

*TESIS DOCTORAL INTERNACIONAL – INTERNATIONAL PHD THESIS*

**CONVERGENCIAS Y DIVERGENCIAS ENTRE MEDIDAS DE  
EVALUACIÓN DE CARGA Y FATIGA MENTAL**

**CONVERGENCES AND DIVERGENCES BETWEEN MENTAL WORKLOAD AND MENTAL  
FATIGUE MEASURES**

MUÑOZ DE ESCALONA FERNÁNDEZ, ENRIQUE

**DIRECTOR**

CAÑAS DELGADO, JOSÉ. J.

**GRUPO DE ERGONOMÍA COGNITIVA**

DEPARTAMENTO DE PSICOLOGÍA EXPERIMENTAL Y FISIOLOGÍA DEL  
COMPORTAMIENTO

PROGRAMA DE DOCTORADO EN CIENCIAS SOCIALES

Mind, Brain, and Behavior Research Center (CIMCYC)



**UNIVERSIDAD  
DE GRANADA**

Granada, Junio de 2021

Editor: Universidad de Granada. Tesis Doctorales  
Autor: Enrique Muñoz de Escalona Fernández  
ISBN: 978-84-1117-128-1  
URI: <http://hdl.handle.net/10481/71617>



# **Acknowledgments**

---

*Agradecimientos*



En primer lugar, agradezco a José J. Cañas, mi director de tesis, por todos estos años de vivencia y aprendizaje juntos. Gracias a ti escribo esta tesis, que resulta ser el culmen y fruto de todos estos años de trabajo, años marcados por esfuerzo y frustraciones, pero también por ilusiones, alegrías y enriquecimiento personal y profesional. Tengo la suerte de haber tenido un director de tesis cercano, ameno, amable y sobretodo amigo, que me ha brindado todo lo que está en su mano para allanarme el camino, que me ha ayudado a ampliar mi círculo de contactos y que me ha ayudado a sembrar los inicios de mi trayectoria profesional.

I would like to thanks Chiara Leva and Luca Longo for hosting me at their research group and for giving me the opportunity to learn and work with them. Chiara really made me feel like if I were at home during my research stay at Dublin (such a wonderful and magical city), she always treated me exceptionally well and I feel very grateful about this.

Thanks to the members of the committee and external evaluators who kindly accepted to evaluate my doctoral dissertation.

Agradezco a mi familia por apoyarme en todas mis decisiones profesionales y compartir mis ilusiones y alegrías. A mis hermanos de los cuales me siento muy orgulloso. A mi madre por velar siempre por mi bienestar, luchar por mi formación profesional y sentirse orgullosa de mí, orgullo compartido por mi padre, a quien también agradezco su cariño, sus palabras siempre amables y la alta autoestima que siempre me ha brindado, al igual que mi madre.

Agradezco a Andrea por ser mi compañera de viaje. Por compartir juntos alegrías, tristezas e innumerables vivencias. Agradezco tu apoyo en los momentos difíciles, tu cariño, tus sonrisas y caricias que hacen de mi mundo un lugar mejor.

Agradezco a mis grandes amigos: Sergio Garrido, por ser un hermano para mí; Gines Navarro y Jorge Torres, compañeros de carrera, por ser amigos excepcionales y excelentes personas; Luis Javier Vázquez y Jose Antonio Ortega, por ser compañeros fieles, de esos en quienes se puede confiar ciegamente. Paco Rodríguez, por haber sido mi luz en mis momentos de mayor oscuridad... Sé que puedo confiar en todos estos amigos y puedo decir, firmemente y con orgullo, que tengo amigos verdaderos.

Sin todos y cada uno de vosotros mi vida estaría más vacía...

Gracias por vuestra existencia.



# **Table of Contents**

---

*Índice de Contenidos*





RESUMEN (in Spanish)	15
OVERVIEW	21
CHAPTER I: INTRODUCTION (in Spanish)	27
1. Introducción	29
2. Definición de Carga Mental y Fatiga Mental	33
3. Metodologías para evaluar la carga y fatiga mental	37
3.1. Medidas de rendimiento o ejecución	40
3.1.1 Ventajas e inconvenientes de las medidas de ejecución	42
3.2. Medidas psicofisiológicas	43
3.2.1 Actividad eléctrica cerebral	44
3.2.2 Actividad metabólica cerebral	49
3.2.3 Frecuencia cardíaca y respiración	52
3.2.4 Respuesta psicogalvánica	54
3.2.5 Parámetros oculares	56
3.2.5.1 Diámetro pupilar	57
3.2.5.2 Patrones de mirada	60
3.2.5.3. Fijaciones	61
3.2.5.4. Parámetros de velocidad	61
3.2.5.5. Parámetros de parpadeo	63
3.2.6 Ventajas e inconvenientes generales de las medidas psicofisiológicas	65
3.3. Medidas subjetivas	67
3.3.1. Medidas de aplicación off-line	69
3.3.2. Medidas de aplicación on-line	76
4. Convergencias y divergencias entre medidas	80
4.1 Asociaciones, disociaciones e insensibilidades	83
4.2 Posibles causas del surgimiento de divergencias entre medidas	87
5. Objetivos e hipótesis de la tesis	92
EMPIRICAL STUDIES	96
CHAPTER II: LATENCY DIFFERENCES BETWEEN MENTAL WORKLOAD MEASURES IN DETECTING WORKLOAD CHANGES	98
Abstract	101
1. Introduction	102
2. Related Work	103
3. Design and Methodology	105

4. Results	112
5. Discussion	114
6. Conclusions	117
References	118
CHAPTER III: TASK DEMAND TRANSITION RATES OF CHANGE EFFECTS ON MENTAL WORKLOAD MEASURES DIVERGENCE	124
Abstract	127
1. Introduction	128
2. Related Work	129
3. Design and Methodology	132
4. Results	140
5. Discussion	143
6. Conclusions	146
References	146
CHAPTER IV: TASK DEMAND TRANSITION PEAK POINT EFFECTS ON MENTALWORKLOAD MEASURES DIVERGENCE	151
Abstract	154
1. Introduction	155
2. Related Work	156
3. Design and Methodology	159
4. Results	168
5. Discussion	172
6. Conclusions	175
References	176
CHAPTER V: INCONSISTENCIES BETWEEN MENTAL FATIGUE MEASURES UNDER COMPENSATORY CONTROL THEORIES	181
Abstract	184
1. Introduction	185
2. Materials and method	190
3. Results	197
4. Discussion	201
5. Resumen	204
References	205

CHAPTER VI: FUNDAMENTAL FREQUENCY AS AN ALTERNATIVE METHOD FOR ASSESSING MENTAL FATIGUE	211
Abstract	214
1. Introduction	215
2. Related Work	216
3. Design and Methodology	224
4. Results	230
5 Discussion	234
6 Conclusions	235
References	235
CHAPTER VII: GENERAL DISCUSSION (in Spanish)	239
1. Introducción	241
2. Divergencias entre medidas de carga mental	242
2.1 La importancia del tiempo	242
2.2 El efecto de las transiciones en carga	243
2.3 El efecto del nivel preexistente de carga mental	246
3. Divergencias entre medidas de fatiga mental	249
4. Exploración de nuevas medidas	252
5. Futuras direcciones de investigación	254
6. Conclusiones finales	259
7. Final conclusions (in English)	262
REFERENCES	266



# Resumen

---



La carga y la fatiga mental son conceptos abstractos que revisten gran importancia en el ámbito de la Ergonomía Cognitiva y las Ciencias del Trabajo. **Una gestión adecuada de ambos constructos resulta fundamental para garantizar el bienestar de los operarios y la seguridad laboral** (Kirschenbaum, Oigenblick & Goldberg, 2000; Rodgers, 2000; Lilley et al., 2002; Josten, Ng-A-Tham & Thierry, 2003; Taylor & Dorn, 2006; Dawson et al., 2011; Sarsangi et al. 2017), **pero también constituyen un criterio clave para optimizar la productividad empresarial** (Priya et al., 2010; Fan & Smith, 2017; Smith & Smith, 2017). Si bien la carga y la fatiga mental como constructos psicológicos son intangibles y, por tanto, no cuantificables de forma directa, la ciencia nos ha desvelado que existen diversos índices que son capaces de reflejar de forma indirecta los cambios que acontecen en la magnitud de dichos constructos (Moray, 1979; Wickens, 2017). Más concretamente, hoy en día está sólidamente establecido que las variaciones en el nivel de carga y fatiga mental soportado por un individuo se reflejan mediante una serie de índices, los cuales pueden clasificarse en tres grandes grupos de medidas: las medidas de rendimiento, las medidas psicofisiológicas y las medidas subjetivas (Cain, 2007; Hancock, 2017; Wickens, 2017). Considerando que todos estos índices reflejan las variaciones que se producen en la magnitud del mismo constructo (carga o fatiga mental), cabría esperar la existencia de una convergencia elevada entre las diferentes medidas, de manera que, con cada una de las medidas de forma independiente, podríamos cuantificar adecuadamente la carga mental (y la fatiga mental) y, por ende, la variación en su magnitud con el paso del tiempo; y además, estas medidas deberían coincidir en reflejar dichas variaciones (convergencia entre medidas). No obstante, **la literatura científica ha demostrado en numerosas ocasiones que esto no siempre acontece, de manera que el surgimiento de divergencias entre medidas de carga y de fatiga mental es más frecuente de lo que cabría esperar** (Yeh & Wickens, 1984; Yeh & Wickens, 1988; Casper, 1990; Horrey, Lesch & Garabet, 2009; Kamzanova, Kustubayeva & Matthews, 2014; Matthews et al., 2015; Wickens, 2017; Hancock, 2017, Hancock & Matthews, 2019). Este hecho constituye un problema para nuestra ciencia, pero sobre todo para nuestra práctica, ya que una adecuada gestión de la carga y la fatiga mental requiere necesariamente ser capaces de evaluar adecuadamente ambos constructos.

Este trabajo de investigación se centra precisamente en la necesidad de disponer de metodologías fiables, válidas y eficaces para evaluar los conceptos de carga y de fatiga



mental. Más concretamente, **el objetivo principal de la tesis doctoral es arrojar luz sobre las causas y circunstancias en las que surgen divergencias entre las diferentes metodologías de evaluación de carga y fatiga mental**, asimismo también se ha abordado sucintamente la **necesidad de obtener nuevas medidas psicofisiológicas que sean capaces de superar las limitaciones tradicionales** que este tipo de medidas suponen.

Esta tesis doctoral está compuesta por un total de 7 capítulos. El **capítulo I**, de cariz introductorio, tiene como objetivo **acercar al lector a los conceptos de carga y fatiga mental, sus métodos de evaluación y al problema de investigación** abordado en la presente tesis, principalmente las divergencias que se producen entre las diferentes medidas de carga y fatiga mental, pero también la necesidad de explorar nuevas medidas psicofisiológicas que superen las limitaciones tradicionales de este tipo de medidas. Los capítulos de van del II al VI contienen las diferentes investigaciones empíricas desarrolladas para arrojar luz al problema de investigación.

El **capítulo II**, recoge un estudio empírico cuyo objetivo fue **explorar la posible relación existente entre el surgimiento de divergencias entre medidas de carga mental y la existencia de diferencias en cuanto a la latencia de respuesta de las diferentes metodologías de evaluación consideradas**. Los resultados aportaron evidencias a favor de la existencia de diferentes latencias temporales entre medidas de carga mental que acontecen particularmente durante picos de elevada carga mental, las cuales contribuyeron a generar divergencias entre medidas. Más concretamente, la medida subjetiva mostró una menor latencia de respuesta que la medida psicofisiológica (diámetro pupilar) durante los picos de mayor demanda cognitiva, coincidiendo también con los picos de peor rendimiento en los participantes.

En el **capítulo III**, se describe un estudio empírico que persiguió por objetivo **analizar el posible efecto sobre el surgimiento de disociaciones entre medidas de carga mental que podría suscitar la tasa de cambio (abrupta vs. lineal) durante las transiciones (ascendentes vs. descendentes) que se producen en los niveles de demanda cognitiva de la tarea**. Los hallazgos de dicho estudio pusieron de manifiesto la existencia de mayores divergencias entre medidas de carga mental durante la condición experimental de “tasa de cambio variable y ascendente”. Estos resultados sugieren la existencia de una

sensibilidad diferencial de las diferentes medidas de carga mental a los cambios abruptos en los niveles de demanda cognitiva de la tarea, que acontecen durante las transiciones en carga, lo cual favorece el surgimiento de divergencias entre medidas.

El **capítulo IV** abarca un estudio empírico que trata de **arrojar luz acerca de si algunas medidas podrían ser más sensibles a los cambios abruptos que se producen en los niveles de demanda cognitiva de la tarea durante las transiciones ascendentes en carga, en función del nivel preexistente de carga mental**, lo cual terminaría afectando al grado de convergencia entre medidas. Los resultados de dicho estudio desvelaron la existencia de diferencias entre medidas en el modo de reflejar los cambios abruptos en carga, en función del nivel de demanda cognitiva preexistente. Más concretamente, la mayor divergencia entre medidas se observó cuando el cambio abrupto en los niveles de demanda cognitiva de la tarea aconteció con un nivel preexistente reducido de demanda cognitiva, debido principalmente a la sobreactivación fisiológica que se produce en el organismo bajo dicha circunstancia y que se vió reflejada a través de la medida psicofisiológica (diámetro pupilar).

El **capítulo V**, presenta un estudio experimental cuyo objetivo consistió en **analizar las divergencias que se producen entre medidas de fatiga mental bajo el efecto del llamado “mecanismo de control compensatorio” propuesto por Hockey en 1997**. Atendiendo a dicho modelo, la fatiga mental percibida actúa como una señal de alarma del organismo frente al agotamiento de los recursos disponibles, así, un individuo puede comenzar a sentirse fatigado mentalmente pese a disponer aún de una cantidad de recursos suficiente para afrontar la tarea de forma satisfactoria, esto es, con una elevada activación fisiológica y un rendimiento adecuado. Nuestros hallazgos en dicho estudio pusieron de manifiesto cómo efectivamente, el mecanismo de control compensatorio entraba en funcionamiento, favoreciendo el surgimiento de divergencias entre medidas de fatiga mental.

En el **capítulo VI**, presentamos un último estudio empírico en el cual exploramos la **posibilidad de utilizar una nueva medida psicofisiológica, la denominada “frecuencia fundamental” de la voz, como una nueva medida alternativa, fiable y válida de fatiga mental, que es capaz de superar las limitaciones más relevantes de las medidas psicofisiológicas**, estas son, el elevado nivel de intrusividad y los elevados

costes y requisitos de implementación. Los resultados de dicho estudio resultan prometedores, y sugieren que dicho parámetro acústico efectivamente resulta sensible al surgimiento de la fatiga mental, por lo que puede ser propuesto como una nueva posible medida alternativa, fiable y válida de fatiga mental.

En última instancia, en el **capítulo VII** presentamos una **síntesis de los resultados más relevantes** obtenidos a lo largo de todos los estudios constituyentes de la presente tesis doctoral, al tiempo que **se señalan las limitaciones de los mismos y las implicaciones más importantes relativas a dichos hallazgos**. Finalmente, abordamos sucintamente las posibles **direcciones futuras de investigación**, en base a los resultados obtenidos, al objeto de poder continuar ahondando en el conocimiento del fenómeno de las divergencias entre las medidas de carga y entre las medidas de fatiga mental, así como el uso de nuevas medidas psicofisiológicas.

Cabe resaltar el hecho de que resulta posible que los lectores de la presente tesis doctoral estimen como redundante parte de la información expuesta en los sucesivos capítulos, lo cual obedece a que dichos capítulos han sido concebidos y, por ende, redactados como artículos de investigación independientes, para poder ser publicados en revistas y libros específicos de investigación.

# Overview

---



Mental workload and mental fatigue are abstract concepts, which are highly relevant in both Cognitive Ergonomics and Work Sciences fields. **Achieving a proper management of both constructs is essential to guarantee occupational health and safety** (Kirschenbaum, Oigenblick & Goldberg, 2000; Rodgers, 2000; Lilley et al., 2002; Josten, Ng-A-Tham & Thierry, 2003; Taylor & Dorn, 2006; Dawson et al., 2011; Sarsangi et al. 2017), **but also to optimize business productivity** (Priya et al., 2010; Fan & Smith, 2017; Smith & Smith, 2017). Although mental workload and mental fatigue as psychological constructs are intangible and, therefore, not directly measurable, science has revealed that there exists certain indexes that are indirectly capable of reflecting variations in the magnitude of these constructs (Moray, 1979; Wickens, 2017). More specifically, nowadays it is well established that mental workload and mental fatigue variations can be reflected by several different indexes, which can be classified into three main groups of measures: performance measures, psychophysiological measures and subjective measures (Cain, 2007; Hancock, 2017; Wickens, 2017). Considering that all these indexes reflect the variations that occur in the magnitude of the same construct (mental workload or mental fatigue), it would be expected the existence of a high convergence between them. This would mean that we could assess mental workload (and mental fatigue) with each different measure independently and thus we could assess how the magnitude of these constructs change over time; furthermore, these measures should coincide in reflecting those variations (convergence between measures). However, **literature research has shown on numerous occasions that this does not always occurs, so that the emergence of divergences between measures is more frequent than we might expect** (Yeh & Wickens, 1984; Yeh & Wickens, 1988; Casper, 1990; Horrey, Lesch & Garabet, 2009; Kamzanova, Kustubayeva & Matthews, 2014; Matthews et al., 2015; Hancock, 2017, Hancock & Matthews, 2019). This fact is a major problem for our science, but especially for our practice, since achieving an adequate management of mental workload and mental fatigue necessarily requires being able to evaluate both constructs adequately.

This research work focuses precisely on the need for reliable, valid and effective methodologies to evaluate mental workload and mental fatigue. More specifically, **the main objective of this doctoral thesis is to shed some light on the causes and circumstances in which divergences arise between the different methodologies for assessing mental workload (and mental fatigue), as well as the need to obtain new**

**psychophysiological measures that are capable of overcoming the traditional limitations** of such measures.

This doctoral thesis comprises 7 chapters. **Chapter I**, which is introductory in nature, aims to **bring the reader closer to the concepts of mental workload and mental fatigue, their assessing methods and the research problem addressed in this thesis:** mainly the divergences that occur between the different mental workload (and mental fatigue) measures, but also the need to explore new psychophysiological measures that overcome the traditional limitations of this type of measures. Chapters II to VI contain 5 empirical studies developed to shed light on the research problem explained above.

**Chapter II** includes an empirical study whose objective was to **explore the possible relationship between the emergence of divergences between mental workload measures and the existence of latency differences between these measures.** The results of this study provided evidence in favour of the existence of latency differences between mental workload measures. Particularly, latency differences between measures emerged during high mental workload peak experience and this would negatively affect convergence between mental workload measures. More specifically, subjective measure showed a lower latency response than the psychophysiological index (pupil diameter) during higher cognitive demand peaks, which coincided with participants' worse performance peaks as well.

**Chapter III** describes an empirical study that aimed to **analyse whether differential sensitivity to rates of change in task demand transitions would affect the convergence between mental workload measures.** These sensitive differences would end up increasing divergences between mental workload measures. The findings of this study revealed the existence of greater divergences between mental workload measures during the "increasing task-load with a variable (abrupt) rate of change" experimental condition. These results suggest the existence of a differential sensitivity of the different mental workload measures to abrupt changes in task-demand levels, which contributes to the emergence of divergences between mental workload measures.

**Chapter IV** covers an empirical study that tries to **shed light on whether some mental workload measures could be more sensitive than others to abrupt increases**

**in task demand, depending on baseline task-load situation.** These sensitive differences would end up increasing divergences between mental workload measures. The results of this study revealed the existence of differences between measures in the way of reflecting abrupt increases in task-load, depending on the pre-existing level of cognitive demand. More specifically, the greatest divergence between measures was observed when the abrupt increase in task-load occurred during reduced pre-existing levels of cognitive demand. This was mainly due to the physiological overactivation that occurs in the organism under this circumstance, which was reflected through our psychophysiological measurement (pupil diameter).

**Chapter V** describes an experimental study whose objective was to **analyze divergences between mental fatigue measures under the effect of the so-called “compensatory control mechanism” proposed by Hockey in 1997.** According to this model, perceived mental fatigue acts as an alarm signal of the organism against available resources depletion, thus, a person may begin to feel mentally fatigued despite still having enough available resources to satisfactorily cope with the task, that is, while still having high physiological activation and adequate performance. Our findings in this study showed how the compensatory control mechanism came into operation, contributing to the emergence of divergences between mental fatigue measures.

