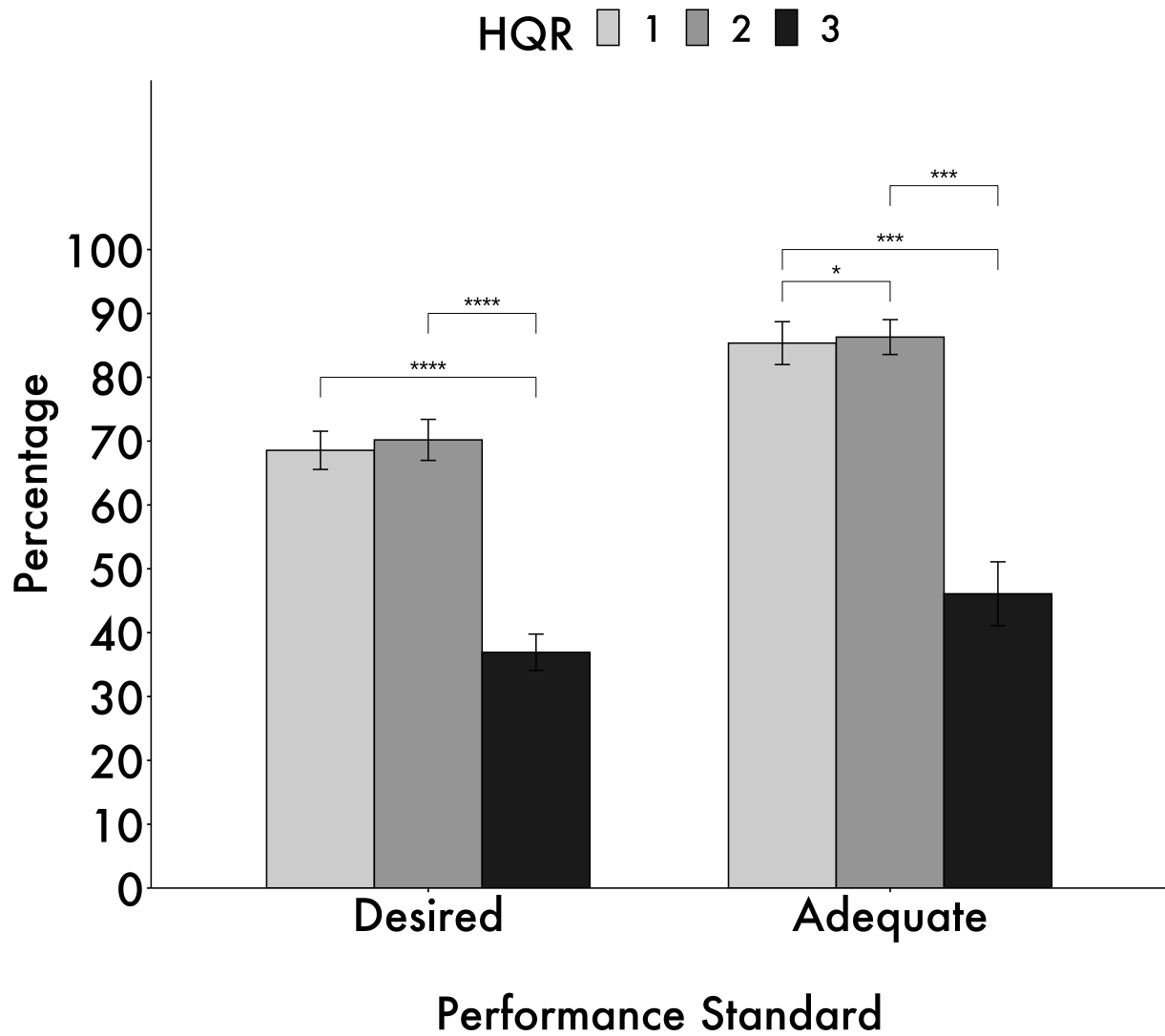


Figure A.5: NROTC Single-axis mean desired and adequate performance by Condition and Level of Handling Quality.