In **Chapter VI**, we present the last empirical study of this thesis, in which **we explore the possibility of considering a new psychophysiological measure, the “fundamental frequency” of the voice, as a new alternative, reliable and valid measure for assessing mental fatigue. The main advantage of this new methodology is that it is capable of overcoming the most relevant limitations of psychophysiological measures,** these are, the high level of intrusiveness and the high costs and implementation requirements. The results of this study are promising, and suggest that said acoustic parameter is indeed sensitive to the emergence of mental fatigue, so it can be proposed as a new possible alternative, reliable and valid measure of mental fatigue.

Finally, in **Chapter VII** we present a **synthesis of the most relevant results** obtained throughout all the constituent studies of this doctoral thesis, while **pointing out their limitations and the most important implications related to said findings.** Finally, based in obtained results, we briefly address the **possible future directions of this**



**research line**, in order to continue increasing the knowledge about convergences and divergences between mental workload and mental fatigue measures, as well as about the use of new alternative psychophysiological measures.

It should be stressed that it is possible that the readers of this doctoral thesis will likely consider the information presented in the successive chapters to be a little redundant, and it is due to the fact that these chapters were drafted as independent research articles, to be published in specific research journals and books

# Chapter I

---

*Introduction/Introducción*



## 1. Introducción

Ginés Navarro trabaja como transportista en una empresa internacional de transporte de mercancías. La empresa, una de las más destacadas del sector, le facilita como herramienta de trabajo un tráiler de marca Premium, con tecnología de detección de fatiga mental, a través de parámetros oculares. El 22 de diciembre de 2019, en plena campaña navideña, Ginés recibe la orden de transportar una carga desde Lyon (Francia) hasta Madrid (España), trayecto que, cubriendo un total de 1238km, se resuelve en aproximadamente 11 horas y 30 minutos de duración, sin contar descansos. La campaña navideña está siendo muy dura, pero Ginés recibe mucha presión por parte de su empresa y sabe que tiene que completar dicho trayecto en el menor tiempo posible. El viaje comenzó a las 9 a.m, y ya han transcurrido más de 10 horas cuando Ginés se aproxima a Aranda de Duero (España). Aunque se había detenido unos minutos a descansar en Burdeos (Francia), la fatiga mental, que ya habría empezado a surgir horas atrás, no había dejado de aumentar desde entonces. El descanso en Burdeos no había sido suficiente y el desgaste acumulado de la campaña de navidad comenzaba a afectarle. Ginés lo sabía, su sistema de alarma cognitiva frente al agotamiento de sus recursos mentales se había activado tiempo atrás, sin embargo, Ginés había decidido continuar, al fin y al cabo, en su empresa le habían indicado que su tráiler estaba equipado con un sistema fiable de detección de fatiga mental y la presión temporal le animaba a continuar con su trayecto. Poco tiempo después, Ginés se sale de la calzada y en un intento sobreactuado con el volante por reorientar el tráiler, se produce un aparatoso accidente que deja el tráiler volcado en la calzada y a Ginés inconsciente en el interior de la cabina. Afortunadamente no había perdido la vida, ni tampoco había causado daños a terceros, si bien, su fortuna podría no haber sido tan halagüeña... ¿Qué había ocurrido?, ¿Por qué el tráiler no le había avisado del surgimiento de fatiga mental?, ¿Por qué Ginés confiaba más en dicha tecnología que en sus propios instintos?, ¿Se podría haber evitado?

Jorge Torres es controlador aéreo y desempeña sus funciones en la torre de control de Torrejón de Ardoz en Madrid (España). Estamos a 25 de agosto, en plena temporada alta de vuelos. Su compañero, se percató de que algo no va bien con él, y le pregunta que si se encuentra sobrecargado. Jorge niega estar sobrecargado, sin embargo, su compañero

observa excesiva sudoración en su rostro, al igual que unas pupilas excesivamente dilatadas. Al cabo de 15 minutos Jorge, que se encuentra monitorizando 12 aviones en su radar, no se percata de una situación de conflicto entre 2 de sus aviones, los cuales se han aproximado tanto que el riesgo de colisión es muy elevado. Por suerte, el compañero de Jorge atiende lo ocurrido y resuelve el conflicto satisfactoriamente y a tiempo. Jorge, abrumado y desorientado, decide tomarse un descanso... ¿Qué había ocurrido?, ¿por qué Jorge afirmó no estar sobrecargado, sin embargo, sus índices psicofisiológicos indicaban lo contrario?, ¿cómo se podría haber evitado dicha situación de peligro?

Las situaciones anteriormente descritas constituyen 2 ejemplos de **circunstancias que se dan a diario en miles de puestos de trabajo y que ponen en riesgo la seguridad de numerosas personas, tanto de los propios trabajadores como de terceros.**

Los conceptos de carga y fatiga mental tienen mucha relevancia y una larga tradición de estudio en el ámbito de prevención de riesgos laborales y ergonomía cognitiva. Hoy día nadie duda acerca de la importancia de una adecuada gestión de la carga y la fatiga mental en el ámbito laboral, no sólo por motivos meramente preventivistas (Kirschenbaum, Oigenblick & Goldberg, 2000; Rodgers, 2000; Lilley et al., 2002; Josten, Ng-A-Tham & Thierry, 2003; Taylor & Dorn, 2006; Da Silva, 2014; Paxion, Galy & Berthelon, 2014; Sarsangi et al. 2017), sino también por motivos económicos e intereses empresariales (Priya et al., 2010; Fan & Smith, 2017; Smith & Smith, 2017). La búsqueda empresarial del máximo beneficio requiere una adecuada gestión de sus recursos humanos. En este sentido, un trabajador sobrecargado es fuente de problemas de seguridad y productividad (Wickens, 2008; Durantin et al., 2014; Wickens, 2017), del mismo modo que un trabajador con infracarga tampoco resulta rentable y entraña riesgos laborales (Endsley & Kiris, 1995; Young & Stanton, 2002). Asimismo, un trabajador fatigado no trabajará de forma óptima, por lo que su eficiencia se verá mermada, redundando esto en posibles accidentes laborales y en una menor productividad (Dawson et al., 2011; Josten, Ng-A-Tham & Thierry, 2003; Taylor & Dorn, 2006; Fan & Smith, 2017). Atendiendo a lo anteriormente expuesto, **la investigación de ambos conceptos en el ámbito de la psicología y las ciencias del trabajo cobra especial relevancia en nuestra sociedad moderna** (Dawson et al., 2011; Longo & Leva, 2017; Longo & Leva, 2018; Longo & Leva, 2019).

Ahora bien, para poder gestionar adecuadamente la carga y fatiga mental de los trabajadores, así como para poder realizar investigación sobre ambos, **debemos ser capaces de cuantificar adecuadamente dichos conceptos mediante alguna metodología.** ¿Cómo podemos medir la carga y la fatiga mental, si son conceptos abstractos e intangibles? y no sólo eso, ¿cómo sabemos cuándo existe sobrecarga e infracarga?, ¿cómo podemos determinar cuándo comienza a experimentar un trabajador el fenómeno de la fatiga mental?

En primera instancia, debe quedar claro que **no existe ningún tipo de metodología capaz de cuantificar de forma directa ambos conceptos.** La carga y la fatiga mental son conceptos abstractos y multidimensionales que, al igual que la mayoría de conceptos tratados en ciencias del comportamiento, no pueden cuantificarse de forma directa. Como representación gráfica, no podemos pesar 1 kilo de carga mental ni medir 50 centímetros de fatiga mental. Sin embargo, **lo que sí podemos hacer para cuantificarlos es observar los reflejos indirectos derivados de ambos conceptos y cuantificarlos** (Moray, 1979; Christodoulou, 2005; Wickens, 2017; Hancock, 2017). A este respecto, existen 3 tipos de metodologías principales para medir carga y fatiga mental, estas son:

- **Índices de desempeño:** esto es, la cantidad y calidad de la tarea desempeñada. Por ejemplo, la tasa de errores, o la tasa de respuesta por unidad de tiempo.
- **Respuestas psicofisiológicas:** son las referentes a los cambios fisiológicos acontecidos por la variación en los niveles de carga mental o fatiga. Por ejemplo, diámetro pupilar, EEG, tasa cardiaca, etc.
- **Estimaciones subjetivas:** correspondiente a la sensación subjetiva de carga mental o fatiga, medida a través de cuestionarios de tipo “on-line” y “off-line”.

Estas medidas son capaces de reflejar estados diferentes de carga y fatiga mental y se les conoce como los tres reflejos primarios de carga o fatiga mental (Wickens, 2017; Hancock 2017). Es de suponer que, si bien los tres tipos de medidas hacen referencia a los mismos conceptos, las tres medidas deben estar relacionadas entre sí, representando de forma equivalente los conceptos aquí expuestos. Esto es, **cabe esperar que las tres medidas muestren convergencia entre sí, de manera que, con cada una de ellas, de**

**forma independiente, podríamos conocer los niveles de carga mental y/o fatiga experimentados. Sin embargo, la literatura bien ha demostrado que esto no es tan claro como podríamos intuir** (Yeh & Wickens, 1984; Yeh & Wickens, 1988; Casper, 1990; Horrey, Lesch & Garabet, 2009; Kamzanova, Kustubayeva & Matthews, 2014; Matthews et al., 2015; Wickens, 2017; Hancock, 2017, Hancock & Matthews, 2019). Más concretamente, **es posible que una medida pueda evolucionar conforme a lo esperado**, a medida que se va modificando la demanda cognitiva de la tarea o se produce el agotamiento de los recursos mentales (“**asociación**” entre la demanda de recursos cognitivos y las variaciones en la magnitud de la medida), **en tanto que otra medida puede evolucionar de forma contraria a lo esperado** (“**disociación**” entre la demanda de recursos cognitivos y las variaciones en la magnitud de la medida) **o simplemente no producirse variación alguna** (“**insensibilidad**” de la medida a los cambios en la demanda de recursos cognitivos). De esta manera, **las medidas de carga (o las medidas de fatiga) pueden converger o divergir entre sí, respectivamente, en función de si dichas medidas coinciden en reflejar el cambio en la demanda cognitiva** (o en el agotamiento de los recursos en el caso de la fatiga) **o, por el contrario, si encontramos que algunas medidas no coinciden entre sí en reflejar los cambios en la magnitud del constructo** (ej. **si alguna o varias medidas no evolucionan conforme a lo esperado, en tanto que otras sí**).

La literatura científica, así como el ámbito aplicado, han demostrado en numerosas ocasiones que el fenómeno de la divergencia entre medidas de carga o fatiga mental es más frecuente de lo que, a priori, podríamos pensar; sin embargo, cabe indicar que, como bien señala Gilbert (2009), la mayoría de divergencias se producen con la medida de carga subjetiva. **Este fenómeno resulta muy problemático para la ciencia y práctica y en los últimos años ha aumentado el interés en su estudio** dentro de la comunidad científica del ámbito de la Ergonomía Cognitiva y los Factores Humanos (Yeh & Wickens, 1984; Casper, 1990; Yeh & Wickens, 1988; Horrey, Lesch & Garabet, 2009; Matthews et al., 2015; Hancock, 2017; Muñoz-de-Escalona & Cañas, 2018; Hancock & Matthews, 2019; Muñoz-de-Escalona, Cañas & van Nes, 2019; Muñoz-de-Escalona, Cañas & Noriega, 2020; Muñoz-de-Escalona et al., 2020). Podemos preguntarnos en este punto **por qué se produce dicha arbitrariedad respecto a la ocurrencia de asociaciones, insensibilidades y disociaciones en distintos tipos de medidas**. Si, como

asumimos todas las medidas deben medir lo mismo, la pregunta lógica es **¿por qué existen esas divergencias?**

Estas son las razones que justifican la investigación realizada en la presente tesis doctoral, en la que se han realizado una serie de estudios para tratar de arrojar luz sobre la ocurrencia y causa de los fenómenos de convergencia y divergencias entre las medidas de carga mental y entre las medidas de fatiga mental. En pocas palabras, nuestro objetivo final en la línea de investigación dentro de la cual se encuadra esta tesis, consiste en tratar de **explicar cuándo, cómo y por qué acontecen las asociaciones, disociaciones e insensibilidades de las diferentes medidas de carga y fatiga mental y en qué medida influyen en el surgimiento de divergencias entre las mismas.** Estas explicaciones deben ayudarnos, por una parte, a avanzar nuestro conocimiento científico sobre los fenómenos de carga y fatiga mental y, por otra parte, permitirnos identificar cuáles son los criterios adecuados para seleccionar una medida concreta para cada ocasión y condición de aplicación determinada.

## 2. Definición de Carga Mental y Fatiga Mental

El término **carga de trabajo mental** es probablemente uno de los conceptos más utilizados en el ámbito de los factores humanos y, por ende, **uno de los conceptos que reviste mayor importancia en su esfera.** La carga mental se define frecuentemente como un concepto ambiguo (Leplat, 2002) o nebuloso (Young et al, 2015), debido en parte a su carácter multidimensional y a las numerosas definiciones otorgadas a lo largo de las décadas. Desde los años 70, década en la que el término comenzó a aparecer en las publicaciones científicas, numerosos términos y definiciones se han ido sucediendo y superponiendo. Los términos esfuerzo mental, coste cognitivo, carga cognitiva, carga atencional, tensión mental o recursos mentales, constituyen un conjunto de términos que se acercan al concepto de carga mental, o lo abordan parcialmente (Leplat, 2002). Dichos términos han sido utilizados indistintamente y equívocamente para referirse al mismo concepto. En cualquier caso, el término carga mental puede ser considerado como uno de los términos más utilizados para describir el fenómeno, el cual ha recibido diferentes definiciones a lo largo de las últimas décadas:



- “Se refiere a la porción de la capacidad de procesamiento de la información de un operador que se requiere para ejecutar las demandas del sistema” (Eggemeier et al., 1991).
- “La carga mental no es una propiedad inherente, sino que emerge de la interacción entre los requerimientos de la tarea, las circunstancias en las que esta se ejecuta, las habilidades, comportamientos y percepciones del operador” (Hart, 1988).
- “Es un constructo hipotético que describe el grado en que los recursos cognitivos que son requeridos para ejecutar una tarea han sido activamente movilizados por el operador” (Gopher & Donchin, 1986).
- “La razón para especificar y evaluar la carga mental reside en cuantificar el coste mental derivado de la ejecución de la tarea, con el objeto de predecir el rendimiento del operador y del sistema” (Cain, 2007).
- “Es el esfuerzo mental que un operador moviliza para controlar o supervisar su tarea, en función de su capacidad para gastar esfuerzo mental” (Curry et al., 1979).
- “Constituye la diferencia entre las capacidades de procesamiento de información del sistema que el rendimiento en la tarea requiere para satisfacer las expectativas de rendimiento y la capacidad disponible en un momento determinado” (Gopher, 1986).

De todas las definiciones anteriormente expuestas, se puede vislumbrar que **el concepto de carga mental se define en términos de los recursos mentales que se encuentran implicados en la realización de una tarea determinada, esto es, como una relación entre los recursos mentales demandados y los recursos mentales disponibles.** El término “capacidad” implica una cantidad finita de recursos (Wickens, 2008), en este caso recursos mentales. Más específicamente, **la carga mental resulta de**

**dividir los recursos demandados** (recursos que son necesarios dedicar a una tarea para realizarla adecuadamente) **entre los recursos disponibles** (recursos de los que una persona dispone para dedicar a la tarea), frente a lo cual esperamos tres situaciones posibles, a saber:

1. **Situación de sobrecarga:** acontece cuando los recursos demandados son mayores que los disponibles.
2. **Situación de equilibrio:** se produce cuando los recursos demandados son equivalentes a los disponibles, con un rango determinado.
3. **Situación de infracarga:** sucede cuando los recursos demandados son muy inferiores a los disponibles.

Si bien la situación de infracarga puede considerarse un problema porque puede resultar en una pérdida de atención y en somnolencia (Endsley & Kiris, 1995; Young & Stanton, 2002), con el consiguiente riesgo de incidentes y accidentes, los problemas realmente importantes para la seguridad acontecen en situaciones de sobrecarga (Wickens, 2008; Durantin et al., 2014; Wickens, 2017). No obstante, lo importante que debemos considerar es que el ser humano intentará en todo momento mantenerse en un equilibrio de recursos demandados y recursos disponibles de tal manera que no peligre la seguridad, su salud no se vea afectada y mantenga un nivel de satisfacción aceptable.

Diferentes factores pueden provocar altos niveles de carga mental: la dificultad en la tarea, presiones temporales, rendimiento, edad, esfuerzo físico, frustración, tensión, fatiga y el tipo de actividad desempeñada pueden ser algunos de ellos (Holmqvist et al., 2011).

**Los recursos demandados dependen directamente del nivel de dificultad de la tarea, es por ello que para poder medir y analizar los recursos demandados, necesitamos cuantificar el nivel de dificultad de la misma** (Wickens, 2008). De otro lado, **los recursos disponibles dependen del nivel de activación, diferencias individuales y factores temporales** (momento del día, tiempo empleado en la tarea, etc.) (Hockey, 2013).

En lo referente al término de fatiga, la definición y la medición de dicho constructo a lo largo de los años ha generado numerosos problemas. La fatiga reduce los niveles de alerta afectando, en última instancia, al rendimiento (Grandjean, 1989), tras un periodo de “alta demanda de actividad cognitiva que requiere una eficiencia mental sostenida” (Loris et al, 2005). Entre las consecuencias de la fatiga, destacan una **disminución del rendimiento, la motivación, la salud, el bienestar y la seguridad laboral** (Grech, Neal, & Yeo, 2009). En este sentido, ha sido considerada como la causa principal de los accidentes laborales (Grandjean, 1989; Dawson et al, 2011; Hopstaken, 2016) y constituye un problema muy frecuente tanto en el ámbito clínico y laboral de las sociedades modernas (González et al., 2005).

Tradicionalmente, la fatiga se ha definido como “un cansancio extremo y persistente, agotamiento o debilidad mental y/o física” (Dittner et al., 2004), que puede afectar al rendimiento. El rendimiento en la tarea puede mejorar o deteriorarse dependiendo, entre otros aspectos, de la cantidad y la calidad de los recursos disponibles (Kahneman, 1973). Asumiendo que los recursos disponibles son limitados y se van agotando, puede surgir el fenómeno conocido como “fatiga mental” (Kahneman, 1973).

La fatiga mental se encuentra ampliamente relacionada con la situación de sobrecarga mental, toda vez que mientras se disponga de suficientes recursos para afrontar una tarea, el desempeño será óptimo, sin embargo, dado que la fatiga mental implica el agotamiento de los recursos mentales disponibles con el paso del tiempo, podemos asumir que **la fatiga mental conducirá a una situación de sobrecarga mental que afectará negativamente al rendimiento. Asimismo, de forma inversa, una situación inicial de sobrecarga mental mantenida en el tiempo (en un individuo con suficientes recursos iniciales), conducirá al surgimiento de la fatiga mental. En suma, la fatiga mental es el término que utilizamos para referirnos al efecto que ejerce el tiempo ejecutando la tarea sobre los recursos disponibles, dando como resultado su paulatino agotamiento con el transcurso del tiempo.**

### 3. Metodologías para evaluar la carga y fatiga mental

A lo largo de estas últimas décadas, los investigadores del ámbito de los Recursos Humanos y la Ergonomía han diseñado numerosos métodos para evaluar los recursos mentales de que dispone un individuo durante la realización de una tarea. Dichos métodos pueden clasificarse atendiendo a varios criterios, de entre los cuales destacaremos dos criterios concretos, estos son, **el momento temporal** en el que se obtiene el registro de los recursos mentales del individuo y **la fuente registro** de los mismos.