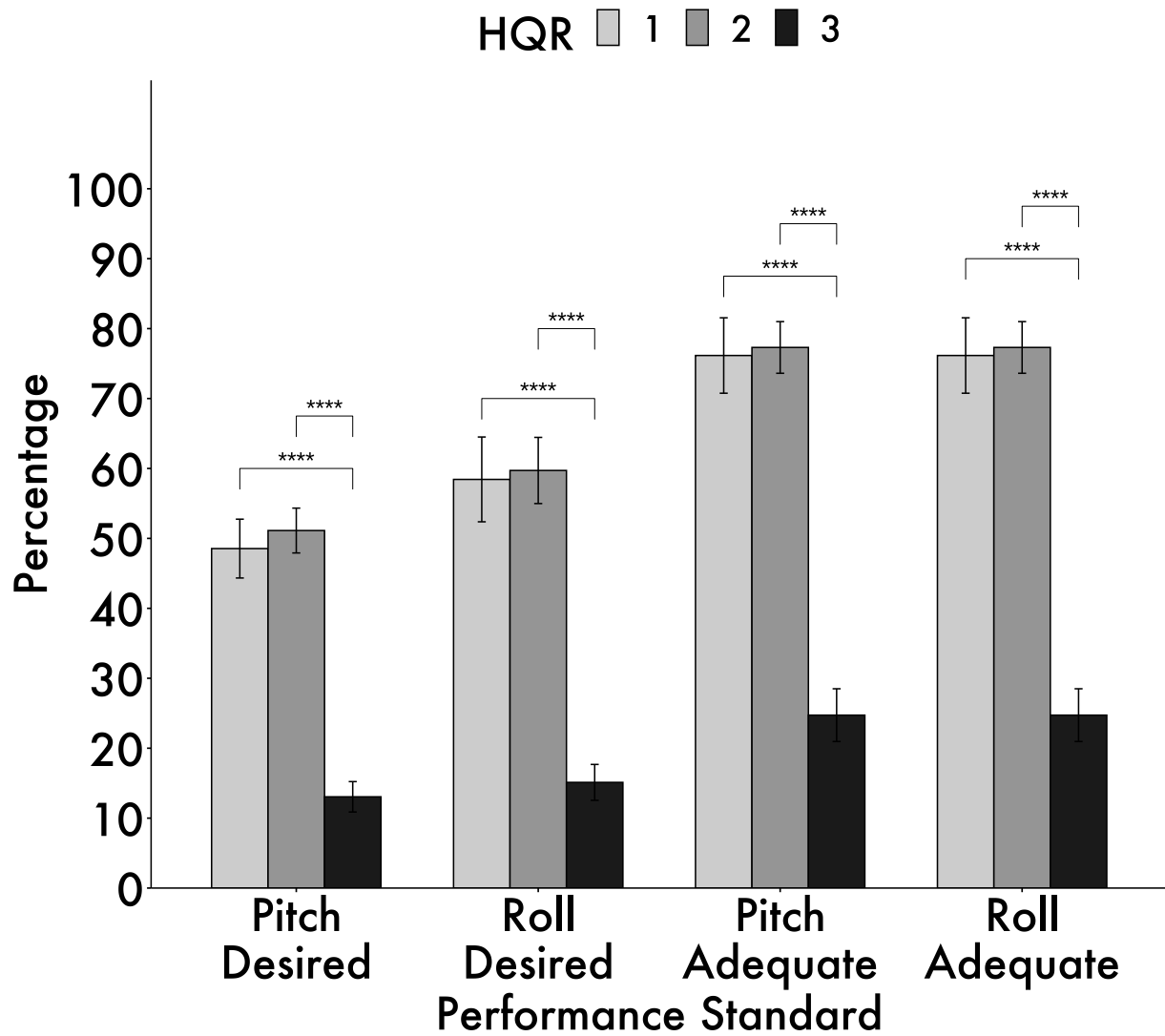


Figure A.6: NROTC Multi-axis mean desired and adequate performance by Condition and Level of Handling Quality.

Subject ID: NA _____

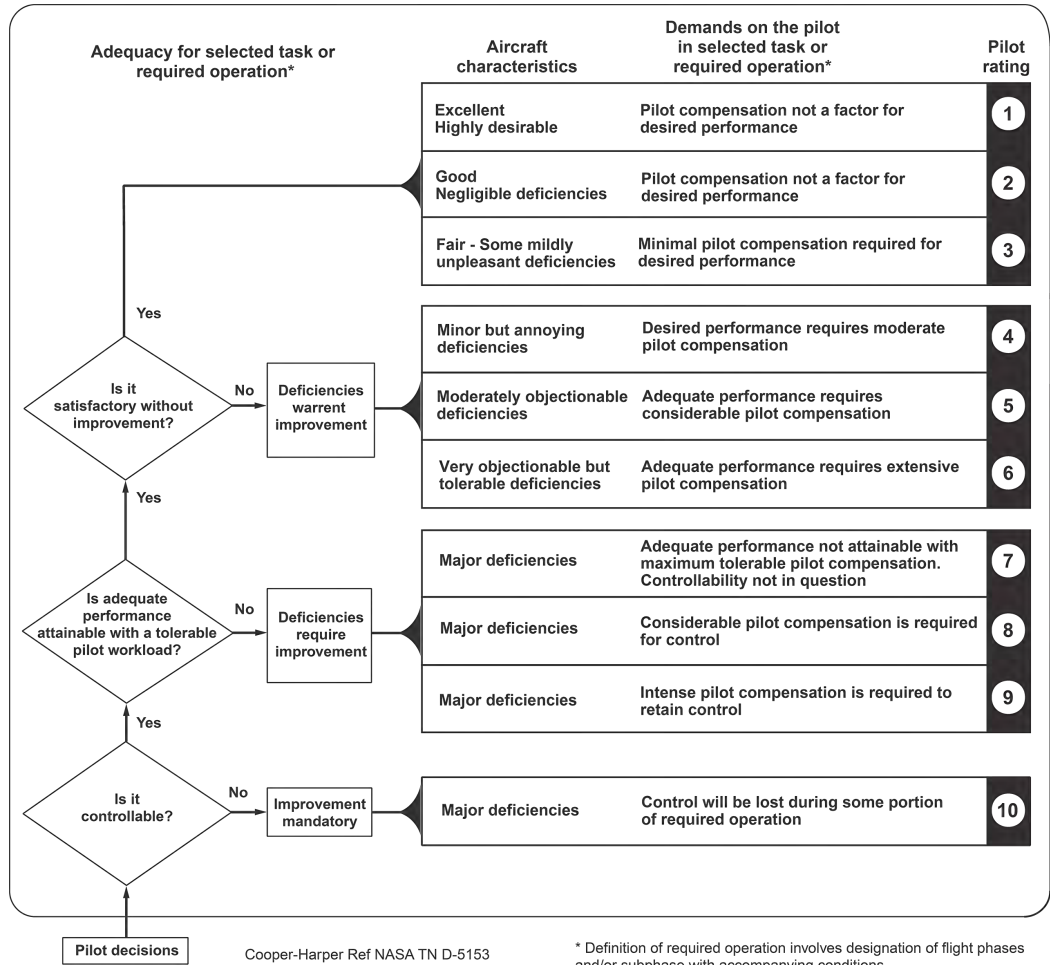


Figure A.7: The Cooper-Harper Rating Scale (Cooper & Harper, 1969)

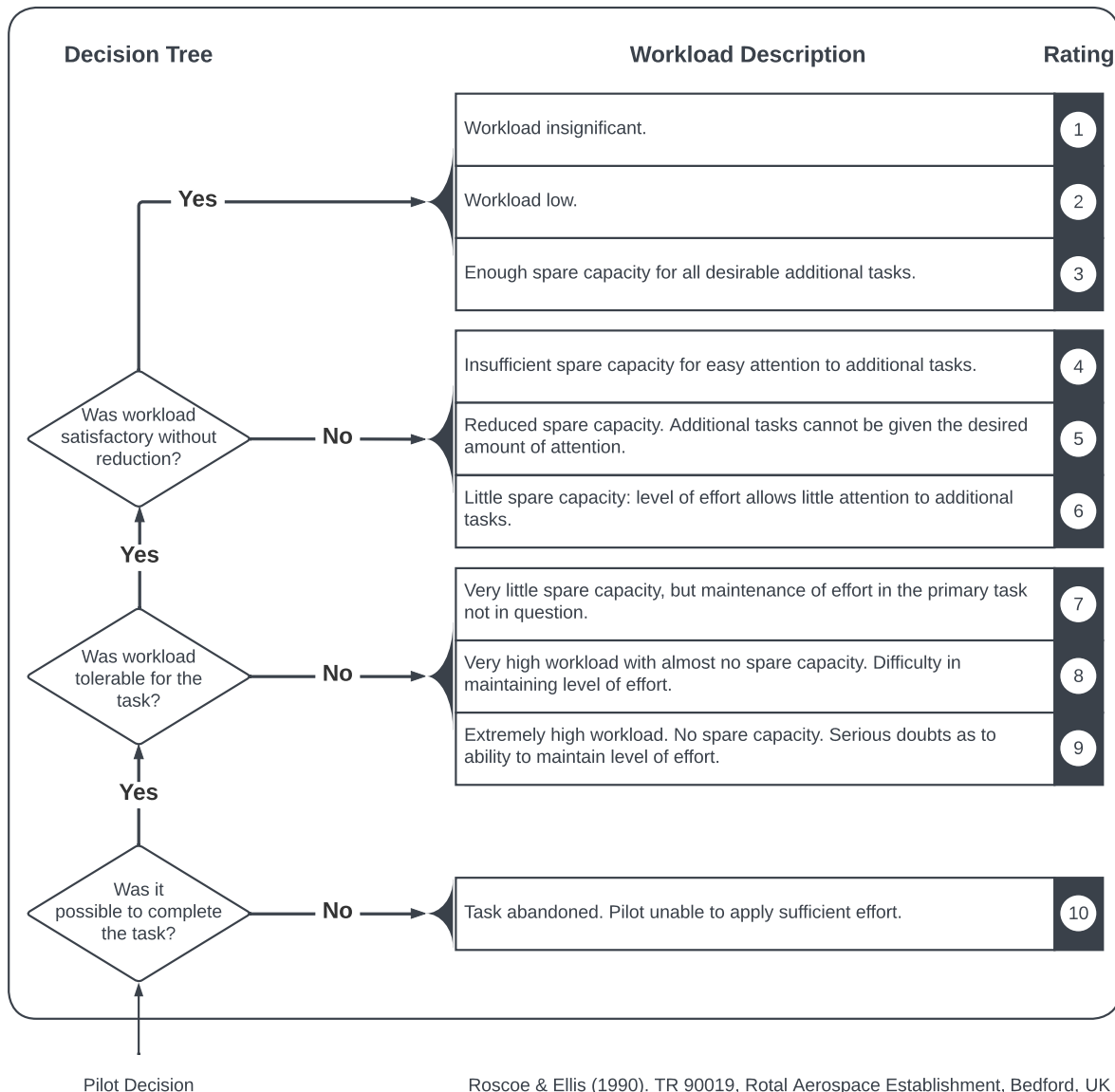
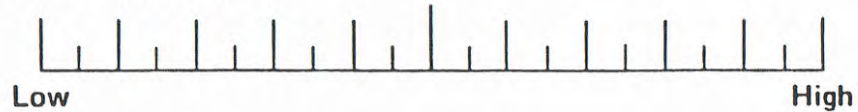