Atendiendo al primer criterio, podemos distinguir entre los denominados “métodos off-line” y los “métodos on-line”. **Los métodos “off-line” consisten básicamente en evaluar la carga o fatiga mental una vez que la tarea ha concluido, en tanto que los métodos “on-line” se diferencian de los primeros en que la evaluación se ejecuta durante la realización de la tarea.** Los métodos off-line tienen la principal ventaja de no interferir con la ejecución de la tarea, pero tienen la desventaja de no permitir obtener una estimación durante la realización de la misma (por lo que depende de la memoria), en tanto que los métodos on-line permiten realizar un registro continuo, lo cual termina afectando a la ejecución del evaluado. **Dependiendo del objetivo de la investigación, el investigador se habrá de decantar por una u otra metodología.** Imaginen que nos interesa conocer la cantidad de recursos restantes de que dispone una persona una vez realizada cierta tarea; estaríamos en esa situación, por ejemplo, cuando deseamos averiguar el grado de fatiga mental de una persona al finalizar su jornada laboral y no cuando nuestro objetivo sea conocer cómo se han ido administrando sus recursos mentales durante el desempeño de su actividad laboral. En esta primera situación, los métodos off-line serían más adecuados, mientras que en la segunda situación resultará necesario acudir al uso de métodos on-line.

Independientemente al momento temporal en el que se realiza el registro de las diferentes medidas y, atendiendo al segundo criterio, **los métodos pueden clasificarse en función de la naturaleza de la fuente de registro de los datos.** Numerosos estudios han tratado de definir y establecer modelos neurofisiológicos, psicosociales e incluso biomatemáticos acerca de la carga y fatiga mental (Dawson et al, 2011), coincidiendo todos ellos en determinar la existencia de **3 componentes bien establecidos: 1) componente**

**de rendimiento, 2) componente fisiológico y 3) componente subjetivo.** Si bien los dos primeros componentes son directamente observables y, por tanto, objetivamente mensurables, el tercero, dado su carácter subjetivo, no resulta posible ser observado ni medido de forma directa. Por lo tanto, **considerando estos tres componentes, la metodología para el estudio de la carga mental y la fatiga incluye tres tipos de medidas, las cuales Hockey y Hancock denominan como las 3 medidas primarias de carga/fatiga mental** (Hockey, 2013, Hancock, 2017), a saber:

- **Medidas de rendimiento:** esto es, la cantidad y calidad de la tarea desempeñada. Por ejemplo, la “tasa de errores”, o la “tasa de respuesta” por unidad de tiempo.
- **Medidas fisiológicas:** son las referentes a los cambios fisiológicos que ocurren como consecuencia de la variación en los niveles de carga y/o fatiga mental. Ejemplos de estos cambios fisiológicos son el “diámetro pupilar”, el “EEG” (las iniciales inglesas del Registro Electroencefalográfico), la “tasa cardíaca”, la “frecuencia fundamental de la voz”, etc.
- **Medidas subjetivas:** es la estimación de la sensación subjetiva de carga o fatiga mental, medida a través de cuestionarios.

A continuación, nos centraremos detalladamente en cada tipo de medida, sin embargo, antes de proceder a detallar los diferentes métodos, resulta conveniente acogernos a unos criterios base que nos sirvan para identificar las ventajas y desventajas de cada método en particular. Así, en la investigación y aplicaciones actuales de los diferentes métodos, se usan los criterios propuestos por O'Donnell y Eggemeier (1986). Tomando como ejemplo la evaluación de la carga mental, estos criterios son los siguientes:

- **Sensibilidad:** se refiere a la capacidad de la técnica para detectar cambios en los niveles de dificultad de la tarea o en sus demandas.
- **Poder de diagnóstico:** un índice debería no sólo identificar cuándo la carga varía, sino también la causa de esta variación. El poder de diagnóstico de una técnica viene dado por su capacidad para detectar el tipo de recursos atencionales que dan

origen a la carga mental de una tarea concreta. Esta información hace posible la implementación de mejores soluciones.

- **Selectividad/Validez:** el índice debería ser selectivamente sensible sólo a las diferencias en las demandas de capacidad y no reflejar los cambios en factores como la carga física o el estrés emocional que pueden no estar relacionados con la carga mental.
- **Grado de intrusión:** el índice no debería interferir con la realización de la tarea cuya carga está siendo evaluada.
- **Fiabilidad:** como cualquier medida de la conducta, un índice de carga debería ser fiable. Si se evalúa la carga mental en momentos diferentes (por ejemplo, si es necesario estudiar los cambios en la carga mental que se producen durante el curso de una tarea u operación), es importante que el índice ofrezca una estimación fiable y consistente de la carga mental.
- **Requisitos de implementación:** incluye aspectos como el tiempo, la instrumentación y el software necesarios para la recogida y análisis de los datos, así como el grado de entrenamiento del operador requerido para poder obtener resultados válidos.
- **Aceptación por el operador:** hace referencia a la percepción que tienen los sujetos de la validez y utilidad del procedimiento.

**Atendiendo a estos criterios, los ergónomos y profesionales de las ciencias del trabajo deciden qué técnica de medida es más apropiada en cada momento y situación.** Por ejemplo, como describiremos más adelante las medidas de ejecución de doble tarea son muy intrusivas, dado que resulta difícil que en una situación real el usuario pueda realizar una tarea simultánea a la tarea principal que constituye su trabajo. En algunos casos, como puede ser el de los pilotos de aeroplano, es incluso peligroso hacerlo; sin embargo, en situaciones donde no existe peligro como en simulaciones, es posible utilizar estas medidas, ya que son muy sensibles y tienen un alto poder de diagnóstico. En

cualquier caso, **la forma más adecuada de realizar una evaluación de carga y/o fatiga mental comprende el uso de al menos una medida de cada tipo de metodología, consiguiendo así una triangulación de información según la cual las carencias de un tipo de medida queda suplida por las ventajas de otras.**

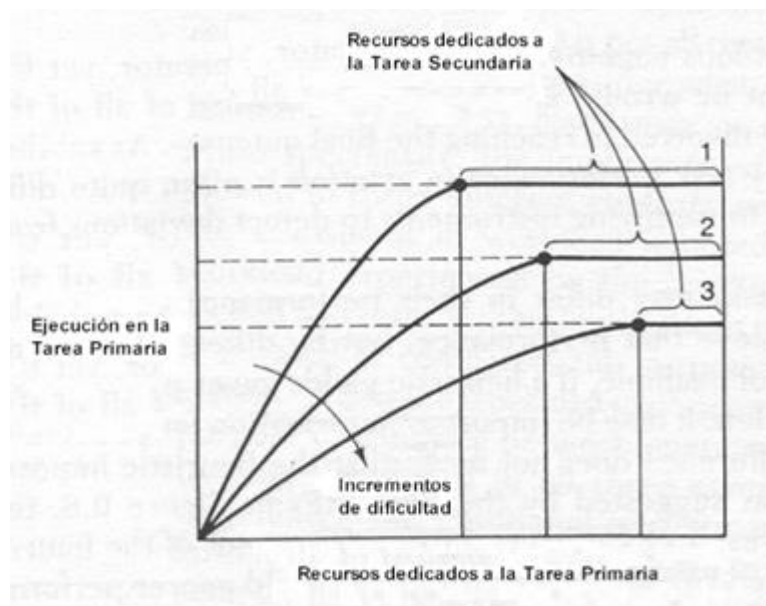
### 3.1. Medidas de rendimiento o ejecución

Las medidas de rendimiento se clasifican en dos grupos principales: **1) medidas de la tarea primaria y 2) medidas de las tareas secundarias.** En primera instancia, atendiendo a los métodos para medir ejecución basados en respuestas discretas de la tarea primaria, tradicionalmente se han considerado dos factores: **la bondad de respuesta y el tiempo de respuesta.** La forma tradicional de medir ejecución en base a respuestas discretas consiste en contabilizar el número de errores o aciertos cometidos durante toda o parte de la ejecución de la tarea primaria. **Se asume que a menor número de errores mejor ha sido la ejecución y viceversa.** De otro lado, también se puede considerar el tiempo de respuesta, esto es, el tiempo que el individuo tarda en emitir su respuesta, de esta manera, **a mayor tiempo de respuesta, peor es la ejecución y viceversa.**

En lo referente a las medidas de las tareas secundarias, se parte de dos supuestos básicos. En primer lugar, se asume que la ejecución de una tarea requiere una cantidad determinada de recursos mentales, de este modo, si se realiza una única tarea, todos los recursos disponibles pueden destinarse por completo a la ejecución de la tarea; sin embargo, **si el número de tareas a realizar es mayor, los recursos disponibles necesariamente deberán ser compartidos entre dichas tareas simultáneas.** El segundo supuesto, pone de manifiesto que, **si un individuo no dispone recursos necesarios para ejecutar satisfactoriamente todas las tareas concurrentes, la ejecución en estas se verá perjudicada** (Wickens, 1984).

De acuerdo con estos supuestos, si queremos evaluar el nivel de recursos que demanda una tarea determinada, a la cual denominamos tarea principal, un procedimiento bien establecido consiste en pedir a la persona que ejecute otra tarea simultáneamente, a la cual denominamos tarea secundaria. La lógica sería la siguiente: si la ejecución en la tarea secundaria es adecuada, podemos asumir que la ejecución en la tarea primaria no necesita

consumir todos los recursos y restan suficientes recursos para ser reemplazados en el procesamiento de la tarea secundaria; si, por el contrario, la ejecución en la tarea secundaria es inadecuada, podemos asumir que la tarea primaria “consume” gran parte de los recursos mentales, quedando una reducida proporción de los mismos dedicada al desempeño de la tarea secundaria (ver Figura 1). Siguiendo esa lógica, **el método de ejecución o rendimiento de “doble tarea”, consiste en evaluar la ejecución en la tarea secundaria para conocer cuál es la demanda de recursos de la tarea primaria.** De este modo, aplicando este método, podemos manipular la complejidad de la tarea primaria (recursos demandados) para poder observar un cambio consecuente en la ejecución de la tarea secundaria. De forma equivalente y siguiendo la lógica inversa, podemos evaluar la ejecución en la tarea primaria, mientras forzamos cierta estabilidad en la ejecución de la tarea secundaria. De esta manera, podemos asumir que la ejecución en la tarea primaria será un reflejo de la demanda de recursos que ésta tiene: **si la ejecución en la tarea primaria es adecuada, podemos asumir que la tarea es sencilla y restan suficientes recursos como para que la intrusión de la tarea secundaria no perjudique el rendimiento en la primaria; si, por el contrario, la intrusión de la tarea secundaria provoca un deterioro en la ejecución de la primaria, podremos inferir que ésta, la primaria, demanda un alto grado de recursos cognitivos.**



**Figura 1.** Relación entre Ejecución y Recursos demandados por las tareas primarias y las tareas secundarias (Wickens, 1984).



Las medidas de ejecución tienen un alto poder diagnóstico porque permiten determinar el origen principal de la demanda conjunta de recursos ya que, de acuerdo al modelo de Recursos Múltiples (Wickens, 2008), cuanto mayor sea la similitud entre los recursos demandados por ambas tareas, mayor será la interferencia entre ellas. De este modo, si la tarea secundaria, por ejemplo, involucra principalmente a procesos memorísticos y observamos que se produce un decremento sustancial en la ejecución de la tarea primaria, podemos asumir que dicha tarea primaria, de forma particular, resulta muy demandante en cuanto a recursos de memoria. Por todo ello, **si partimos de un conjunto de tareas donde están claramente especificados los procesos cognitivos implicados, resulta posible aplicar la técnica de doble tarea para diagnosticar el origen concreto de una sobrecarga mental determinada.** A modo de ejemplo, imaginemos que pensamos que un nuevo diseño de cierta interfaz resulta capaz de disminuir las demandas de procesamiento de Memoria de Trabajo respecto a la versión previa de dicha interfaz. Para probar nuestra hipótesis podemos seguir una estrategia consistente en pedir al participante que, al tiempo que utiliza dicha interfaz, ejecute una de dos tareas secundarias: una en la cual se involucra principalmente la “Memoria de Trabajo” y otra con un elevado componente “Perceptual”. Si nuestra hipótesis es correcta debemos encontrar que el incremento en la demanda de memoria de trabajo suscitado por la tarea secundaria tendrá un menor efecto sobre la nueva interfaz respecto a la versión previa, mientras que el efecto de la demanda de la tarea perceptual debería ser el mismo en ambas interfaces.

### 3.1.1 Ventajas e inconvenientes de las medidas de ejecución

En primer lugar, tal y como se comenta con anterioridad, podemos decir que **la mayor ventaja de las medidas de ejecución es su alto valor diagnóstico.** En ciertas ocasiones no es suficiente con evaluar el grado de carga mental experimentado, sino que lo que se pretende conocer es el proceso cognitivo que se encuentra implicado en mayor medida en la tarea primaria. Por esta razón, como hemos podido ver, si el investigador diseña varias tareas secundarias, cada una implicando un proceso cognitivo diferente (percepción, memoria, etc.), podrá realizar un diagnóstico diferencial de las causas de la sobrecarga, lo cual en última instancia permitirá adoptar las medidas preventivas apropiadas.

De otro lado y como inconveniente principal, podemos destacar el **elevado nivel de intrusividad** del método de doble tarea. Si bien ya resulta difícil que en una situación real el usuario pueda realizar una tarea simultánea a la tarea principal que requiere su trabajo, en algunos casos, como puede ser el de los pilotos dentro de una cabina o el de un controlador aéreo en una posición de control, hacerlo puede resultar incluso arriesgado. Sin embargo, en situaciones donde no existe peligro, como cuando se usan simuladores, es posible aplicar estas medidas ya que son muy sensibles y tienen un alto poder de diagnóstico. En lo concerniente a las medidas de la tarea primaria, la principal desventaja reside en que **puede resultar una estrategia de evaluación poco fiable**, habida cuenta de que una ejecución adecuada no implica necesariamente la inexistencia de sobrecarga o infracarga mental. Por ejemplo, resulta posible que un operario se encuentre muy sobrecargado y/o fatigado y al mismo tiempo no cometa ningún tipo de error, es por esto que el uso de este tipo de medidas (basadas en la bondad de respuesta y en el tiempo de reacción), no resulta suficiente para determinar fielmente el grado de carga y/o fatiga mental experimentado por el individuo evaluado.

### 3.2. Medidas psicofisiológicas

Las medidas psicofisiológicas de carga y fatiga mental comenzaron a ser propuestas por la comunidad científica a tenor de que la investigación comenzara a demostrar la existencia de una relación entre dichos estados psicológicos y ciertos índices de actividad fisiológica. Estos índices fisiológicos **reflejan la actividad cerebral, la cual es responsable del consumo y el agotamiento de los recursos mentales durante la ejecución de tareas cognitivas.**

Para establecer una clasificación de las medidas que han sido propuestas en la literatura, tendremos en cuenta el criterio fundamental del objeto de medición, esto es, si se mide directamente la actividad cerebral o, por el contrario, se mide un parámetro de la actividad fisiológica que se ve afectada por la actividad cerebral. Si bien todas las medidas fisiológicas constituyen índices de la actividad cerebral, cabe explicitar que **algunas de estas medidas implican el registro directo de la actividad neuronal, en tanto que otras constituyen registros de parámetros controlados por la actividad neuronal, por lo que constituyen un reflejo indirecto de la misma.** Por ejemplo, los parámetros

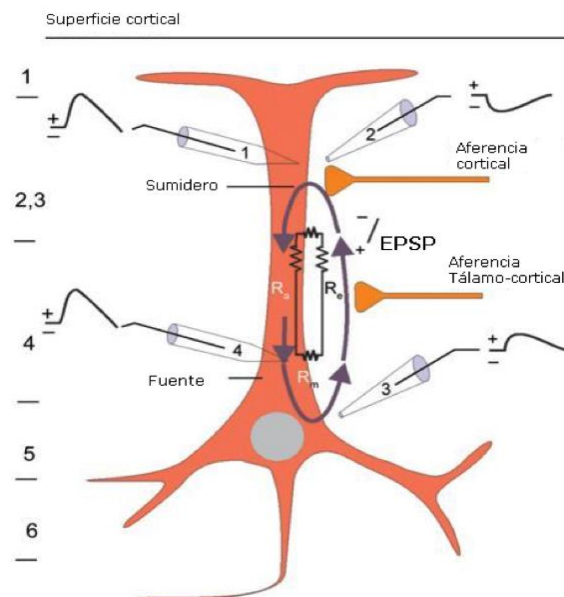
relacionados con la actividad ocular están controlados por el cerebro porque realmente podemos considerar que los ojos son parte del sistema nervioso. Sin embargo, al medir el diámetro de la pupila o las fijaciones y los movimientos sacádicos, no medimos directamente la actividad eléctrica o metabólica de las neuronas, sino las consecuencias indirectas de dicha actividad.

En cualquier caso, debemos señalar que **la mayoría de medidas psicofisiológicas lo que realmente miden es el nivel de activación general del sistema cognitivo** (recursos disponibles). Este nivel de activación se ha tomado tradicionalmente como un valor confiable del nivel de carga mental ya que cabe suponer que, a mayor nivel de activación fisiológica, mayor carga mental se encuentra experimentando el individuo; del mismo modo, cabe suponer que el nivel de activación fisiológica irá disminuyendo a medida que transcurre el tiempo en la tarea, se agotan los recursos y, por ende, comienza a aparecer la fatiga mental. Sin embargo, de acuerdo a la definición de carga mental como la razón entre los recursos demandados y los recursos disponibles, el grado de carga mental puede ser mayor o menor dependiendo de que se produzca una variación tanto en los recursos disponibles como en los demandados. Por tanto, **un valor de carga mental tomado a partir de los parámetros fisiológicos que indican activación, significa que la activación del sistema cognitivo (recursos disponibles) tiene un determinado valor, pero no nos confirma necesariamente el grado de recursos demandados por la tarea.**

### 3.2.1 Actividad eléctrica cerebral

La actividad eléctrica cerebral se registra con la técnica conocida como electroencefalograma (“EEG”, por sus siglas en inglés) y ha sido muy estudiada como medida de carga mental (Murata, 2005; Berka et al., 2007; So et al., 2017; Kartali et al., 2019; Di Flumeri et al., 2019; Acerra et al., 2019). La técnica no es nueva y se ha venido utilizando durante décadas en Psicología, Psiquiatría y Neurología para, fundamentalmente, diagnosticar trastornos neuronales. Sin embargo, también se conoce que **la actividad eléctrica cerebral refleja la intensidad de la actividad neural que se produce durante la realización de tareas cognitivas**, es por esto que hace años que viene explorándose la posibilidad de utilizarla para evaluar carga mental.

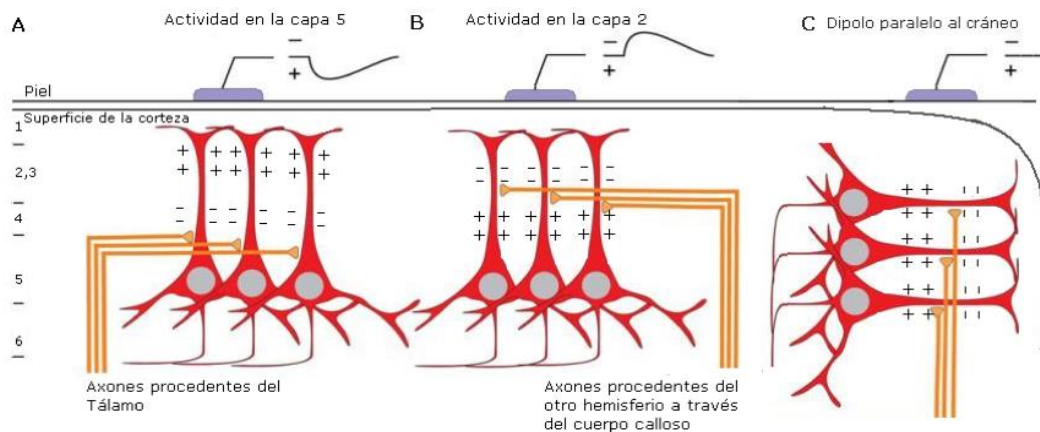
El EEG es una técnica que permite registrar un conjunto de potenciales eléctricos cerebrales a través del contacto de los electrodos con la superficie del cuero cabelludo, cuya actividad no resulta de los potenciales de acción neuronales, sino de la suma de los potenciales postsinápticos de grupos de neuronas. En otras palabras, **la señal EEG refleja las diferencias entre potenciales eléctricos que acontecen en las dendritas de las neuronas apicales de la corteza cerebral** (Kandel & Schwartz, 2000). Estas células están presentes sólo en la corteza cerebral y se caracterizan por estar orientadas siempre de forma perpendicular a la tangente de la superficie de la corteza (véanse las Figuras 2 y 3). Cabe señalar que tanto el propio tejido neuronal, como las membranas meníngeas, el hueso del cráneo y la piel actúan como un filtro “paso-bajo”, lo que genera que los componentes de alta frecuencia de los potenciales de acción disminuyen rápidamente al incrementarse la distancia a la cual se sitúa el electrodo de registro. Se considera, también, que es bastante improbable que la neuroglia (células de sostén del sistema nervioso central) contribuya a la actividad eléctrica registrada en la superficie del cuero cabelludo, aunque, dado su potencial de membrana, puede contribuir a los cambios lentos que se observan a nivel superficial (Kandel & Schwartz, 2000).