Figure A.8: The 'Bedford' Workload Scale (Roscoe & Ellis, 1990)

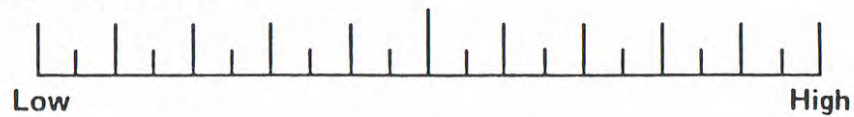
Subject ID: _____ Task ID: _____

RATING SHEET

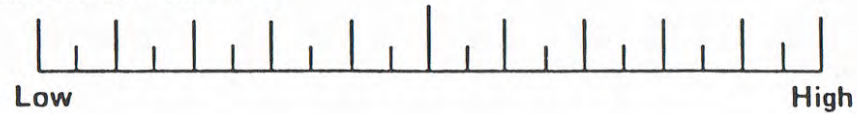
MENTAL DEMAND



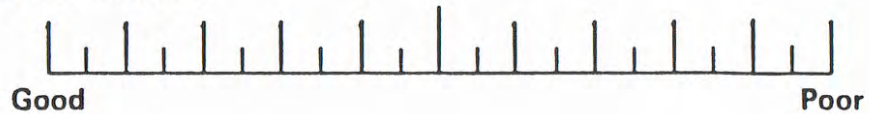
PHYSICAL DEMAND



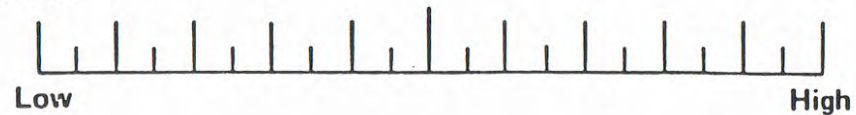
TEMPORAL DEMAND



PERFORMANCE



EFFORT



FRUSTRATION

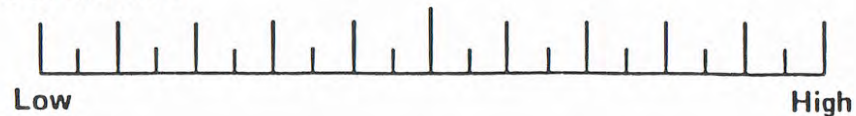


Figure A.9: The National Aeronautics and Space Administration Task Load Index (Hart & Staveland, 1988)

References

- Andreassi, J. L. (2006). *Psychophysiology: Human behavior & physiological response*. New York, NY, Psychology press. <https://doi.org/10.4324/9780203880340>
- Bacon, S. J. (1974). Arousal and the range of cue utilization. *Journal of Experimental Psychology*, 102(1), 81–87. <https://doi.org/10.1037/h0035690>
- Baker, S. N. (2007). Oscillatory interactions between sensorimotor cortex and the periphery. *Current Opinion in Neurobiology*, 17(6), 649–655. <https://doi.org/10.1016/j.conb.2008.01.007>
- Başar-Eroglu, C., Başar, E., Demiralp, T., & Schürmann, M. (1992). P300-response: Possible psychophysiological correlates in delta and theta frequency channels. a review. *International journal of psychophysiology*, 13(2), 161–179. [https://doi.org/10.1016/0167-8760\(92\)90055-g](https://doi.org/10.1016/0167-8760(92)90055-g)
- Baskett, B. J. (2000). *Aeronautical design standard performance specification handling qualities requirements for military rotorcraft* (Report). Army Aviation and Missile Command Redstone Arsenal, Alabama.

- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>
- Blanken, C. L., Hoh, R. H., Mitchell, D. G., & Key, D. L. (2008). *Test guide for ads-33e-prf* (Report). HOH AERONAUTICS INC LOMITA CA.
- Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review*, 64(3), 205–15. <https://doi.org/10.1037/h0047313>
- Brown, I. (1964). The measurement of perceptual load and reserve capacity. *Occupational Medicine*, 14(1), 44–49. <https://doi.org/10.1093/occmed/14.1.44>
- Busk, J., & Galbraith, G. C. (1975). EEG correlates of visual-motor practice in man. *Electroencephalography and Clinical Neurophysiology*, 38(4), 415–22. [https://doi.org/10.1016/0013-4694\(75\)90265-5](https://doi.org/10.1016/0013-4694(75)90265-5)
- Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, 74, 221–232. <https://doi.org/10.1016/j.apergo.2018.08.028>
- Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked eeg activities. *American Journal of EEG Technology*, 25(2), 83–92. <https://doi.org/10.1080/00029238.1985.11080163>
- Chen, X.-J., & Kwak, Y. (2022). Contribution of the sensorimotor beta oscillations and the cortico-basal ganglia-thalamic circuitry during value-

- based decision making: A simultaneous eeg-fmri investigation. *NeuroImage*, 257, 119300. <https://doi.org/https://doi.org/10.1016/j.neuroimage.2022.119300>
- Cooper, G. E., & Harper, R. P. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities*. National Aeronautics; Space Administration.
- Deeny, S. P., Haufler, A. J., Saffer, M., & Hatfield, B. D. (2009). Electroencephalographic coherence during visuomotor performance: A comparison of cortico-cortical communication in experts and novices. *Journal of Motor Behavior*, 41(2), 106–16. <https://doi.org/10.3200/JMBR.41.2.106-116>
- Deeny, S. P., Hillman, C. H., Janelle, C. M., & Hatfield, B. D. (2003). Cortico-cortical communication and superior performance in skilled marksmen: An eeg coherence analysis. *Journal of Sport and Exercise Psychology*, 25(2), 188–204.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- de Waard, D. (1996). *The measurement of drivers' mental workload* (Doctoral dissertation). University of Groningen. s.n.
- DiDomenico, A., & Nussbaum, M. A. (2008). Interactive effects of physical and mental workload on subjective workload assessment. *International journal of industrial ergonomics*, 38(11-12), 977–983.

- Dyke, F., Godwin, M. M., Goel, P., Rehm, J., Rietschel, J. C., Hunt, C. A., & Miller, M. W. (2014). Cerebral cortical activity associated with non-experts' most accurate motor performance. *Human Movement Science*, 37, 21–31. <https://doi.org/doi.org/10.1016/j.humov.2014.06.008>
- Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 66(3), 183–201. <https://doi.org/10.1037/h0047707>
- Engel, A. K., & Fries, P. (2010). Beta-band oscillations—signalling the status quo? *Current Opinion in Neurobiology*, 20(2), 156–165. <https://doi.org/https://doi.org/10.1016/j.conb.2010.02.015>
- Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00154>
- Gaillard, A. W. (1993). Comparing the concepts of mental load and stress. *Ergonomics*, 36(9), 991–1005. <https://doi.org/10.1080/00140139308967972>
- Gentili, R. J., Jaquess, K. J., Shuggi, I. M., Shaw, E. P., Oh, H., Lo, L. C., Tan, Y. Y., Domingues, C. A., Blanco, J. A., Rietschel, J. C., Miller, M. W., & Hatfield, B. D. (2018). Combined assessment of attentional reserve and cognitive-motor effort under various levels of challenge with a dry eeg system. *Psychophysiology*, 55(6), e13059. <https://doi.org/10.1111/psyp.13059>
- Gevins, A. (1997). High-resolution eeg mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and prac-

tice. *Cerebral Cortex*, 7(4), 374–385. <https://doi.org/10.1093/cercor/7.4.374>

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, 4(1-2), 113–131. <https://doi.org/10.1080/14639220210159717>

Hancock, G. M., Longo, L., Young, M. S., & Hancock, P. A. (2021). Mental workload. In G. Salvendy & W. Karwowski (Eds.), *Handbook of human factors and ergonomics* (pp. 203–226). Hoboken, NJ, John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119636113.ch7>