**Figura 2.** Flujo de corriente en una neurona apical de la corteza cerebral. La corriente fluye dentro de la dendrita en el lugar de generación del potencial postsináptico excitatorio (EPSP), creando una entrada de corriente (sumidero, el lugar donde la neurona se despolariza). La salida de corriente cierra el circuito, lo que genera una fuente de corriente en la dendrita (o en el soma). El tamaño del cambio de voltaje es predicho por la ley de Ohm ( $V=IR$ ). La resistencia de la membrana ( $R_m$ ) es mayor que la del líquido extracelular y, por ello, el voltaje registrado por un electrodo intracelular (1) es mayor y de polaridad opuesta al registrado por un electrodo extracelular (2), suponiendo que ambos están localizados cerca del sumidero. (Adaptado de Kandel & Schwartz, 2000).

En cualquier caso, **la actividad electroencefalográfica es debida a la actividad de grupos de neuronas de la corteza cerebral**. Por ello, se asume que el EEG refleja la suma de los potenciales postsinápticos de grandes grupos de neuronas, por lo que dicha técnica tiene el **gran inconveniente de no ofrecer una gran resolución espacial**. En apoyo de esta idea está el hecho de que el registro intracerebral de la actividad eléctrica neuronal tiende a coincidir en forma y características con el EEG. Es conveniente tener en cuenta que el cerebro es un conductor en volumen, esto es, capaz de transmitir un campo eléctrico desde una fuente a través del tejido hasta lugares distantes. Esto implica que el EEG constituye, por tanto, una medida de la corriente extracelular resultante de la suma de potenciales postsinápticos de grandes grupos de neuronas, **que pueden estar situados relativamente lejos del electrodo que los registra**. Se ha estimado que cada electrodo “capta” alrededor de 6 cm<sup>2</sup> de corteza (Kandel & Schwartz, 2000).

Cuando se sincronizan muchos inputs de grupos de neuronas, tal y como se muestra en la Figura 4, se genera un dipolo transitorio el cual se registra como una positividad (imagen "A" en la Figura 3) o como una negatividad (imagen "B" en la Figura 3) en la superficie. Este tipo de dipolos orientados verticalmente se registran con facilidad en el EEG, a diferencia de otros dipolos, como el que se muestra en imagen "C" en la Figura 3, que, debido a su localización en una circunvolución cerebral y su orientación ortogonal a la piel, no se registran con tanta facilidad (Kandel & Schwartz, 2000).



**Figura 3.** Organización laminar de la corteza cerebral, con las dendritas apuntando a la superficie y los somas en capas profundas (Adaptado de Kandel & Schwartz, 2000).

El proceso de análisis depende básicamente de si se quiere realizar una aproximación centrada en el dominio del tiempo, o se prefiere una orientación centrada en el dominio

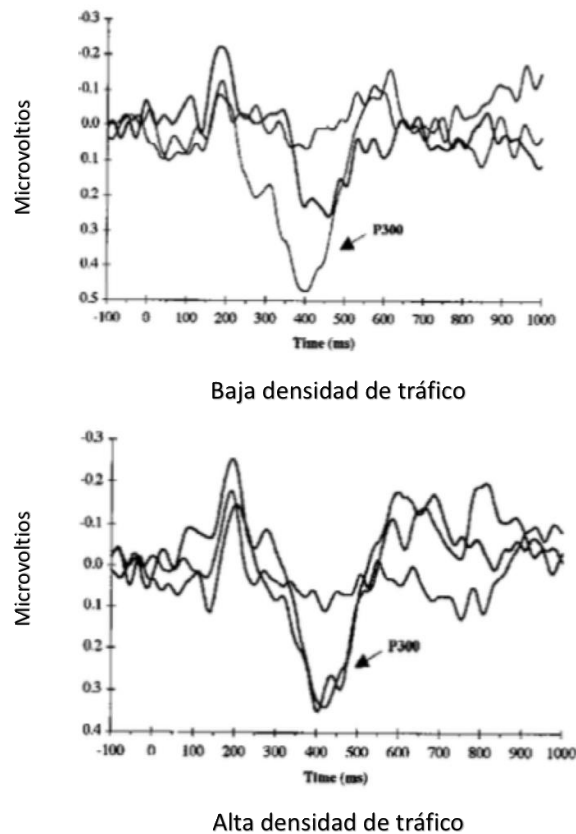
de la frecuencia. La elección entre una y otra depende de los objetivos del estudio. **Lo más frecuente, y también lo más sencillo, es utilizar una aproximación clásica basada en el tiempo, en la que, en esencia, se trata de obtener los potenciales evocados por eventos (ERP, por sus siglas en inglés), y compararlos entre distintas condiciones experimentales.** En la aproximación basada en la frecuencia el objetivo fundamental consiste en obtener las características de amplitud y fase presentes en el EEG bruto y realizar comparaciones entre condiciones experimentales.

En el análisis por el tiempo **el interés fundamental reside en observar la respuesta de una onda registrada por un electrodo, como consecuencia de la ocurrencia de un evento.** Por esta razón se habla de una medida conocida como “Potenciales Evocados” (ERP), los cuales constituyen cambios de potencial (en el orden de unos pocos milivoltios) que se producen milisegundos después de la presentación de una señal.

Los ERPs se componen de una letra ("N" o "P") seguida de un número. La letra "N" determina que se ha producido un cambio de potencial a negativo, en tanto que la letra "P" determina que dicho cambio se ha producido a positivo. El número hace referencia al tiempo, en milisegundos, que ha pasado desde que ocurrió del evento (ej. 100 mse, 200 mse o 300 mse). A modo de ejemplo, "N200" designa que se ha producido un cambio a negativo a los 200 mse, en tanto que "P300" un cambio a positivo a los 300 mse.

**Cabe explicitar que de todos los ERPs posibles en nuestro campo de estudio cobra especial relevancia los ERP comprendidos en el rango de “P300”.** Los ERPs P300 se ven afectados en amplitud y latencia por varios factores: la probabilidad de aparición del objetivo, la incertidumbre del sujeto sobre si el estímulo es un objetivo o no y el esfuerzo que el sujeto dedica a la tarea que realiza. Su latencia se incrementa cuando la categorización del estímulo es más difícil, sin embargo, no se ve afectada por el tiempo de respuesta (mayor o menor) que se requiere del sujeto. Asimismo, se ha demostrado que la amplitud del P300 se reduce conforme aumental el nivel de carga mental (Käthner et al., 2014), es por ellos que podemos considerar al **P300 como un índice adecuado para medir carga mental.** En el ámbito del control aéreo, el P300 ha sido relacionado con la alta carga mental desde hace varias décadas. Por ejemplo, la Figura 4 muestra los resultados obtenidos por Cabon et al., (1997) en dos condiciones (alta y baja) de densidad

de tráfico aéreo. Puede observarse que **en la condición de alta densidad la amplitud del P300 se reduce.**

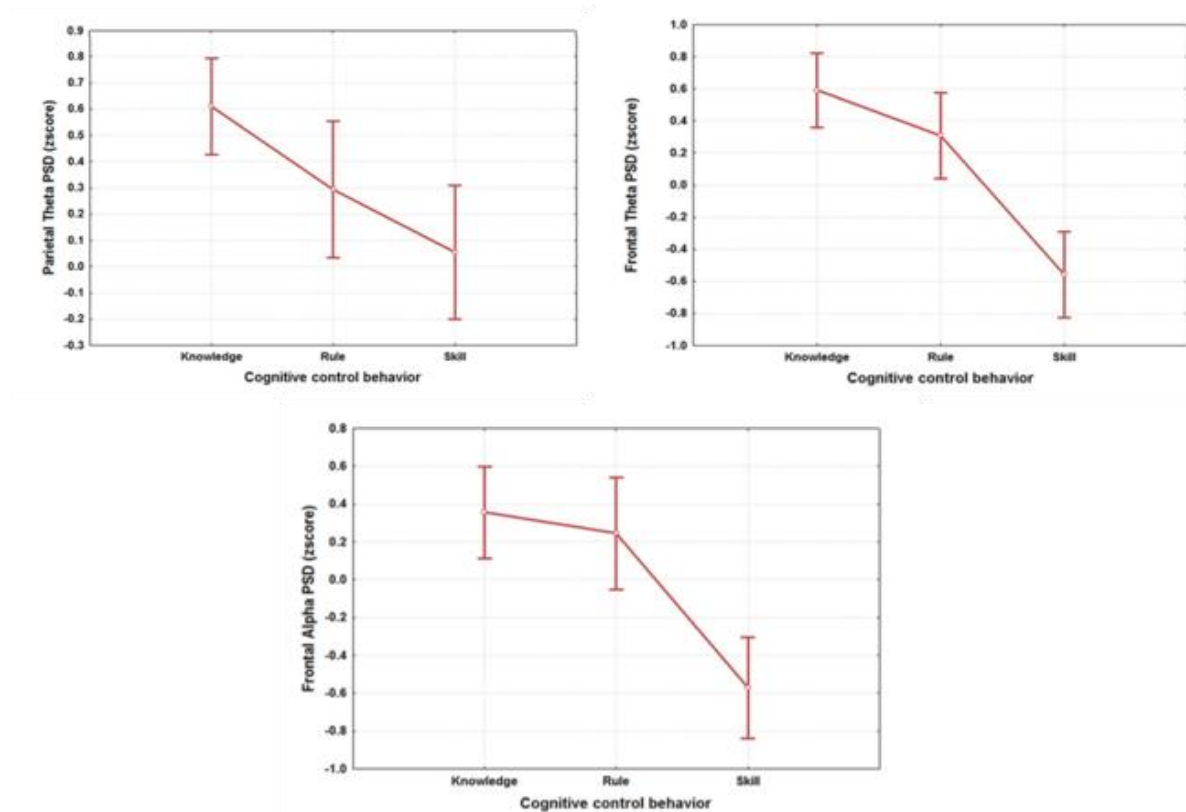


**Figura 4.** Retraso en el P300 como consecuencia de una situación de tráfico denso en el estudio de Cabon et al., (1997).

El análisis del EEG en el dominio de la frecuencia tiene características similares al que se realiza en el dominio del tiempo. **Consiste en determinar los cambios que acontecen en el espectro de potencia (o de amplitud) y de fase que suceden como consecuencia de la presentación de un estímulo (Time-lock to the stimulus) o de la ejecución de una respuesta (Time-lock to the response).** Se obtiene, por tanto, una imagen bidimensional de los cambios, llamada “Perturbaciones en el espectro evocadas por eventos” (ERSP, por sus siglas en inglés). En esencia, el analista clásico tiende a obtener promedios de esos cambios que luego remitirá a análisis estadísticos tipo t de Student o análisis de varianza.

Un ejemplo de análisis por frecuencias lo encontramos en el estudio reciente realizado por Borghini et al. (2017) para identificar las frecuencias que mejor correlacionan con los modos de control definidos por Rasmussen (1983): habilidades (Skills), reglas (Rules) y

conocimiento (Knowledge). Utilizando un análisis estadístico de clasificación, identificaron que las ondas "Theta Frontal", "Theta Parietal" y "Alpha Frontal" permiten identificar el modo de control que se encuentra ejerciendo el participante. Algunos de sus resultados de su estudio pueden verse en la Figura 5.



**Figura 5.** Identificación de los modos de control SRK de Rasmussen (1983) utilizando las frecuencias Theta Frontal, Theta Parietal y Alpha Frontal en el estudio de Borghini et al., (2017).

### 3.2.2 Actividad metabólica cerebral

Aunque el cerebro representa solo alrededor del 2% del peso corporal promedio de un adulto, recibe el 15% del gasto cardíaco, representando entre el 20 y el 25% del consumo total de oxígeno y el 25-30% de la utilización de glucosa del cuerpo (Clark, 1994). Esto ha hecho que desde hace muchos años se haya trabajado en la hipótesis de la relación entre la actividad mental y el consumo de glucosa en el cerebro, es decir, **se asume que, a mayor actividad mental en una zona concreta en el cerebro, se produce un mayor consumo de glucosa.** Esta hipótesis está basada en el principio de acoplamiento neurovascular, también conocido como respuesta hemodinámica o respuesta dependiente del nivel de oxígeno en sangre y, aunque no ha sido aún confirmada completamente



(Hockey, 2013), es la base para técnicas de investigación muy utilizadas actualmente. Entre estas técnicas **podemos destacar el "fNIR (Functional Near Infrared spectroscopy)" y el "fMIR (Functional magnetic resonance imaging)".**

En primera instancia, la espectroscopia funcional de infrarrojo cercano (fNIR) es una modalidad de neuroimagen no invasiva diseñada para monitorizar el cambio hemodinámico, que puede ayudar a identificar los correlatos neuronales del funcionamiento del cerebro humano en función de diferentes eventos y tareas. El método se basa en el principio básico de que la luz que incide sobre el organismo, viaja a través del tejido biológico y se somete a dispersión múltiple y absorción parcial (Cope et al., 1988). Dentro del rango de luz cercano al infrarrojo, el tejido biológico es lo suficientemente transparente como para que el análisis espectroscópico permita la detección de los dos cromóforos principales, "HbO<sub>2</sub>" (Hemoglobina oxigenada) y "Hb" (Hemoglobina desoxigenada). Existen tres implementaciones distintas de "fNIR", concretamente espectroscopía de resolución temporal (TRS, por sus siglas en inglés), dominio de frecuencia y espectroscopia de onda continua (CW, por sus siglas en inglés) (Strangman, Boas, & Sutton, 2002). En los sistemas "TRS", se aplican pulsos de luz incidente extremadamente cortos al tejido, y se miden las distribuciones temporales de los fotones, que transportan información sobre la dispersión y absorción del tejido. En los sistemas de dominio de frecuencia, la fuente de luz se modula en amplitud a diferentes frecuencias (decenas a cientos de mHz), posteriormente se realiza una medición de la señal detectada con respecto a la señal incidente, para poder, así, caracterizar las propiedades ópticas del tejido a partir de la disminución que se registra en la amplitud de la señal y su desplazamiento (Boas, 2002). En tercer lugar, respecto a los sistemas "CW", la luz se aplica continuamente al tejido a amplitud constante y posteriormente se registra la atenuación de la amplitud de la luz incidente. De los tres sistemas el "CW" es el más utilizado en los campos aplicados por su portabilidad, su facilidad de uso y el hecho de que no es una técnica demasiado invasiva (Izzetoglu et al., 2004). **Una característica importante de la fNIR es que, como sugieren resultados recientes, es capaz de detectar cambios en la magnitud de carga mental, lo que se ha denominado como "transiciones en la carga mental".** En un estudio reciente, McKendrick (2016) desveló que la **actividad del córtex prefrontal se atenúa con la sobrecarga y que se puede establecer una función polinómica cúbica entre la carga cognitiva y dicha actividad**

registrada mediante "fNIR", lo cual puede ser utilizado para medir transiciones en carga mental.

De otro lado, el "fMIR" es una técnica basada en el hecho de que la hemoglobina desoxigenada "Hb" tiene ciertas propiedades magnéticas (paramagnética), a diferencia de la hemoglobina oxigenada "HbO<sub>2</sub>", que es virtualmente resistente al magnetismo (diamagnética). Esta diferencia conduce a una señal mejorada que se puede mapear para mostrar qué neuronas se encuentran activas simultáneamente. **Pese a las ventajas de esta técnica por su precisión, se ha descartado de momento para las investigaciones aplicadas debido a los elevados requisitos para su implementación que la hacen inviable en ámbitos aplicados.** En la Figura 6 puede verse un equipo de fMIR y se puede comprobar que no podría ser utilizado para medir la carga o fatiga mental de un operario.



Figura 6. Equipo de fMIR.

En cualquier caso, a pesar de que el uso de las medidas basadas en la actividad metabólica del cerebro es cada vez más popular en muchos ámbitos, **existen serias dudas respecto a su validez.** Estas dudas han sido expuestas por Hockey (2013) quien, después de realizar una amplia revisión acerca de las investigaciones que se han realizado para probar la correlación entre actividad cerebral y el consumo de glucosa, concluye lo siguiente:

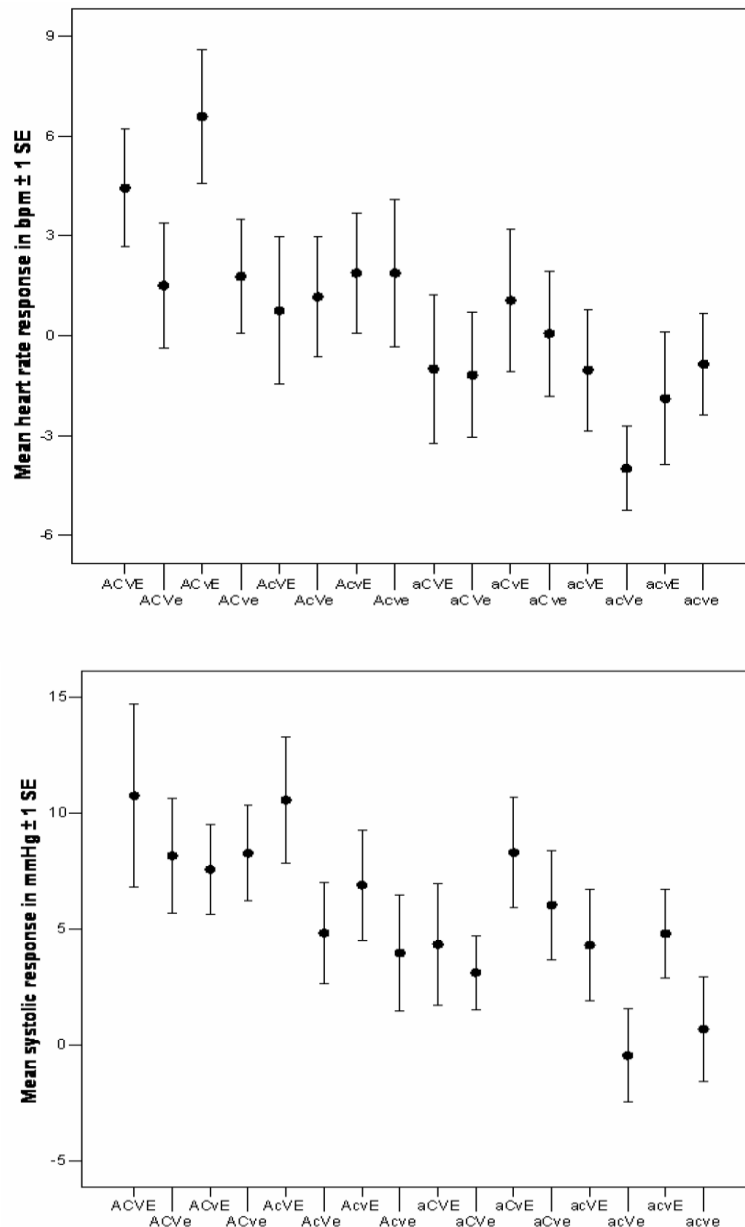
- Es cierto que el cerebro utiliza mucha energía, pero también resulta cierto que la mayor parte es usada para mantenimiento del tejido cerebral (incluso durante el descanso).
- La actividad mental puede o no incrementar la demanda de energía, pero su impacto es muy pequeño.
- Algunas tareas mentales implican un cambio de demanda de energía de un área a otra del cerebro, pero el cambio implica no más de un 1% del consumo de energía de un área a otra.
- En general, no existe una evidencia clara de que las tareas que demandan más procesamiento cognitivo consuman más energía.