Hancock, P. A. (1989). The effect of performance failure and task demand on the perception of mental workload. *Applied Ergonomics*, 20(3), 197–205. [https://doi.org/https://doi.org/10.1016/0003-6870\(89\)90077-X](https://doi.org/https://doi.org/10.1016/0003-6870(89)90077-X)

Hancock, P. A., & Szalma, J. L. (2006). Stress and neuroergonomics. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work*. Oxford, Oxford University Press.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31(5), 519–537. <https://doi.org/10.1177/001872088903100503>

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience*, 7, 83. <https://doi.org/10.3389/fnint.2013.00083>

- Harper Jr., R. P., & Cooper, G. E. (1986). Handling qualities and pilot evaluation. *Journal of Guidance, Control, and Dynamics*, 9(5), 515–529. <https://doi.org/10.2514/3.20142>
- Hart, S. G. (1986). Nasa task load index (tlx).
- Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Human mental workload*. (pp. 139–183). Oxford, England, North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hatfield, B. D., Costanzo, M. E., Goodman, R. N., Lo, L. C., Oh, H., Rietschel, J. C., Saffer, M., Bradberry, T., Contreras-Vidal, J., & Haufler, A. (2013). The influence of social evaluation on cerebral cortical activity and motor performance: A study of "real-life" competition. *International Journal of Psychophysiology*, 90(2), 240–9. <https://doi.org/10.1016/j.ijpsycho.2013.08.002>
- Hatfield, B. D., & Hillman, C. H. (2001). The psychophysiology of sport: A mechanistic understanding of the psychology of superior performance. In R. N. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *The handbook of sport psychology* (pp. 362–386). New York, NY, John Wiley.
- Hatfield, B. D., Jaquess, K. J., Lo, L.-C., & Oh, H. (2020). The cognitive and affective neuroscience of superior athletic performance. In *Handbook of sport psychology* (pp. 487–512). [https://doi.org/doi.org/10.1002/9781119568124.ch23](https://doi.org/10.1002/9781119568124.ch23)

- Hatfield, B. D., & Kerick, S. E. (2007). The psychology of superior sport performance: A cognitive and affective neuroscience perspective. In G. Tenenbaum & R. Eklund (Eds.), *Handbook of sport psychology* (pp. 84–109). <https://doi.org/10.1002/9781118270011.ch4>
- Hebbar, P. A., Bhattacharya, K., Prabhakar, G., Pashilkar, A. A., & Biswas, P. (2021). Correlation between physiological and performance-based metrics to estimate pilots' cognitive workload. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.555446>
- Hess, R. A. (1977). Prediction of pilot opinion ratings using an optimal pilot model. *Human Factors*, 19(5), 459–475. <https://doi.org/10.1177/001872087701900503>
- Hockey, G. R. (1997). Compensatory control in the regulation of human performance under stress and high workload; a cognitive-energetical framework. *Biological psychology*, 45(1-3), 73–93. [https://doi.org/10.1016/s0301-0511\(96\)05223-4](https://doi.org/10.1016/s0301-0511(96)05223-4)
- Hockey, G. R., Nickel, P., Roberts, A. C., & Roberts, M. H. (2009). Sensitivity of candidate markers of psychophysiological strain to cyclical changes in manual control load during simulated process control. *Applied Ergonomics*, 40(6), 1011–8. <https://doi.org/10.1016/j.apergo.2009.04.008>
- Jaquess, K. J., Gentili, R. J., Lo, L. C., Oh, H., Zhang, J., Rietschel, J. C., Miller, M. W., Tan, Y. Y., & Hatfield, B. D. (2017). Empirical evidence for the relationship between cognitive workload and attentional reserve.

International Journal of Psychophysiology, 121, 46–55. <https://doi.org/doi.org/10.1016/j.ijpsycho.2017.09.007>

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, 4(186). <https://doi.org/10.3389/fnhum.2010.00186>

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Citeseer.

Kayser, J. (2009). Current source density (CSD) interpolation using spherical splines - CSD toolbox (version 1.1).