Por lo tanto, tomando estas conclusiones de Hockey (2013) debemos decir que, **aunque los resultados obtenidos con las medidas metabólicas son muy prometedores, resulta necesario arrojar más luz sobre estas mediante más investigación, antes de poder considerarlas como una medida válida, fiable y eficaz de carga mental y fatiga mental.**

### 3.2.3 Frecuencia cardíaca y respiración

Los parámetros relacionados con la frecuencia cardíaca y la respiración han sido muy utilizados para medir carga mental durante los últimos 20 años (Meshkati, 1988; Jorna, 1992; Wientjes, 1992; Roscoe, 1992; Henelius et al., 2009; Delliaux et al., 2019). Por ejemplo, en el ámbito del control de tráfico aéreo, nos encontramos el estudio que Vogt, Hagemann & Kastner (2006) realizaron para explorar las posibles correlaciones entre los parámetros de la actividad cardiovascular y la carga mental en controladores de torre y en ruta. Estos autores encontraron **correlaciones positivas significativas entre respuestas cardiovasculares y la cantidad de aviones bajo control**, especialmente la densidad de aviones, sus velocidades y su evolución (ascendiendo o descendiendo). La frecuencia cardíaca, la presión arterial sistólica, la concentración autoinformada y el

malestar fueron significativamente mayores en las simulaciones ejecutadas con 12 aeronaves controladas, respecto a las simulaciones en las que había solo 6. Una cantidad alta versus baja de conflictos potenciales entre aeronaves en la configuración en ruta también causó aumentos de la frecuencia cardíaca, presión arterial sistólica, concentración autoinformada y malestar. Un ejemplo de los resultados puede verse en la Figura 7.



**Figura 7.** Cambios promedio y errores estándar en la frecuencia cardíaca y presión sistólica ( $n=21$ ) como una función de la dificultad de la tarea en ruta en el estudio de Vogt, Hagemann & Kastner (2006). Las condiciones experimentales deben leerse como: A, a: 12 vs. 6 aviones continuamente bajo control; C, c: 6 vs. 2 conflictos potenciales en un punto; V, v: 66% vs. 33% movimientos verticales; E, e: 2 vs. 0 errores de los pilotos en un punto.

Sin embargo, el uso de la frecuencia cardíaca ha sido criticado a menudo por sus limitaciones (por ejemplo, Kahneman, 1973) y las mediciones respiratorias todavía son bastante poco comunes en la investigación aplicada. Además, **una serie de problemas técnicos asociados con la adquisición de estos dos índices hace que sea imposible aplicarlos en entornos de trabajo**. Dos fuentes comunes de problemas metodológicos para estos índices, en entornos aplicados, provienen del habla y la actividad muscular (Jorna, 1992; Porges & Byrne, 1992). Así, por ejemplo, es imposible realizar una evaluación de un controlador de tránsito aéreo o un operador de una refinería limitando sus movimientos o sus interacciones verbales. Además, como sucede con las mediciones fisiológicas más comunes, los índices cardiorrespiratorios **no logran resultados confiables porque carecen de especificidad y tienen una hiper-hipo sensibilidad a diferentes factores psicológicos, emocionales y ambientales no relacionados con la carga de trabajo** (Roscoe, 1992). Estos problemas son importantes porque en contextos aplicados es imposible definir cargas de tareas estables y, por lo tanto, es imposible definir relaciones causales claras (Mulder, 1992).

#### 3.2.4 Respuesta psicogalvánica

La piel es una barrera selectiva que sirve para evitar que las sustancias externas contaminen el organismo y para que otras sustancias desechables que circulan por la sangre sean expulsadas hacia el exterior. Estos procesos se llevan a cabo mediante el mantenimiento de un balance homeostático de agua y temperatura que se produce mediante la vasoconstricción y vasodilatación, así como a través de la sudoración. Todo este proceso está controlado por centros cerebrales y, por ende, de alguna forma refleja su actividad. Por esta razón, **se llevan muchos años midiendo la actividad de los procesos que ocurren en la piel como método para medir ciertos procesos mentales**.

Una forma tradicional de medir los procesos fisiológicos que ocurren en la piel ha sido mediante la **medición de la concentración salina que tiene el sudor en un momento determinado** (Dawson, Schell & Fillion, 2007). El sudor es segregado por dos tipos de glándulas: la "apocrina" y la "ecrina", no obstante, la que ha interesado a los psicofisiólogos ha sido la "ecrina", ya que ejerce una función termorreguladora y se ve afectada por factores psicológicos. El procedimiento de medición de la concentración

salina del sudor implica el registro de la conductancia eléctrica entre dos puntos de la piel: se colocan dos electrodos en dos puntos cercanos de la piel (por ejemplo, en dos puntos del brazo) y se hace pasar una pequeña corriente eléctrica. De este modo, **al medir la conductancia se está midiendo, en última instancia, la concentración salina de la piel, de manera que, a mayor concentración salina, mayor conductancia se produce.** Tradicionalmente se suelen obtener dos índices: el nivel de conductancia y la respuesta de conductancia a la presentación de un estímulo. En la Figura 8 se pueden ver los componentes de un equipo de registro de respuesta psicogalvánica.



**Figura 8.** Registro de la respuesta psicogalvánica de la piel.

El registro psicogalvánico de la piel ha sido utilizado en Factores Humanos para medir carga mental, dando como resultado un aumento de la conductancia de la piel que se produce con el incremento de esta (Gould et al., 2009; Díaz & Robles 2018). La **principal desventaja de esta metodología reside en la dificultad para medir la conductancia de la piel cuando la persona se está moviendo.** Por ejemplo, Vogt, Hagemann & Kastner (2009) utilizaron la conductancia de la piel, así como un conjunto de medidas de carga mental en su estudio con controladores de ruta. Los resultados encontrados por estos autores mostraron ausencia de correlación entre la conductancia de la piel y la carga mental de los controladores, posiblemente debido a las dificultades que encontraron al

medir conductancia cuando la persona se encuentra en movimiento, como es el caso de un controlador realizando su tarea de control. **Este problema ocurre también cuando se mide la tasa cardiaca y la respiración** (Yoshino et al., 2007).

### 3.2.5 Parámetros oculares

Durante los últimos años, con el incremento en accesibilidad de los sistemas de monitorización del comportamiento ocular conocido como “eye-tracking”, se ha venido incrementando el interés por las métricas psicofisiológicas oculares. Cabe señalar que, en ciertos ámbitos, como el ámbito ATM (Air Traffic Management, por sus siglas en inglés), el interés ha sido muy elevado, dada la necesidad específica de utilizar sistemas no invasivos de monitorización de la carga mental y la fatiga en tiempo real.

Mediante el registro de la conducta ocular, **pueden estimarse métricas que reflejan fatiga, somnolencia, distracción visual, distracción cognitiva y carga de trabajo** (Lee & Cuijpers, 2013; Lenné & Jacobs, 2016). A modo de ejemplo la desviación estándar de la mirada es una métrica comúnmente utilizada para reflejar carga mental, habida cuenta de que los operarios muestran un estrechamiento del campo visual en respuesta a una mayor carga cognitiva (Wang et al., 2014). De forma similar, el diámetro pupilar también se ha visto relacionado con el nivel de carga mental (Hoeks y Levelt, 1993; Mittner et al., 2014). **Los sistemas de eye-tracking permiten monitorizar y registrar en tiempo real todo el conjunto de parámetros oculares que comprenden el comportamiento visual:** áreas de fijación, frecuencia y duración de parpadeos, velocidad de movimientos sacádicos, diámetro pupilar, etc. La metodología de eyetracking parte de la hipótesis “ojo-mente”, según la cual si un individuo se encuentra mirando algo es porque tiene su atención centrada en ese algo (Just & Carpenter 1980). En la Figura 9 y 10 se pueden ver los componentes de un equipo de eyetracking de tipo fijo y portátil respectivamente.



**Figura 9.** Ejemplo de sistema eye-tracking fijo.



**Figura 10.** Ejemplo de sistema de eye-tracking portátil.

A continuación, se expondrán los principales parámetros oculares utilizados para la evaluación de la actividad cerebral.

#### 3.2.5.1 Diámetro pupilar

La pupila, abertura situada en el centro de nuestro iris, es la encargada de regular la luz entrante del exterior a nuestro globo ocular, de la que dependen sus cambios de tamaño, generalmente comprendido entre 2 mm (condiciones de alta luminancia) y 8 mm (oscuridad). El iris, músculo y tejido pigmentado que rodea la pupila, es el responsable de los cambios en el tamaño pupilar a través de dos grupos de músculos (Beatty & Lucero-Wagoner, 2000). El primero de ellos, el dilatador pupilar, tiene forma radial y su



contracción produce una retracción del iris que provoca un aumento del diámetro de la pupila. Está principalmente innervado por fibras del sistema nervioso simpático y la noradrenalina es el neurotransmisor implicado en las sinapsis entre las fibras simpáticas eferentes postganglionares. El segundo músculo es el esfínter pupilar. Tiene sus fibras organizadas en forma de anillo y su contracción refleja una disminución de la pupila por la expansión del iris. En este caso, la innervación procede de las fibras parasimpáticas postganglionares y el neurotransmisor implicado es la acetilcolina (Beatty & Lucero-Wagoner, 2000).

En general, se admite que la dilatación pupilar correlaciona de modo muy elevado con la actividad del locus coeruleus (LC), hasta el punto de considerarse como un indicador directo de la actividad de esta zona del tallo cerebral (Murphy et al., 2011). El LC es una región clave en el sistema noradrenérgico (Laeng, Sirois & Gredeback, 2012).

El reflejo psicosensoresal ha sido considerado en numerosas ocasiones como uno de los componentes de la respuesta de orientación (Lynn, 1966; Sokolov, 1963). Sin embargo, **la relación fundamental histórica ha sido establecida entre la dilatación pupilar y la activación y, por tanto, de la carga mental y el esfuerzo mental**, a modo de ejemplo:

- Hess & Polt (1960) encontraron que el diámetro pupilar aumentaba a medida que incrementaba el nivel de dificultad de un conjunto de multiplicaciones.
- En 1965, Hess obtuvo los mismos resultados con una tarea de deletreo de palabras, de manera que conforme aumentaba la dificultad de la tarea, mayor dilatación pupilar se producía.
- Lo mismo acontece con tareas de recuerdo, a mayor dificultad, mayor dilatación pupilar (Elshtain & Schaefer, 1968; Peavler, 1974).
- En tareas de cálculo mental de diferente índole se repetían los mismos resultados (Bradshaw, 1967; Payne, Parry & Harasymiw, 1968; Schaefer et al., 1968), así como en tareas de procesamiento continuo (Bradshaw, 1968).

Para obtener una revisión extensa de estos y otros trabajos, se puede consultar Goldwater (1972) y Beatty (1982).

Ahora bien, fueron los estudios de Kahneman (1973) enmarcados dentro de su teoría atencional, los que alcanzaron mayor repercusión (Beatty & Kahneman, 1966; Kahneman & Beatty, 1967; Kahneman, Onuska & Wolman, 1968; Kahneman, Peavler & Onuska, 1968), llegando a proponerse a partir de entonces la utilización de la **dilatación pupilar como indicador adecuado del grado de procesamiento cognitivo** (recursos atencionales) implicado en el desarrollo de una tarea determinada. Dichos estudios fueron el origen de numerosos trabajos aparecidos en la década de los 90, los cuales **utilizaban el diámetro pupilar como índice de esfuerzo mental** (Matthews et al., 1991; Backs & Walrath, 1992; Hyönä, Tommola & Alaja, 1995; Granholm et al., 1996).

En la actualizad, **numerosos estudios continúan confirmando la relación entre la activación y diámetro pupilar** (Iqbal, Zheng & Bailey, 2004; Verney, Granholm & Marshall, 2004; Porter, Troscianko & Gilchrist, 2007; Privitera et al., 2010; Reiner & Gelfeld, 2014; etc.). Cabe destacar los hallazgos de Steinhauer et al., (2004), en lo referente al control del sistema nervioso ejercido sobre los procesos de cambio pupilar como consecuencia diferencial del esfuerzo cognitivo o del procesamiento de información emocional: estos autores señalan como responsables de la dilatación pupilar acontecida en el primer caso (a diferencia del segundo), a la inhibición cortical de las fibras parasimpáticas del núcleo oculomotor.

A tenor de las líneas anteriores, **queda explicitado que el diámetro de la pupila es un buen indicador del nivel de activación en tareas cognitivas** (Jones et al., 2010; Piquado, Isaacowitz & Wingfield, 2010; Muñoz-de-Escalona & Cañas, 2018; Muñoz-de-Escalona, Cañas & van Nes, 2019; Muñoz-de-Escalona, Cañas & Noriega 2020). Sin embargo, existen algunas dificultades metodológicas comunes a todos estos procedimientos que merecen atención. Entre todas estas dificultades destaca por su importancia el control de los efectos de la luminosidad sobre el diámetro pupilar, debido a que pueden llegar a contaminar los efectos de la carga mental si no se mantienen controlados.

Tal y como se comenta más arriba, **el diámetro pupilar depende primordialmente de las condiciones de luminosidad**: el reflejo fotomotor contrae la pupila en respuesta a los incrementos en luminosidad y viceversa. El diámetro pupilar puede oscilar varios milímetros en respuesta al reflejo fotomotor, en contraste, los cambios relacionados con el efecto de la carga mental son de orden inferior, usualmente entre 0.1 y 0.5 mm (Beatty, 1982). De esta manera, estos pequeños cambios pueden quedar enmascarados por el efecto de la luminosidad. Así, **la adecuada predicción de carga mental a partir del diámetro pupilar queda supeditada a la capacidad para disociar los diferentes efectos ejercidos sobre dicha medida**. A este respecto, la solución más utilizada en investigación pasa por mantener las condiciones de iluminación constantes tanto en la sala como en los monitores, de manera que el efecto de la luminosidad queda controlado. Sin embargo, si bien esta solución no resulta problemática en condiciones de laboratorio, en condiciones reales lo más probable es que no pueda garantizarse.

#### 3.2.5.2 Patrones de mirada

La literatura científica ha desvelado la existencia de una **correlación negativa entre carga mental y la amplitud visual, de manera que cuando mayor es la carga mental experimentada, menor resulta la amplitud del patrón de movimiento ocular** (Tsai et al., 2007). Este hallazgo resulta consistente con el fenómeno de “visión en túnel”, según el cual el comportamiento visual de escaneo disminuye a medida que aumenta el nivel de demanda cognitiva de la tarea. De forma anecdótica, resulta bastante común escuchar a los ATCo (Air Traffic Controller, por sus siglas en inglés) decir que mantienen su mirada focalizada en el radar en periodos de alta demanda cognitiva. Esto es, los ATCo afirman centrar su mirada y su atención hacia elementos concretos del radar a expensas de obviar otro tipo de información (fenómeno conocido como “visión en túnel”) (Kuo et al., 2017). Este fenómeno también ha sido reportado en el contexto de la automoción: los conductores que circulan realizando tareas de alta demanda cognitiva emplean más tiempo centrando su atención en el centro de la carretera, a expensas de un peor comportamiento de escaneo visual del entorno (Victor, Harbluk, y Engstöm, 2005).

### 3.2.5.3. Fijaciones

Definimos las fijaciones como el tipo de comportamiento ocular que se utiliza para observar un área visible específica. **La frecuencia de las fijaciones (nº de fijaciones por unidad de tiempo) incrementa con el aumento en carga mental y disminuye cuando el individuo se encuentra en una situación de sobrecarga mental** (He et al., 2012).

La duración de las fijaciones hace referencia al tiempo que toma el ojo focalizando un punto específico y se considera el parámetro más utilizado en la investigación del comportamiento ocular, de acuerdo a Holmqvist y colaboradores (2011). En el ámbito del control del tráfico aéreo Callan (1999) realizó una investigación donde encontró que las altas frecuencias de fijaciones largas (mayores de 500 ms) indicaban que los controladores estaban dedicados a actividades de resolución de problemas. Los resultados de estos autores mostraban que **las fijaciones largas estaban relacionadas con las violaciones de los controladores que podían llevar a conflictos. Por el contrario, las fijaciones cortas indicaban un procesamiento eficiente.**

### 3.2.5.4. Parámetros de velocidad

Los movimientos rápidos de los ojos, conocidos como movimientos sacádicos, son movimientos balísticos que permiten acercar sucesivas regiones de interés a la fovea, y son cruciales para navegar en el mundo visual, desde leer e interactuar con objetos hasta realizar tareas de monitoreo visual exigentes como el tráfico aéreo o el control de la plataforma nuclear (Wang, 1998). Estos movimientos tienen un inicio, una dirección y, sobre todo una velocidad. Aunque era conocido desde hace años en algunas áreas de la medicina, sobretodo en Psiquiatría, que esta velocidad puede ser afectada por procesos mentales, no ha sido hasta recientemente que los investigadores en Factores Humanos han mostrado interés en la velocidad como parámetro relacionado con la carga mental.

En general, los parámetros de la dinámica de los movimientos sacádicos se han utilizado como marcadores de la ejecución en la tarea (Galley, 1998). Además, la investigación empírica también ha mostrado relaciones entre la dinámica sacádica y el estado de activación (Galley, 1989; Galley, 1993; Galley & Andrès, 1996; Schleicher et

al., 2008). Si bien **la amplitud sacádica y la latencia sacádica se pueden utilizar como indicadores del rendimiento, la velocidad sacádica se relaciona con el estado de activación en las tareas de rendimiento visual** (App & Debus, 1998). Por lo tanto, es razonable hipotetizar que, en tareas visuales de larga duración (por ejemplo, monitorización de dispositivos), la velocidad sacádica podría cambiar de acuerdo con el estado de activación. Sin embargo, como han sugerido App & Debus (1998), la velocidad sacádica (media) no se ha utilizado como un indicador de estado mental porque depende fuertemente de la amplitud sacádica y la dirección de la órbita, dos variables que en contextos reales generalmente no están controladas. Por esta razón, **se ha propuesto que sea el Pico de Velocidad el que pueda ser utilizado para medir activación, ya que no está vinculado a priori por una fórmula matemática a la amplitud o duración sacádica y podría representar un posible índice sensible de la complejidad de la tarea**. La velocidad media, el pico de velocidad (cuando la aceleración es cero) y la duración del movimiento aumentan en función de la amplitud sacádica, una relación conocida como secuencia principal (“Main sequence”) (Zuber et al., 1965; Bahill et al., 1975). Este pico de velocidad se ha utilizado para interrelacionar varias hipótesis relativas a la generación y el control de movimientos sacádicos (Bahill et al., 1975). Durante los primeros 20 ms (más o menos) de un movimiento sacádico, la velocidad tiende a ser la misma independientemente de la posición del objetivo, sin embargo, durante los siguientes 80 ms (o más), la posición del objetivo afecta la aceleración sacádica, que aumenta hasta un punto, antes de que la velocidad disminuya ligeramente y se mantenga hasta alcanzar el objetivo. La velocidad máxima es el punto en el cual la aceleración se convierte en negativa, es decir, el punto en el que se alcanza la velocidad máxima del movimiento sacádico. Esta velocidad máxima es independiente de la duración del movimiento sacádico ya que no está vinculado a priori por una relación matemática. La medición de la velocidad máxima no depende de los umbrales utilizados para definir los puntos de inicio y final de una sacada, mientras que la duración de la sacada sí lo hace (Becker, 1989).

**La investigación está mostrando que el pico de velocidad puede ser un buen indicador de la carga mental** (Di Stasi et al., 2010; Di Stasi et al., 2010; Di Stasi et al., 2012; Di Stasi et al., 2013). Sin embargo, existe un problema instrumental que está retrasando el uso de este parámetro en la investigación aplicada, **problema que se deriva del hecho de que para poder medirlo resulta necesario disponer de costosos equipos**

**de registro de movimientos oculares que han de caracterizarse por ofrecer una elevada resolución** (por encima de 500 Hz). Además, actualmente, los sistemas de registro de altas frecuencias de muestreo requieren que la persona mantenga la cabeza inmovilizada, impidiendo esto el comportamiento natural del operario. En otras palabras, en situaciones aplicadas donde la persona debe mover su cabeza para ejecutar sus funciones, como puede ser el caso de un piloto, no es posible utilizar estos equipos de registros actualmente.