Kayser, J., & Tenke, C. E. (2006a). Principal components analysis of laplacian waveforms as a generic method for identifying ERP generator patterns: I. evaluation with auditory oddball tasks. *Clinical Neurophysiology*, 117(2), 348–368. <https://doi.org/10.1016/j.clinph.2005.08.034>

Kayser, J., & Tenke, C. E. (2006b). Principal components analysis of laplacian waveforms as a generic method for identifying ERP generator patterns: II. adequacy of low-density estimates. *Clinical Neurophysiology*, 117(2), 369–380. <https://doi.org/10.1016/j.clinph.2005.08.033>

King, R., & Schaefer, A. (2011). The emotional startle effect is disrupted by a concurrent working memory task. *Psychophysiology*, 48(2), 269–272. <https://doi.org/https://doi.org/10.1111/j.1469-8986.2010.01062.x>

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Research Reviews*, 53(1), 63–88. <https://doi.org/10.1016/j.brainresrev.2006.06.003>

- Klyde, D. H., Lampton, A. K., Mitchell, D. G., Berka, C., & Rhinehart, M. (2021). A new approach to aircraft handling qualities prediction. In *Aiaa scitech 2021 forum*. <https://doi.org/10.2514/6.2021-0178>
- Klyde, D. H., Pitoniak, S. P., Schulze, P. C., Ruckel, P., Rigsby, J., Fegely, C. E., Xin, H., Fell, W. C., Brewer, R., Conway, F., Mulato, R., Horn, J., Ott, L. C. R., & Blanken, C. L. (2020). Piloted simulation evaluation of tracking mission task elements for the assessment of high-speed handling qualities. *Journal of the American Helicopter Society*, 65(3), 1–23. <https://doi.org/10.4050/jahs.65.032010>
- Knyazev, G. G. (2007). Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neuroscience and Biobehavioral Reviews*, 31(3), 377–395. <https://doi.org/10.1016/j.neubiorev.2006.10.004>
- Knyazev, G. G. (2012). Eeg delta oscillations as a correlate of basic homeostatic and motivational processes. *Neuroscience and Biobehavioral Reviews*, 36(1), 677–695. <https://doi.org/10.1016/j.neubiorev.2011.10.002>
- Kuc, A., Grubov, V. V., Maksimenko, V. A., Shusharina, N., Pisarchik, A. N., & Hramov, A. E. (2021). Sensor-level wavelet analysis reveals eeg biomarkers of perceptual decision-making. *Sensors*, 21(7), 2461. <https://doi.org/10.3390/s21072461>
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional se-

- lection. *Science*, 320(5872), 110–113. <https://doi.org/10.1126/science.1154735>
- Lone, M., & Cooke, A. (2014). Review of pilot models used in aircraft flight dynamics. *Aerospace Science and Technology*, 34, 55–74. <https://doi.org/10.1016/j.ast.2014.02.003>
- McDonnell, J. D. (1968). *Pilot rating techniques for the estimation and evaluation of handling qualities* (Report). SYSTEMS TECHNOLOGY INC HAWTHORNE CA.
- McRuer, D. T., & Jex, H. R. (1967). A review of quasi-linear pilot models. *IEEE Transactions on Human Factors in Electronics*, HFE-8(3), 231–249. <https://doi.org/10.1109/thfe.1967.234304>
- McRuer, D. T., Clement, W. F., Thompson, P. M., & Magdaleno, R. E. (1990). *Minimum flying qualities. volume 2. pilot modeling for flying qualities applications* (Report). SYSTEMS TECHNOLOGY INC HAWTHORNE CA.
- McRuer, D. T., & Krendel, E. S. (1974). *Mathematical models of human pilot behavior* (Report). ADVISORY GROUP FOR AEROSPACE RESEARCH and DEVELOPMENT NEUILLY-SUR-SEINE (FRANCE).
- McRuer, D. T., Myers, T. T., & Thompson, P. M. (1989). Literal singular-value-based flight control system design techniques. *Journal of Guidance, Control, and Dynamics*, 12(6), 913–919. <https://doi.org/10.2514/3.20500>
- Memon, W. A., White, M. D., Padfield, G. D., Cameron, N., & Lu, L. (2022). Helicopter handling qualities: A study in pilot control compensation. *The*

Aeronautical Journal, 126(1295), 152–186. <https://doi.org/10.1017/aer.2021.87>

Miller, M. W., Groman, L. J., Rietschel, J. C., McDonald, C. G., Iso-Ahola, S. E., & Hatfield, B. D. (2013). The effects of team environment on attentional resource allocation and cognitive workload. *Sport, Exercise, and Performance Psychology*, 2(2), 77–89. <https://doi.org/10.1037/a0030586>

Miller, M. W., Rietschel, J. C., McDonald, C. G., & Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *Int J Psychophysiol*, 80(1), 75–8. <https://doi.org/10.1016/j.ijpsycho.2011.02.003>

Moray, N. (1979). Models and measures of mental workload. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 13–21). Boston, MA, Springer US. https://doi.org/10.1007/978-1-4757-0884-4_2

Mulder, G. (1986). The concept and measurement of mental effort. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 175–198). Dordrecht, Springer Netherlands. https://doi.org/10.1007/978-94-009-4448-0_12

Oldfield, R. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia*, 9(1), 97–113. [https://doi.org/http://dx.doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/http://dx.doi.org/10.1016/0028-3932(71)90067-4).

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educa-*

tional Psychologist, 38(1), 63–71. https://doi.org/10.1207/S15326985EP3801_

8

Padfield, G., Charlton, M., Jones, J., & Bradley, R. (1994). Where does the workload go when pilots attack manoeuvres? An analysis of results from flying qualities theory and experiment, In *20th european rotorcraft forum*, Amsterdam, The Netherlands, NLR.

Paul, R., & Rhinehart, M. (2020). Exploring pilot workload using inceptor time histories, In *Vertical flight society's 76th annual forum and technology display*, Virtual.

Paul, R., Rhinehart, M., Cahill, K., & Smith, Z. (2023). Investigation of pilot inceptor metrics as correlates to qualitative workload, In *Vfs handling qualities technical meeting*, NAVAIR.

Pfurtscheller, G., & Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalography and Clinical Neurophysiology*, 42(6), 817–826. [https://doi.org/doi.org/10.1016/0013-4694\(77\)90235-8](https://doi.org/doi.org/10.1016/0013-4694(77)90235-8)

Pogosyan, A., Gaynor, L. D., Eusebio, A., & Brown, P. (2009). Boosting cortical activity at beta-band frequencies slows movement in humans. *Current Biology*, 19(19), 1637–1641. <https://doi.org/10.1016/j.cub.2009.07.074>

Polich, J. (2007). Updating p300: An integrative theory of p3a and p3b. *Clin Neurophysiol*, 118(10), 2128–48. [https://doi.org/10.1016/j.clinph.2007.](https://doi.org/10.1016/j.clinph.2007.04.019)

04.019

- Rietschel, J. C., Miller, M. W., Gentili, R. J., Goodman, R. N., McDonald, C. G., & Hatfield, B. D. (2012). Cerebral-cortical networking and activation increase as a function of cognitive-motor task difficulty. *Biological Psychology*, 90(2), 127–133. <https://doi.org/doi.org/10.1016/j.biopsycho.2012.02.022>
- Rietschel, J. C. (2011). *Psychophysiological investigation of attentional processes during motor learning* (Doctoral dissertation). University of Maryland, College Park.
- Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight: A decade of practical use* (Report). Royal Aerospace Establishment Farnborough (United Kingdom).
- Saul, L. J., & Davis, H. (1933). Action currents in the central nervous system. *Archives of Neurology and Psychiatry*, 29(2), 255–259. <https://doi.org/10.1001/archneurpsyc.1933.02240080045004>
- Sauseng, P., Griesmayr, B., Freunberger, R., & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of eeg theta oscillations. *Neuroscience & Biobehavioral Reviews*, 34(7), 1015–1022. <https://doi.org/doi.org/10.1016/j.neubiorev.2009.12.006>
- Shaw, E. P., Rietschel, J. C., Hendershot, B. D., Pruziner, A. L., Miller, M. W., Hatfield, B. D., & Gentili, R. J. (2018). Measurement of attentional reserve and mental effort for cognitive workload assessment under various task demands during dual-task walking. *Biological Psychology*, 134, 39–51. <https://doi.org/10.1016/j.biopsycho.2018.01.009>

- Vogel, W., Broverman, D. M., & Klaiber, E. L. (1968). EEG and mental abilities. *Electroencephalography and Clinical Neurophysiology*, 24, 166–175. [https://doi.org/10.1016/0013-4694\(68\)90122-3](https://doi.org/10.1016/0013-4694(68)90122-3)
- Weinberg, R. S., & Hunt, V. V. (1976). The interrelationships between anxiety, motor performance and electromyography. *Journal of Motor Behavior*, 8(3), 219–224. <https://doi.org/10.1080/00222895.1976.10735075>
- Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70–73. <https://doi.org/10.1109/tau.1967.1161901>
- Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *International Journal of Aviation Psychology*, 12(1), 3.
- Winkler, I., Haufe, S., & Tangermann, M. (2011). Automatic classification of artifactual ica-components for artifact removal in eeg signals. *Behavioral and Brain Functions*, 7(1), 30. <https://doi.org/10.1186/1744-9081-7-30>
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5), 459–482. <https://doi.org/doi:10.1002/cne.920180503>
- Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, 58(1), 1–17. <https://doi.org/10.1080/00140139.2014.956151>

Zarjam, P., Epps, J., & Chen, F. (2011). Characterizing working memory load using EEG delta activity, In *2011 19th european signal processing conference*.

Zarjam, P., Epps, J., Chen, F., & Lovell, N. H. (2013). Estimating cognitive workload using wavelet entropy-based features during an arithmetic task. *Computers in Biology and Medicine*, 43(12), 2186–2195. <https://doi.org/10.1016/j.combiomed.2013.08.021>