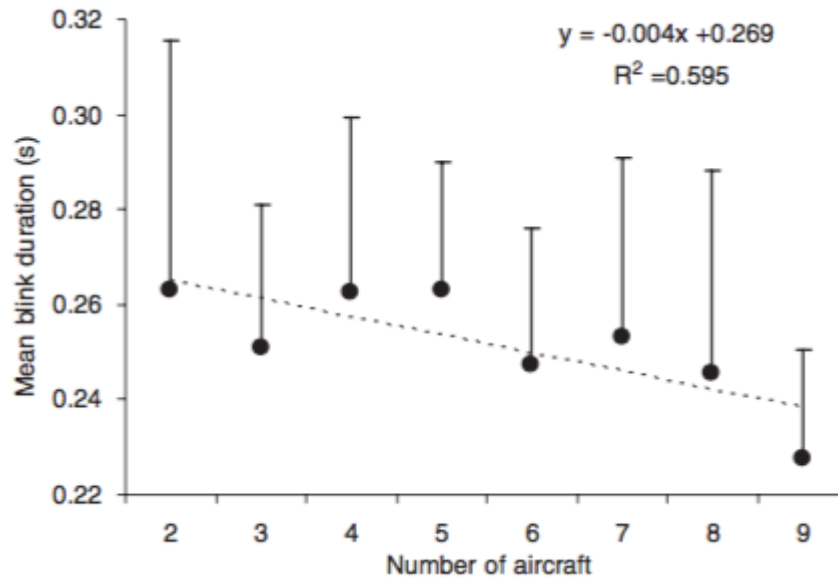
#### 3.2.5.5. Parámetros de parpadeo

Holmqvist et al., (2011) define el término parpadeo como el fenómeno en que las pestañas comienzan a descender para cerrar el globo ocular, momento en el cual el diámetro pupilar y el reflejo corneal no pueden ser medidos. Holmqvist enfatiza el hecho de que el parpadeo constituye una medida de movimiento palpebral, más que una medida ocular propiamente dicha. De otro lado, Bodala et al., (2016) consideran el movimiento del párpado como un parpadeo cuando el diámetro pupilar es inferior a 0.5. En cualquier caso, podemos considerar que se ha producido un parpadeo cuando el valor del diámetro pupilar es igual o cercano al cero. La relación entre carga mental y el comportamiento de parpadeo ha sido estudiada en numerosos estudios. El parpadeo se considera generalmente como una medida útil tanto de carga mental como de fatiga (Benedetto et al., 2011). De forma general, **las medidas de parpadeo comprenden tanto el número como la frecuencia de parpadeos, aunque es la frecuencia de parpadeo la que suele utilizarse con más frecuencia.**

El término frecuencia de parpadeo, a veces denominado tasa de parpadeo, hace referencia al número de parpadeos que se producen por unidad de tiempo (generalmente segundos o minutos). Schneider & Deml (2016) lo definen como el número de parpadeos que se producen en un periodo predefinido de tiempo. Holmqvist et al., (2011) han señalado que la media de parpadeos en una tarea típica de lectura ronda los 3-7 parpadeos por minuto, en tanto que dicha frecuencia aumenta hasta los 15-30 parpadeos en tareas que no implican lectura. **La mayoría de investigaciones realizadas ponen de manifiesto la existencia de una relación inversa entre la frecuencia de parpadeo y el nivel de carga mental experimentado.** A modo de ejemplo, Zheng et al., 2012, en un

experimento con 23 cirujanos, desvelaron que aquellos participantes que tenían una menor frecuencia de parpadeo obtenían valores más elevados de carga mental. De forma parecida, Ledger (2013) encontró en un estudio con 30 estudiantes de psicología que la tasa de parpadeo descendía conforme aumentaba el nivel de carga mental.

Con respecto a la duración del parpadeo, Holmqvist et al., (2011) la definen como el tiempo que toma el párpado en cerrar el globo ocular y regresar a su posición original. La duración del parpadeo constituye un parámetro más sensible y adecuado para predecir carga mental comparado con la frecuencia de parpadeo (Benedetto et al., 2011). Estos autores señalan que el aumento en el número de parpadeos de corta duración es indicativo de la presencia de alta carga mental. En otras palabras, **la duración media del parpadeo tiende a disminuir con el incremento en carga mental**. La explicación de este fenómeno reside en el hecho de que las personas tendemos a disminuir la duración del parpadeo en condiciones de alta demanda cognitiva, al objeto de disminuir lo máximo posible la pérdida de información causada por el proceso de parpadeo. En favor de esta explicación están los resultados de recientes estudios que han revelado la existencia de un mayor intervalo temporal entre parpadeos en una tarea de identificación visual comparado a una tarea de detección visual (Bodala et al., 2016). En suma, **la duración del parpadeo, así como la frecuencia de parpadeo, disminuyen en respuesta al incremento en dificultad** (Benedetto et al., 2011). También se ha encontrado que la duración y frecuencia del parpadeo disminuyen en respuesta al incremento en dificultad en un juego de identificación visual (Van orden et al., 2001). La referencia clásica es el estudio de Ahlstrom & Friedman-Berg (2006) que mostró que los ATCo muestran una menor duración del parpadeo en respuesta al aumento en la dificultad de la tarea. El objetivo del estudio era comparar diferentes displays para la presentación de información meteorológica en cuanto a la carga mental que los ATCos soportaban con cada display. Como se puede ver en la Figura 11, los resultados desvelaron que la duración media de los parpadeos disminuía conforme aumentaba el tráfico aéreo en el monitor.



**Figura 11.** Duración del parpadeo en función del número de aviones en la pantalla en el estudio de Ahlstrom & Friedman-Berg (2006).

Finalmente, hay que señalar que existen una serie de factores que afectan al parpadeo. Más concretamente, **se ve afectado por factores como la sequedad de ojos, la polución, el uso de monitores, el tiempo empleado en la tarea, el momento del día, etc.**, que podrían contaminar la recogida de datos fiables y válidos.

### 3.2.6 Ventajas e inconvenientes generales de las medidas psicofisiológicas

Considerando a las medidas fisiológicas en su conjunto, podemos tener la impresión de que existen numerosas dificultades para obtener, analizar e interpretar este tipo de medidas. Esto ha llevado a los investigadores a plantearse si verdaderamente los registros fisiológicos nos pueden aportar alguna información útil que no pueda ser obtenida mediante medidas subjetivas o de rendimiento, por su naturaleza de registro más sencilla y económica. Sin embargo, **lo que la investigación básica y aplicada nos desvela es la gran relevancia de considerar el registro de este tipo de medidas, dado que nos permite obtener información única y útil acerca de la actividad del sistema nervioso en relación con la carga y fatiga mental, que no es posible obtener con las medidas de ejecución o las medidas subjetivas.**

Podemos destacar varias **ventajas** de las medidas fisiológicas. En primer lugar, considerando su naturaleza multidimensional, estas medidas **nos permiten estudiar y**



**evaluar diferentes perspectivas de la carga y la fatiga mental.** A modo de ejemplo, consideremos el subgrupo de medidas constituido por las medidas fisiológicas del sistema nervioso central (ej. electroencefalograma (EEG), potenciales evocados (ERPs), resonancia magnética funcional (fMIR)): cada una de estas técnicas refleja aspectos únicos de la carga mental, pudiéndose además subdividirse en varios índices para ofrecer análisis más detallados del procesamiento de la información. Por ejemplo, los ERPs se descomponen tradicionalmente en un conjunto de componentes categorizados en base a criterios espaciales y temporales. Asimismo, diferentes aspectos de estos componentes, tales como la latencia y amplitud, han mostrado ser diferencialmente sensibles a dimensiones cronométricas y energéticas del procesamiento de la información humana. Otra de las ventajas más relevantes de este tipo de medidas es que **pueden mostrarse sensibles a los cambios físicos (abruptos y temporales) de carga mental.** En última instancia, si bien las medidas subjetivas de carga mental pueden verse afectadas por factores como la deseabilidad social, **las medidas psicofisiológicas gozarán de mayor validez interna,** toda vez que este tipo de factores psicosociales no tendrán cabida, dado el carácter objetivo de las medidas psicofisiológicas.

En cualquier caso, no debemos olvidar las obvias **desventajas** de las medidas fisiológicas. Primeramente, si bien el coste de los sistemas de registro fisiológicos ha disminuido significativamente durante las últimas décadas, bien es cierto que la necesidad de disponer de un equipamiento especializado para recabar datos fisiológicos **resulta más costoso que el uso de metodologías de registro de datos subjetivos o de rendimiento** (mayores requisitos de implementación). Asimismo, los procedimientos de estimación no se encuentran tan estandarizados como en el caso de las medidas subjetivas y de rendimiento, **requiriendo un mayor nivel de experiencia y de conocimientos técnicos** para poder interpretar los datos fisiológicos. Por ejemplo, debemos resaltar el hecho de que las medidas fisiológicas tienen su propia escala temporal: los cambios en amplitud y latencia de los potenciales evocados, por ejemplo, así como las variaciones que acontecen en la tasa cardiaca, ocurren en varios cientos de milisegundos tras producirse un cambio en las estrategias del operador; sin embargo, en el otro extremo, las medidas metabólicas cerebrales toman un lapso de 30 segundos a varios minutos para reflejar variaciones en los niveles de carga mental. Otra de las mayores desventajas tiene que ver con la intrusividad de los métodos de medida fisiológicos. A este respecto, si bien en la actualidad los equipamientos de medición resultan cada vez más simples y menos

intrusivos, no cabe duda de que **la intrusividad con este tipo de medidas es superior**. Además, este tipo de sistemas de registro pueden suscitar comportamientos antinaturales en los participantes, lo cual contaminaría los datos adquiridos. A modo de ejemplo, para la obtención de la actividad eléctrica cerebral “EEG” es requisito indispensable la colocación de un gorro de electrodos, con el consecuente perjuicio para los participantes, quienes verán su comodidad y libertad condicionadas por esta metodología, redundando asimismo en una menor aceptación por parte del operador.

En definitiva, pese a la existencia de ciertas desventajas inherentes a este tipo de medidas, sobre todo de índole metodológico, nosotros consideramos que, como han señalado muchos autores (por ejemplo, Ahlstrom et al., 2006), **las medidas fisiológicas están en un continuo desarrollo tecnológico de tal manera que se convertirán en cruciales para poder evaluar carga y fatiga mental en tiempo real, dada la gran sensibilidad y objetividad que ofrecen** respecto a cambios en el nivel de carga y fatiga mental experimentado por los operarios tal y como se ha venido demostrando en la literatura científica durante las últimas décadas.

### 3.3. Medidas subjetivas

Las medidas subjetivas de carga y fatiga mental **evalúan la percepción subjetiva de carga o fatiga que tiene una persona en un momento determinado o durante un periodo temporal determinado**. En otras palabras, se demanda del individuo que informe verbalmente o por escrito, utilizando alguna escala numérica de intervalo, cuál es la percepción subjetiva de carga o fatiga mental que se encuentra experimentando en ese momento o en un periodo temporal concreto. Su diseño, validación y uso se ha debido a dos factores fundamentales: **1) no requieren grandes requisitos para su implementación y 2) muchos investigadores creen que la percepción subjetiva de carga es un dato importante en sí mismo que, además, no se puede inferir directamente de los registros fisiológicos o de la ejecución en la tarea**. En este sentido, el cómo se siente una persona no tiene por qué reflejarse en su ejecución o en un registro fisiológico concreto, de este modo, el uso de medidas subjetivas nos permite obtener un reflejo de la opinión directa que la persona se forma acerca del esfuerzo mental suscitado por las tareas o del grado de fatiga mental que experimenta. Según Yeh & Wickens

(1988), las medidas subjetivas son una proyección de la introspección de las personas que refleja las demandas impuestas en la Memoria de Trabajo.

Existen muy diversos tipos de medidas subjetivas que difieren en varias características (Tsang & Wilson, 1997) y, por tanto, podemos clasificarlas utilizando diferentes criterios. Uno de los criterios más significativos es la dimensionalidad de la medida, de modo que disponemos de **medidas unidimensionales y multidimensionales**. Atendiendo, por ejemplo, a la evaluación subjetiva de carga mental, **las medidas unidimensionales proporcionan una medida global de la carga mental mientras que las multidimensionales proveen de información diagnóstica acerca de la naturaleza o de los factores que componen el constructo**. Dentro de las medidas subjetivas unidimensionales encontramos la Escala de Cooper & Harper (1969), la Escala de Bedford (Roscoe, 1987; Roscoe & Ellis, 1990) o la Escala de carga global de Vidulich & Tsang (1987) entre otras (para una revisión detallada de estas técnicas consultar Rubio & Díaz, 1999). Sin embargo, debido a que casi todos los autores coinciden en el carácter multidimensional del constructo, en la mayoría de las aplicaciones se usan las medidas multidimensionales más relevantes como son el "SWAT" (Subjective Workload Assessment Technique, por sus siglas en inglés), "NASA-TLX" (NASA Task Load Index, por sus siglas en inglés) y el Perfil de Carga de Trabajo.

Sin embargo, desde el punto de vista de la medición de carga mental o fatiga en los ambientes aplicados, el criterio de clasificación más importante es el momento de aplicación del método de evaluación. De esta manera, tal y como hemos señalado en líneas anteriores de este apartado, **podemos diferenciar entre métodos "off-line" (al finalizar la tarea) y los métodos "on-line" (durante la ejecución de la tarea)**. Los inconvenientes principales de los métodos off-line aplicados concretamente a las medidas subjetivas, son dos principalmente. En primer lugar, los **métodos off-line dependen de la memoria**, esto es, el operario debe recordar cuál era su nivel de carga o fatiga durante la realización de la tarea. En los métodos off-line el momento exacto de hacer la estimación puede variar desde minutos después de terminar la tarea hasta varias horas o incluso días. El intervalo de tiempo entre el final de la tarea y la emisión de la respuesta de la carga o fatiga percibida hará que la introspección dependa en mayor o menor medida de la memoria episódica, de este modo, cuanto mayor sea el tiempo transcurrido, más dependiente será la introspección de la memoria y, por tanto, menos fiable será la

estimación de la carga o la fatiga mental. En segundo lugar, los métodos off-line **no nos informan sobre las variaciones en los niveles de carga o fatiga que experimenta el individuo durante la ejecución de la tarea**. Cabe explicitar que obtener información acerca de dichas variaciones nos arroja luz acerca de aspectos específicos de la tarea que son causas de sobrecarga, infracarga o incluso de fatiga mental. Por todo lo anteriormente explicitado, **existe cierta preferencia por el uso de los métodos on-line**, sobretodo en ambientes aplicados, **cuya principal limitación es su cierto grado de intrusividad**.

### 3.3.1. Medidas de aplicación off-line

Una de las medidas multidimensionales off-line más destacada de carga mental es la denominada "**SWAT**" que fue elaborada por Reid & Nygren (1988), autores que asumían el constructo de carga mental como multidimensional. Tras una amplia revisión de la literatura, estos autores defendieron la existencia de **tres variables definitorias del constructo, las cuales terminarían por conformar su instrumento: tiempo, esfuerzo mental y estrés** (ver Tabla 1). Comenzando por el tiempo, éste es considerado casi por todos los autores como el componente más relevante (Kahneman, 1973). La dimensión de esfuerzo mental sería el segundo elemento más importante y estaría relacionado con la complejidad y la dificultad de las tareas implicadas en procesos como cálculo, toma de decisiones, atención a fuentes de información, recuperación, mantenimiento de información en la memoria, etc. La tercera dimensión es el estrés, que parece relacionado con variables como la motivación, el entrenamiento, la fatiga, la salud o el estado emocional y que es definida por los autores como "algo que contribuye a la confusión, frustración y ansiedad del operador". Cada una de dichas dimensiones se estima en un rango de 1 a 3 valores, correspondientes a las descripciones verbales de la Tabla 1.

**Tabla 1.** Dimensiones del SWAT, desarrollado por Reid & Nygren, (1988).

<b>Tiempo</b>
1. Normalmente sobra tiempo. Las interrupciones o solapamientos entre las actividades son infrecuentes o nunca ocurren.
2. Ocasionalmente sobra tiempo. Las interrupciones o solapamientos entre las actividades son frecuentes.

<p>3. Nunca o casi nunca sobra tiempo. Interrupciones o solapamientos entre las actividades son muy frecuentes o se producen siempre.</p>
<p><b>Esfuerzo Mental</b></p>
<p>1. Se requiere muy poco esfuerzo o concentración mental consciente. La actividad es casi automática y requiere muy poca o ninguna atención.</p> <p>2. Se requiere un nivel moderado de esfuerzo o concentración mental consciente. La complejidad de la actividad es moderadamente alta debido a incertidumbre, imprevisión o falta de familiaridad. Se requiere un nivel de atención considerable.</p> <p>3. Se requiere un nivel alto de esfuerzo y concentración. La actividad es muy compleja y requiere toda la atención.</p>
<p><b>Estrés</b></p>
<p>1. Niveles muy bajos de confusión, riesgo, frustración o ansiedad que pueden tolerarse con facilidad.</p> <p>2. Niveles moderados de estrés debido a confusión., frustración, ansiedad. Para mantener el nivel adecuado de rendimiento es necesario hacer un esfuerzo significativo.</p> <p>3. Niveles muy intensos de estrés debido a confusión, frustración y ansiedad. Se requiere un grado de autocontrol extremo.</p>

La técnica se aplica siguiendo un procedimiento de dos fases. Durante la primera fase, se realiza una evaluación del nivel de carga mental que le genera al individuo los 27 niveles de la escala, los cuales resultan de combinar entre sí los 3 niveles de cada una de las dimensiones: tiempo, esfuerzo y estrés. Estos 27 niveles deben ser ordenados por el sujeto de mayor a menor en función del nivel de carga que le suscita cada nivel. Posteriormente, se aplica un análisis de medida conjunta, obteniendo una escala de intervalo que asigna una puntuación de 0 a 100 a cada una de las 27 combinaciones. En las investigaciones en las que se compara la carga mental de varios grupos de personas, la técnica permite agrupar a los sujetos en función de la dimensión a la que otorgaron más importancia en la ordenación y, de esta forma, se obtiene una escala de carga mental diferente para cada grupo. En la segunda fase, una vez realizada la tarea o tareas, el operario evalúa cada tarea en las tres subescalas, otorgando una puntuación de 1 a 3. Más tarde, dichas puntuaciones se transforman en una puntuación de carga global, mediante el uso de la escala obtenida en la fase precedente. Así se pretende disminuir la variabilidad entre sujetos en la estimación de la importancia de cada dimensión de carga mental.

Hart & Staveland (1988) propusieron una escala de evaluación multidimensional, conocida como "NASA-TLX". Esta escala está **compuesta por 6 factores y se combina información acerca de la importancia subjetiva de cada factor y su magnitud en una tarea determinada**. Para su elaboración, Hart & Staveland presentaron un total de 19 factores a un grupo de personas dedicadas a un amplio rango de ocupaciones, quienes debían determinar cuál de dichos factores estaban o no relacionados con la carga mental. Como resultado se descartaron 5 factores, restando un total de 14 factores relacionados con el constructo. A continuación, se procedió a evaluar la experiencia de otro grupo de sujetos con esos 14 factores en una variedad de tareas de laboratorio y tareas de simulación de vuelo, reduciéndose el número de factores relevantes a 10: carga mental global, dificultad de la tarea, presión de tiempo, ejecución propia, esfuerzo físico, esfuerzo mental, frustración, estrés, fatiga y tipo de actividad. En última instancia, estos factores fueron evaluados en una serie de diferentes tareas experimentales (tarea simple, tareas duales y simulación, entre otras) y sus resultados fueron utilizados para escoger los 6 factores que finalmente componen a la escala, estos son: demanda mental, demanda física, demandas temporales, ejecución, esfuerzo y nivel de frustración (ver Tabla 2).

La aplicación de la técnica NASA-TLX consta, al igual que en el caso de SWAT, de dos fases, siendo la primera de estimación del grado de relevancia de cada dimensión para cada individuo y la segunda de evaluación de la magnitud de cada escala. Durante la primera, se les presenta a los participantes un conjunto de "pares de factores" (15 en total, resultantes de realizar comparaciones binarias de las seis dimensiones) y se les demanda elegir de cada uno de los pares de factores aquel que ellos estiman más determinante como fuente de carga mental en dicha tarea en particular. De este modo, en base a la cantidad de ocasiones en las que el participante selecciona cada dimensión en las comparaciones binarias, se asigna un peso específico (0-5) a cada factor. Durante la segunda fase, que se realiza tras la realización de la tarea, el participante debe otorgar un valor de magnitud (0-100) de carga mental para cada uno de los 6 factores de forma individual. Posteriormente, los resultados de ambas fases se combinan para obtener una medida global de carga mental, además de una medida relativa del grado de carga mental suscitado por cada factor. De este modo, la combinación de los pesos provee un indicador sensible de la carga global entre diferentes tareas o diferentes niveles de una tarea, en

tanto que, los pesos y las magnitudes de las evaluaciones de cada factor proveen importante información diagnóstica acerca de la fuente específica de carga mental.

**Tabla 2.** Factores del NASA-TLX, desarrollado por Hart & Staveland, (1988).

Factor	Descripción
Demanda Mental	Cantidad de Actividad Mental y Perceptiva que requiere la tarea (p.e. pensar, decidir, calcular, buscar, etc.).
Demanda Física	Cantidad de actividad física que requiere la tarea (p.e. pulsar, empujar, girar, deslizar, etc.).
Demanda Temporal	Nivel de presión temporal sentida. Razón entre el tiempo requerido y el disponible.
Rendimiento	Hasta qué punto el individuo se siente satisfecho con su nivel de rendimiento.
Esfuerzo	Grado de esfuerzo mental y físico que tiene que realizar el sujeto para obtener su nivel de rendimiento.
Nivel de Frustración	Grado de inseguridad, estrés, irritación, descontento, etc., sentido durante la realización de la tarea

Finalmente, la técnica conocida como **“Perfil de Carga de Trabajo” (Workload Profile)** fue desarrollada por Tsang & Velázquez (1996) basándose en el modelo de Recursos Múltiples de Wickens (1984). Las dimensiones que evalúa son, por tanto, las hipotetizadas en su modelo: **procesamiento perceptivo/central, selección de respuesta y ejecución, procesamiento espacial, procesamiento verbal, procesamiento visual, procesamiento auditivo, output manual y output verbal** (Ver Tabla 3).

El procedimiento de aplicación es sencillo y consta de una sola fase: se pide a los sujetos que estimen la proporción de recursos atencionales empleados de cada tipo (0-1), después de realizar un conjunto de tareas. La simple suma de las dimensiones de carga mental no parece ser la manera óptima de combinar las evaluaciones multidimensionales, por lo que, aún no se tiene un método claro para combinar las dimensiones y obtener un índice global de carga mental.

**Tabla 3.** Dimensiones del Perfil de Carga de Trabajo, desarrollado por Tsang & Velázquez, 1996.

<b>Estructuras de Procesamiento</b>
Procesamiento de Respuesta: son recursos atencionales requeridos para seleccionar y llevar a cabo respuestas.
<b>Códigos de procesamiento</b>
(1) Procesamiento perceptivo/central: son recursos atencionales requeridos para actividades como percibir (detectar, reconocer e identificar objetos), recordar, solucionar problemas y tomar decisiones.
(2) Procesamiento espacial: algunas tareas son espaciales en naturaleza. Conducir, por ejemplo, requiere prestar atención a la posición del coche, la distancia entre la posición y la distancia de la señal de stop, etc.
(3) Procesamiento verbal: otras tareas son verbales en naturaleza. Por ejemplo, leer requiere procesamiento de material verbal y lingüístico.
<b>Modalidad de input</b>
Procesamiento Visual. Algunas tareas se realizan basándose en la información visual recibida. Por ejemplo, ver la televisión es una tarea que consume recursos visuales

La validación de este conjunto de técnicas multidimensionales ha revelado los siguientes resultados (Nygren, 1991; Rubio & Díaz, 1999; Tsang & Velázquez, 1996):

1. La **técnica del Perfil de Carga de Trabajo proporciona más información diagnóstica acerca de la naturaleza de las tareas que otras técnicas multidimensionales** como el SWAT o NASA-TLX y además está basada en un modelo teórico de carga mental que permite hacer predicciones a priori.
2. Las medidas subjetivas **son económicas, fáciles de aplicar, el grado de intrusión es mínimo y son ampliamente aceptadas por los operadores.**
3. El **NASA-TLX tiene utilidad preferentemente como modelo de predicción general de la carga mental experimentada.** Sin embargo, su procedimiento para pesar las dimensiones es ineficaz y podría ser ignorado.



4. El **SWAT es muy sensible para captar las diferencias individuales en carga mental** gracias al uso de la metodología de medida conjunta.
5. Las dimensiones del Perfil de Carga de Trabajo son diferentes de las dimensiones del NASA-TLX y del SWAT, sin embargo, las tres han resultado ser válidas para evaluar la carga mental, por tanto, **cada una debe estar dando cuenta de un aspecto diferente del constructo multidimensional.**

Hay que tener en cuenta que esta validación de las medidas subjetivas se ha llevado a cabo por comparación (correlación) con las medidas de ejecución y las medidas fisiológicas. Es decir, cuando se considera, por ejemplo, que la técnica SWAT es sensible a las diferencias individuales, se afirma que es capaz de detectar dichas diferencias en aquellos casos en los que las otras medidas criterio también las detectan. En cualquier caso, Yeh & Wickens (1988) han mostrado que la relación entre ejecución y carga mental obtenida mediante técnicas subjetivas no es perfecta. De otro lado, la percepción subjetiva de carga mental aumenta cuando se produce una elevada demanda de recursos de la Memoria de Trabajo, mientras que, cuando la demanda es baja, las técnicas subjetivas no son capaces de detectar adecuadamente los cambios en carga mental.

En lo que respecta a las **medidas off-line de fatiga, tradicionalmente la fatiga ha sido evaluada mediante escalas unidimensionales**, a través de una simple pregunta acerca de cómo de cansado se siente el individuo (Goldmark, 1912; Folkard, 1985; Rimehaug & Svebak, 1987; Lee et al., 1991; Rabinach, 1992; Monk & Öberg et al., 1994; Okogbaa et al., 1994; Schaeffer et al., 1995). Este tipo de instrumentos unidimensionales permiten a los investigadores obtener una estimación de la intensidad de la fatiga percibida, sin embargo, no ofrecen información sobre su cualidad. Tal y como apunta Ahsberg (1998), incluso aunque se puede experimentar el mismo nivel de fatiga tras realizar tareas de distinta índole, la percepción de fatiga podría ser de diferente naturaleza en función del tipo de tarea desempeñada. A este respecto, **un acercamiento multidimensional permite precisamente evaluar la intensidad y la cualidad de la fatiga percibida** (Kinsman & Weiser, 1976). Las características de la fatiga cambian dependiendo del contexto específico, por lo que varios estudios se han desarrollado precisamente con el objetivo de identificar las dimensiones involucradas en la percepción de la fatiga en contexto laboral.

Los aspectos cualitativos de la fatiga más frecuentemente descritos son aquellos definidos como “Fatiga física” y “Fatiga mental” (Chalder et al., 1993; Grandjean, 1989), en tanto que otros aspectos relacionados con el constructo que han sido investigados son la “Somnolencia” (Gillberg et al., 1994) y el “Malestar” (Cameron, 1996). Un estudio realizado con trabajadores ferroviarios identificó dos dimensiones de la fatiga: “Activación debilitada” y “Motivación debilitada” (Kashiwagi, 1969). De otro lado, Matthews & Desmond (1998) identificaron cuatro dimensiones básicas de la fatiga como resultado de la exposición prolongada a una tarea de simulación de conducción: “Aburrimiento”, “Fatiga visual”, “Incomodidad” y “Fatiga muscular”. Mientras que Saito & Kashiwagi (1970) identificaron, por su parte, las siguientes dimensiones de un conjunto de 30 síntomas de fatiga en una muestra de 9575 trabajadores industriales: “Somnolencia y embotamiento”, “Síntomas mentales” y “Proyección de degradación física”. Posteriormente, dichas dimensiones fueron aplicadas en un estudio con 17.789 participantes en el que se halló que la primera dimensión (“Somnolencia y embotamiento”) era común en un gran número de profesiones, mientras que la segunda (“Síntomas mentales”) y la tercera (“Proyección de degradación física”) eran comunes en actividades de carácter mental y físicas, respectivamente (Yoshitake, 1978).

Uno de los instrumentos destacados para evaluar fatiga es el “**Multidimensional Fatigue Inventory**” (MFI) (Smets et al., 1995; Smets et al., 1996), originalmente desarrollado para la evaluación de fatiga en pacientes con cáncer pero que, sin embargo, posteriormente ha sido empleado con éxito en el ámbito laboral (Mansour, Riad, & Moussa, 2010). Otro instrumento destacado y ampliamente utilizado es el “**Swedish Occupational Fatigue Inventory**” (SOFI) desarrollado por Ahsberg et al., (1997), el cual propone 5 dimensiones obtenidas a partir de 95 expresiones verbales que describen las sensaciones de la fatiga y que fueron evaluadas por 705 trabajadores de 14 profesiones diferentes. En la tabla 4 se pueden observar dichas dimensiones, las cuales se denominan: “Falta de energía”, “Esfuerzo físico”, “Malestar físico”, “Falta de motivación” y “Somnolencia”.

**Tabla 4.** Descripción de las dimensiones de fatiga propuestas por el “Swedish Occupational Fatigue Inventory (SOFI; Ahsberg et al., 1997).

Dimensión	Descripción
Falta de energía	Describe un sentimiento general de disminución de fortaleza

Esfuerzo físico	Describe sensaciones percibidas por todo el cuerpo que pueden ser el resultado del trabajo dinámico y, hasta cierto grado, los síntomas del agotamiento metabólico
Malestar físico	Describe sensaciones corporales más localizadas como resultado de carga isométrica o estática.
Falta de motivación	Describe sentimientos relacionados con la falta de involucración y entusiasmo en la tarea.
Somnolencia	Describe sensaciones de cansancio/somnolencia

De acuerdo a los autores, la estructura subyacente del instrumento corresponde a una nueva descripción cualitativa y cuantitativa de las dimensiones físicas (Esfuerzo físico y Malestar físico) y mentales (Falta de motivación y Somnolencia) de la fatiga percibida. Junto a estos, el factor “Falta de energía” corresponde a una dimensión de fatiga que comparte características tanto físicas como mentales. Originalmente, el inventario SOFI estaba compuesto de 25 expresiones (5 por cada dimensión) relacionadas con respuestas fisiológicas, cognitivas, motoras y emocionales. El proceso de evaluación consiste en pedir a los participantes que evalúen en una escala de 11 puntos hasta qué grado dichas expresiones pueden ser aplicadas para describir sus sensaciones en dicho momento.

### 3.3.2. Medidas de aplicación on-line

En muchas ocasiones el interés de los investigadores no consiste en obtener el grado de carga mental y/o fatiga general experimentado por una persona durante la totalidad de la tarea, sino que el interés puede residir en conocer el modo en que la carga mental va evolucionando a lo largo de la tarea como consecuencia de eventos que acontecen durante su desarrollo o, quizás, en determinar el momento de surgimiento de fatiga mental y/o su evolución con el transcurso del tiempo. Por esta razón, se han desarrollado **las medidas denominadas “on-line” que nos permiten obtener una estimación continua de carga y/o fatiga mental, pero que, sin embargo, tienen el gran inconveniente de ser muy intrusivas**, habida cuenta de que requieren la demanda constante de la estimación de carga y/o fatiga, de manera que el individuo debe retirar periódicamente su atención de la tarea, aunque sea durante breves periodos de tiempo.

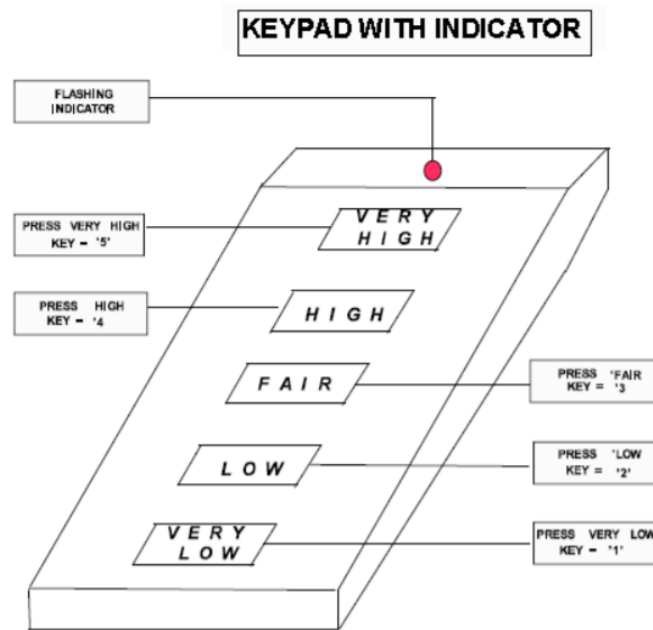
Si bien, tal y como queda explicitado en líneas anteriores, las medidas off-line pueden dividirse en medidas unidimensionales y multidimensionales, **las medidas on-line deben ser de naturaleza unidimensional**. El sentido de esto es el siguiente, las medidas off-line se aplican al finalizar la/s tarea/s de interés, por lo que obtener este tipo de medida no perjudicará la ejecución del operario, de este modo, las medidas multidimensionales pueden utilizarse sin ningún inconveniente pese a ser más intrusivas aportando, además, la ventaja de ofrecer un análisis más detallado sobre las fuentes de carga y/o fatiga mental. Sin embargo, cuando el objetivo de nuestra investigación nos obliga a tener que adquirir una medida de carga y/o fatiga mental en tiempo real, debemos acogernos a las medidas unidimensionales, dada la menor intrusividad relativa a la obtención de una única medida global de carga y/o fatiga mental.

**La escala subjetiva de carga on-line más utilizada por defecto es la denominada “Escala ISA” (The Instantaneous Self-Assessment Scale)** (Brennan, 1992). Dicha escala fue desarrollada en el ATMDC (Air Traffic Management Development Centre, National Air Traffic Services) para evaluar los niveles de carga mental de los ACTo en el ámbito ATM y constituye una de las más frecuentemente utilizadas en las simulaciones en tiempo real en dicho ámbito (Kirwan et al., 1997); sin embargo, se trata de una escala muy sencilla que puede utilizarse con gran sencillez en cualquier ámbito.

**La escala ISA consiste en realizar una evaluación continua (normalmente intervalos de 2 a 5 minutos) al operario de un sistema, o a un individuo que ejecuta una tarea determinada, del nivel de carga mental que ha experimentado desde el último intervalo evaluado.** La escala original tiene un **rango de puntuaciones posibles de tipo Likert de 1 (mínimo nivel de carga mental) a 5 (máximo nivel de carga mental)** (ver Tabla 5), sin embargo, se trata de una escala muy flexible que puede ser modificada según los intereses del investigador (por ejemplo, aumentando el rango a 1-7 o 1-9) y puede utilizarse en cualquier lugar, momento y ámbito. Dicha escala puede ser aplicada mediante lápiz y papel y con una alarma que suene cada cierto intervalo temporal y que sirve al operario de señal para estimar su nivel de carga del último intervalo. Asimismo, existen aplicaciones disponibles de la escala para Android e iOS e incluso un teclado desarrollado exclusivamente para emitir la respuesta denominado “Sistema ERGO” y utilizado en el ámbito ATM Figura 12).

**Tabla 5.** Rango de respuesta de la Escala ISA con su definición de nivel de carga, capacidad de respuesta y descripción.

Nivel	Carga mental	Recursos disponibles	Descripción
1	Infracarga	Muy numerosos	Poco o nada que hacer. Bastante aburrido
2	Relajada	Amplios	Más tiempo del necesario para completar las tareas. El tiempo pasa lento.
3	Cómoda	Algunos	El controlador tiene suficiente trabajo para mantenerse estimulado. Todas las tareas se encuentran bajo control.
4	Alta	Muy pocos	Algunas tareas no esenciales son pospuestas. No se puede trabajar manteniendo este nivel por mucho tiempo. El controlador trabaja en el límite. El tiempo pasa rápido.
5	Sobrecarga	Inexistentes	Algunas tareas no se resuelven. El controlador se siente sobrecargado y siente que no tiene el control de la situación.



**Figura 12.** Representación gráfica del “Sistema ERGO” para aplicar la Escala ISA en el ámbito ATM.

Las **ventajas** de dicha escala son numerosas:

- La escala proporciona un **método simple, económico y rápido** de analizar los niveles de carga mental subjetiva a lo largo de la ejecución de una/s tarea/s determinada/s.

- **Aprender a utilizarla requiere muy poco entrenamiento.**
- Dependiendo del intervalo de recogida, **permite una evaluación quasi-continua** de los niveles subjetivos de carga mental.
- Los **resultados son inmediatamente utilizables** con mínimo análisis requerido.
- El **método es genérico**, pudiendo ser aplicado en cualquier dominio.
- **Requiere muy pocos recursos.**
- Pese a ser un método intrusivo para la tarea primaria, **se trata del método menos intrusivo de todos los métodos de evaluación subjetiva de carácter on-line.**

Entre las **desventajas** del método, encontramos las siguientes:

- **Puede interferir con el rendimiento y puede significar una fuente adicional de carga mental**, especialmente en configuraciones de elevada carga.
- El carácter subjetivo del método lo hace **susceptible de sufrir fenómenos psicológicos** como la influencia de la deseabilidad social (responder aquello que el operario piensa que socialmente se desea que se responda) o la tendencia a repetir siempre los mismos valores (por ejemplo, valores intermedios).
- Obtener las puntuaciones **requiere de la cooperación de la persona evaluada**, quien puede dejar de prestar atención a dicha escala si los niveles de carga mental de la tarea primaria son muy elevados.

En última instancia, **al objeto de evaluar fatiga mental subjetiva on-line, se puede utilizar una versión modificada de la escala ISA**, respetando las características de la misma, pero modificando el objeto de medición. Esto es, se trata de la misma escala de

medida, pero ahora, el constructo evaluado pasa a ser el nivel de fatiga mental experimentado por el participante.

Cabe hacer mención a la **importancia que tiene la etapa de entrenamiento** con dicha escala, no tanto con el uso de la misma que resulta sencillo, sino con la comprensión del constructo evaluado. A este respecto, la persona que está siendo evaluada debe comprender el significado del constructo evaluado, bien sea carga o fatiga mental y debe aprender a utilizar, siempre que sea posible, un rango lo más amplio posible de la escala (mayor sensibilidad). En otras palabras, debe aprender a identificar qué percepción subjetiva corresponde con cada nivel de la escala, a fin de poder obtener resultados sensibles, válidos y fiables. Para ello, resulta fundamental el proceso de formación que el evaluador debe ofrecer al evaluado a fin de tratar de reflejar lo más fielmente posible las percepciones subjetivas del individuo respecto a sus niveles de carga y/o fatiga mental.

#### 4. Convergencias y divergencias entre medidas

Una vez explicitadas, en el capítulo anterior, las principales medidas utilizadas para evaluar carga y fatiga mental por los investigadores y profesionales del ámbito de la Ergonomía y Ciencias del Trabajo, pasamos en este capítulo a explicar el principal problema que puede surgir cuando pretendemos realizar una evaluación de dichos constructos y que es el objeto principal motivador de la presente tesis doctoral: **las divergencias que se producen entre las diferentes metodologías de evaluación de carga y fatiga mental.**

En primera instancia, consideremos el ejemplo evidente de las **medidas físicas**, como, por ejemplo, las medidas de peso, distancia, volumen, etc. En este primer caso, podemos afirmar que existe una equivalencia exacta entre las diferentes medidas de magnitud. Por ejemplo, podemos decir que Sebastián pesa 74 kilogramos, mide 1.82 metros y ha caminado 8 kilómetros; pero también podemos afirmar que nuestro amigo Sebastián pesa 163,14 libras, mide 8.52 pulgadas y ha caminado 4,97 millas. Pese a utilizar un sistema de medición diferente, podemos afirmar que existe una equivalencia directa y exacta entre diferentes medidas de magnitud, por lo que **encontramos una convergencia perfecta entre medidas.** Sebastián podrá ganar peso, crecer, o caminar mayor distancia y esto se

verá reflejado en los diferentes sistemas de medida, pero siempre de forma equivalente y exacta.

Sin embargo, **cuando lo que pretendemos cuantificar son constructos psicológicos, como es el caso de la carga o la fatiga mental, los problemas de evaluación se multiplican.** La psicometría es la disciplina encargada del estudio de la evaluación (medición y cuantificación) de los procesos psicológicos y las capacidades cognitivas. **El principal problema al que se enfrenta la psicometría es que su objeto de evaluación está constituido por constructos mentales y, por ende, no son físicamente mensurables.** No se puede medir directamente la inteligencia o el neuroticismo y, por tanto, las leyes matemáticas no aplican (por ejemplo, no podemos decir que Sebastián es el doble o la mitad de inteligente que Gustavo). A este respecto, **si bien los constructos mentales no pueden ser evaluados directamente, lo que sí se puede evaluar son los reflejos derivados de los cambios en la magnitud del constructo.** Por ejemplo, sabemos que de forma general, una persona más inteligente debería ser capaz de resolver un mayor número de problemas en menor cantidad de tiempo que otra menos inteligente, de este modo, la cantidad de problemas resueltos y el tiempo de resolución constituirían, en este ejemplo, 2 reflejos indirectos (medidas de rendimiento) del nivel de inteligencia de cierta persona y es, precisamente en esto, en lo que se basan los test de inteligencia para ofrecer una puntuación de coeficiente intelectual (CI) determinada. Uno de los problemas fundamentales a los que se enfrenta la psicometría es que no siempre las diferentes medidas del mismo constructo coinciden entre sí al reflejar las variaciones en los niveles del mismo.

Para continuar entrando en más detalle, pasemos ahora a considerar los constructos que nos ocupan en esta tesis, estos son, la carga y la fatiga mental. Lejos de lo que a priori cabría pensar, **los diferentes índices, tanto de carga como de fatiga mental, no necesariamente coinciden entre sí al reflejar los cambios en la magnitud del constructo.** Como hemos visto en el apartado anterior, existen tres categorías de medidas fundamentales para cuantificar la carga y la fatiga, estas son, las medidas de rendimiento, las psicofisiológicas y las subjetivas. Tomemos como ejemplo el constructo de carga mental. Cabría esperar que, a medida que aumenta la demanda cognitiva de una tarea (lo que supondría un incremento en carga mental), la persona que estamos evaluando mostraría:



1. **Una disminución en su rendimiento:** el individuo iría haciendo reajustes constantes entre su actividad y el nivel de carga cognitiva para ir adaptándose a la situación, sin embargo, a partir de cierto grado de demanda cognitiva, su nivel de carga mental aumentará de tal manera que le resultaría difícil mantener una ejecución óptima en la tarea, produciéndose así una disminución de su rendimiento.
2. **Un incremento en los niveles de activación del sistema nervioso central y mejora en las estrategias del procesamiento de la información:** debido al incremento progresivo en los niveles de demanda cognitiva y fruto de los reajustes internos que acontecen en el organismo para hacer frente a la nueva situación, habría de producirse un incremento en los niveles de activación del sistema nervioso central, lo cual se vería reflejado en diversos índices psicofisiológicos. Por ejemplo, el nivel de sudoración aumentaría, aumentando la conductividad de la piel, aumentaría la frecuencia cardiorrespiratoria, la pupila se dilataría y los potenciales eléctricos cerebrales de ciertas áreas del cerebro cambiarían. Del mismo modo, se producirían otro tipo de cambios psicofisiológicos que no guardan relación directa con los niveles de activación del sistema nervioso central, pero sí constituyen estrategias inconscientes para mejorar el procesamiento de la información. Por ejemplo, la tasa y duración de parpadeos disminuirá, a fin de poder incorporar la mayor cantidad de información posible (INPUT) a los procesos cognitivos, y la mirada seguirá un patrón determinado que favorezca, de nuevo, el procesamiento de la información. Por ejemplo, aumentará la velocidad de los movimientos sacádicos o se reducirá la amplitud de los movimientos oculares, para focalizar la atención justo donde resulta necesario. En otras palabras, el organismo del individuo se reajustaría de forma óptima para poder hacer frente a la nueva situación más demandante, y esto se vería reflejado en un conjunto de cambios psicofisiológicos.
3. **Un incremento en los niveles subjetivos de carga mental:** el cerebro monitoriza constantemente y de forma inconsciente el estado actual de carga (y fatiga) mental al que está sometido el organismo. De este modo, al

producirse un incremento en el nivel de demanda cognitiva el individuo podrá, mediante un proceso de introspección, determinar de forma subjetiva su nivel de carga mental actual y podrá evaluarlo, reportando, en cierta escala de medida, que se ha producido un incremento en sus niveles de carga mental como consecuencia del aumento en los niveles de demanda cognitiva de la tarea.

Sin embargo, la investigación científica y la práctica han demostrado que los diferentes índices (de carga y fatiga mental) no evolucionan como cabría esperar en respuesta a variaciones en la magnitud del constructo. La comunidad científica muestra cada vez un **mayor número de estudios que ofrecen evidencia acerca de la ausencia de convergencia entre las diferentes medidas** (Yeh & Wickens, 1984; Yeh & Wickens, 1988; Casper, 1990; Horrey, Lesch & Garabet, 2009; Kamzanova, Kustubayeva & Matthews, 2014; Matthews et al., 2015; Wickens, 2017; Hancock, 2017, Hancock & Matthews, 2019). Se han reportado casos en los que las tres categorías de medidas se muestran sensibles a los cambios en la magnitud del constructo y evolucionan conforme a lo esperado, pero, también, casos en los que alguna o ninguna de las medidas evoluciona conforme a lo esperado, o incluso de forma contraria a lo que cabría esperar.

La existencia de divergencia entre las diferentes medidas constituye uno de los mayores problemas a los que debe enfrentarse actualmente la Ergonomía y las Ciencias del Trabajo. Al objeto de poder continuar con el desarrollo de una sociedad moderna, **resulta necesario poder evaluar la carga y la fatiga mental de una forma adecuada** tanto en investigación, **para continuar aumentando nuestro nivel de conocimientos** en el campo, como sobretodo en la práctica, **para evitar accidentes laborales y aumentar la productividad y el bienestar de los trabajadores**; sin embargo, para poder evaluar la carga y la fatiga, **es necesario disponer de metodologías fiables y válidas.**

#### 4.1 Asociaciones, disociaciones e insensibilidades

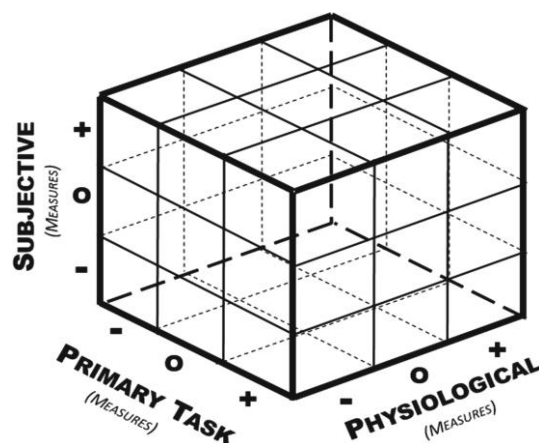
Si los constructos a los que llamamos carga mental y fatiga mental están bien definidos y establecidos en nuestra ciencia, esperaríamos que todas las medidas que han mostrado ser sensibles para reflejar dichos constructos converjan, hasta una proporción determinada, con los cambios en la magnitud de los mismos. Bien es cierto que la

convergencia entre medidas nunca podrá llegar a ser perfecta ni exacta, dadas las dificultades inherentes a la evaluación de constructos mentales, que no son directamente observables; sin embargo, se esperaría encontrar un elevado grado de convergencia entre las diferentes categorías de medidas. **La ausencia de convergencia entre diferentes medidas que deberían ser capaces de reflejar de forma unitaria los cambios en la magnitud del constructo que reflejan está relacionada, según Hancock (2017), con el surgimiento de disociaciones y/o insensibilidades en las medidas de carga mental, cosa que también acontece en el caso de la fatiga mental.** A continuación, pasaremos a explicar los conceptos de asociación, insensibilidad y disociación. Nos centraremos en el concepto de carga para ejemplificar la explicación.

Imaginemos una situación en la que el grado de dificultad de una tarea aumenta, por ejemplo, imaginemos un piloto de avión comercial que se encuentra pilotando un avión de pasajeros y se debe enfrentar a una tormenta virulenta, al tiempo que evaluamos la carga mental del piloto con las tres categorías de medidas, esto es, mediante la escala ISA (medida subjetiva), contabilizando los errores del piloto (medida de rendimiento) y registrando su tasa cardiaca (medida psicofisiológica). Bajo dicha circunstancia, cabe esperar que el nivel de demanda cognitiva de la tarea aumente y, si el nivel de recursos disponibles del piloto es sobrepasado por las circunstancias, se producirá una sobrecarga mental. En esta situación de sobrecarga habríamos de encontrar un deterioro en el rendimiento del piloto (ej. el piloto podría fallar al informar al controlador de su posición), especialmente si la dificultad de la tarea continúa aumentando con un nuevo evento indeseable, como por ejemplo con la congelación de cierto instrumento del avión. Asimismo, esperamos que su percepción subjetiva de carga siga la misma progresión que el deterioro en el rendimiento, el piloto habría aumentado sustancialmente sus puntuaciones en la escala ISA de carga mental (ej. de 2 a 5). En última instancia, cabría esperar un incremento de su tasa cardiaca, indicativo esto de la existencia de una mayor activación fisiológica. Los ejemplos anteriores constituyen casos de lo que denominamos como “asociaciones”. **Las asociaciones acontecen cuando las medidas de carga mental van en la dirección de lo esperado con respecto a la variación en los niveles de demanda cognitiva de la tarea.** Sin embargo, la literatura científica ha demostrado repetidamente a lo largo de estos años de investigación que no siempre encontramos asociaciones entre medidas, sino que también podemos encontrar “disociaciones” e “insensibilidades”. Continuando con el ejemplo anterior, podría ser que, con el

incremento en carga mental, el piloto cometiera un mayor número de errores pero que, sin embargo, al observar la escala ISA, nos sorprenderíamos al descubrir que sus reportes de carga mental subjetiva permanecen estables. En este último caso estaríamos hablando de la **existencia de una insensibilidad, habida cuenta de que la medida subjetiva no ha variado con el incremento en los niveles de demanda cognitiva de la tarea**. En última instancia, podría ocurrir que el piloto, pese al evidente incremento en la dificultad de la tarea, muestre una disminución de su diámetro pupilar, lo cual **iría completamente en contra de lo esperado**. Esta última circunstancia se denomina **disociación**. Así, en función de la respuesta que encontramos en las diferentes medidas de carga, podemos encontrarnos con una o varias asociaciones, disociaciones e insensibilidades, según las medidas de carga mental cambien conforme a lo esperado, contrariamente a lo esperado o, simplemente, se mantengan estáticas (respectivamente).

Para una mayor comprensión de estos conceptos, en la Figura 13 podemos observar las posibles variaciones de las tres categorías de medida ante cambios en la demanda cognitiva de la tarea. Como resultado del incremento en la demanda cognitiva de la tarea, las medidas de rendimiento, las medidas subjetivas y las psicofisiológicas pueden aumentar (+), permanecer igual (0) o disminuir (-). **Cuanto las respuestas de carga mental se ajustan a los cambios en demanda cognitiva encontramos asociaciones, cuando los reflejos de carga mental no varían se producen insensibilidades y cuando sí varían, pero de forma contraria a lo esperado, se producen disociaciones** (Yeh & Wickens 1988).



**Figura 13.** Posibles respuestas de las tres categorías de medidas (rendimiento, psicofisiológicas y subjetivas) ante variaciones en los niveles de demanda cognitiva de la tarea, Hancock (2017).

Cabe señalar que la investigación desvela que la mayoría de insensibilidades y disociaciones implican a la medida de carga subjetiva (Gilbert, 2009). Asimismo, las asociaciones se reportan en la literatura científica mucho más frecuentemente que las disociaciones o insensibilidades, aunque no se conoce si esto responde a un efecto “file-drawer” o “sesgo en las publicaciones científicas”, según el cual existe una tendencia a publicar únicamente los resultados positivos y coherentes con nuestras expectativas (Hancock, 2017). En cualquier caso, **la existencia de disociaciones e insensibilidades en las medidas resultan bastante frecuentes y son muy problemáticas**, tanto para la interpretación científica de los constructos de carga y fatiga mental, como para su evaluación en los ámbitos aplicados. Por todo lo anterior, **las divergencias entre medidas constituyen uno de los mayores problemas a los que se enfrenta la evaluación de los constructos de carga y fatiga mental, y deben ser investigadas** para arrojar luz tanto sobre las circunstancias en las que toman lugar como sobre los motivos por los que acontecen.

De otro lado, la literatura científica ha desvelado la existencia de divergencias, debido a la aparición de disociaciones o insensibilidades, no solo entre las 3 categorías de medidas, sino incluso entre diferentes tipos de medidas dentro de cada categoría. Por ejemplo, numerosos estudios han desvelado la existencia de divergencias entre diversas medidas psicofisiológicas bien establecida (Kamzanova et al., 2014; Matthews et al., 2015). A modo de ejemplo, el diámetro pupilar y la tasa de parpadeo son medidas psicofisiológicas que han mostrado ser sensibles a los cambios en los niveles de demanda cognitiva de la tarea, sin embargo, a veces ambos parámetros pueden verse afectados por ciertos factores extraños (ej. luminosidad, humedad relativa, etc.) que impactarían negativamente sobre dichos índices, favoreciendo así la aparición de divergencia entre dichas medidas y otras medidas psicofisiológicas que no se ven afectadas por citados factores.

Por lo tanto, después de esta revisión, debemos preguntarnos **por qué y cuándo se produce la ocurrencia de asociaciones e inconsistencias (insensibilidades y disociaciones) entre distintos tipos de medidas, que darán lugar a divergencias entre las mismas**. Esta pregunta está en la agenda actual de los investigadores, puesto que de su respuesta dependerá en gran medida nuestro conocimiento sobre los constructos de

carga y fatiga mental, pero, sobre todo, la forma cómo evaluemos la carga y fatiga mental soportada por un operador en los ámbitos aplicados.

#### 4.2 Posibles causas del surgimiento de divergencias entre medidas

Los investigadores han identificado múltiples posibles causas sobre este fenómeno, las cuales, de forma general, pueden deberse a una **falta de especificidad o a una falta de diagnosticidad** (Novstrup, Goan & Heaton, 2018). La falta de especificidad explicaría la ocurrencia de disociaciones e insensibilidades debido a la existencia de factores ajenos (variables extrañas) a la carga/fatiga mental que afectan a dichas medidas. La falta de diagnosticidad, por su parte, explicaría la ocurrencia de inconsistencias entre medidas en base a la naturaleza multidimensional de dichos constructos, esto es, diferentes medidas podrían estar reflejando diferentes aspectos de dichos constructos. Cabe recordar que los constructos de carga y de fatiga son multidimensionales y la falta de especificidad y diagnosticidad como causas fundamentales de la existencia de divergencias entre medidas sería consistente con la teoría de los recursos múltiples de Wickens (2008).

**Hancock (2017), por su parte, ofrece un conjunto de posibles razones acerca de la ocurrencia de disociaciones e insensibilidades respecto al constructo de carga mental**, las cuales el define coloquialmente como las “AIDS” (Acquired immunodeficiency syndrome, por sus siglas en inglés) de la carga mental.

En primer lugar, una de las posibles causas puede estar relacionada con la **escala de tiempo considerada entre medidas**. Existe una fuerte razón para creer que cada uno de los métodos de medida de carga mental posee su propia e inherente escala temporal y que algunas disociaciones e insensibilidades están supeditadas a dichas diferencias temporales. En otras palabras, la existencia de una escala temporal natural e intrínseca para cada uno de los diferentes métodos de evaluación de la carga mental implica que **algunos índices de carga mental reflejarán los cambios en la magnitud del constructo de forma instantánea, en tanto que, otros, mostrarán una mayor latencia** entre las variaciones en los niveles de demanda cognitiva y su manifestación. Algunas medidas son el resultado de experiencias a lo largo de toda la tarea, misión u operación. Otras ocurren en milisegundos. Por lo tanto, la metodología usada para medir carga mental debe

distinguir estas diferencias existentes en las latencias para asegurar que las disociaciones no sean categóricas, sino debidas a errores en la escala temporal utilizada en los análisis. Imaginemos, por ejemplo, que vamos a evaluar el cambio que acontece en carga mental en un operario cuando se produce un aumento en los niveles de demanda cognitiva de la tarea. Imaginemos que utilizamos 3 medidas de carga mental: una medida de rendimiento (nº de errores en la tarea), una medida psicofisiológica (tasa de parpadeos) y una medida subjetiva (escala ISA). Resulta posible que la medida psicofisiológica considerada sea capaz de reflejar dicho aumento en carga mental (mostrando una menor tasa de parpadeos) varios intervalos antes de que el individuo comience a mostrar un mayor número de errores y un reporte subjetivo más elevado de carga mental. Es decir, **se estarían produciendo divergencias entre medidas como consecuencia de una mayor inmediatez de algunas medidas (tasa de parpadeo) en reflejar los cambios en la magnitud del constructo con respecto a otras medidas que poseen una mayor latencia de respuesta** (nº de errores y escala ISA); sin embargo, si consideramos la escala temporal de cada una de las medidas, con sus propias latencias de respuesta, podríamos caer en la cuenta de que en realidad no se están produciendo divergencias entre las diferentes medidas, sino que en realidad las medidas convergen entre sí, solo que no en tiempo real, sino en distintos momentos temporales.

Otra de las posibles causas de este fenómeno tiene que ver con la **sensibilidad diferencial de las medidas de carga mental a la tasa de cambio en la magnitud de la demanda cognitiva**. Uno de los principios mejor establecidos que existen en las ciencias del comportamiento es que los seres humanos frecuentemente mostramos tener una mayor sensibilidad al cambio que a los niveles absolutos de un estímulo determinado. En palabras de Hancock (2017), tiene sentido pensar que la tasa de cambio en los niveles de demanda cognitiva puede resultar más determinante en la carga mental percibida que cualquier nivel absoluto y estable de demanda cognitiva. En otras palabras, un operario va a tender a reportar, por ejemplo, niveles más elevados de carga cuando se produce un cambio de baja a alta carga (contraste) que si el operario ya se encontraba en una situación estable de alta carga. Sin embargo, podemos ir un paso más allá y pensar que también podría ser posible que exista una sensibilidad diferencial de las diferentes medidas de carga mental a la tasa de cambio en los niveles de demanda cognitiva. Dicho de otro modo, **algunas medidas podrían mostrarse más sensibles que otras a los cambios en los niveles de demanda cognitiva** (transiciones en carga), de tal manera que, cuando el

individuo experimenta cambios abruptos en dichos niveles, las medidas más sensibles al cambio mostrarían una reacción más acentuada que aquellas menos sensibles. **Dichas diferencias entre las distintas medidas podrían dar lugar a la aparición de disociaciones e insensibilidades afectando, en última instancia, a la convergencia entre medidas de carga mental.** Volviendo al ejemplo anterior, en el cual registramos en el operario una medida de rendimiento (nº de errores en la tarea), una medida psicofisiológica (tasa de parpadeos) y una medida subjetiva (escala ISA); imaginemos dos situaciones posibles: una primera situación en la que el aumento en carga se produce de forma progresivo, esto es, de forma lineal (gráfica ascendente lineal) y una segunda situación en la que el incremento en carga se produce de forma abrupta, esto es, por picos de carga (gráfica ascendente en forma de sierra). Cabe preguntarnos si el grado de convergencia será el mismo en ambas situaciones. Puede que, en la primera situación, los niveles de convergencia entre medidas sean mayores que en la segunda, dada la existencia de dicha sensibilidad diferencial de las diferentes medidas de carga a la tasa de cambio en los niveles de demanda cognitiva, cuyo efecto tomaría lugar en la segunda situación de forma particular: puede que el número de errores manifieste adecuadamente los cambios abruptos en carga, así como la tasa de parpadeos y, sin embargo, puede que la escala ISA nos desvele la existencia de una percepción ascendente en los niveles de carga mental, pero no tenga suficiente sensibilidad para desvelar la existencia de dichos cambios abruptos.

En cualquier caso y respecto a lo anteriormente expuesto, cabe destacar que no debe considerarse únicamente el posible efecto diferencial de las tasas de cambio en los niveles de demanda cognitiva de la tarea, sino también las condiciones (los niveles de carga base) en las que acontecen dichos cambios abruptos en carga mental. Un operario puede experimentar un incremento abrupto en carga cuando los niveles de demanda cognitiva son muy reducidos, pero también puede experimentarlo cuando los niveles de demanda cognitiva se acercan al límite de lo asumible. **Las circunstancias en las que acontece dicho incremento abrupto en carga también habrían de afectar al modo en que los índices de carga mental reflejan dichos cambios y, por ende, también a la posible falta de convergencia entre medidas.** Recurriendo nuevamente a nuestro ejemplo anterior, consideremos 2 nuevas circunstancias, en ambas circunstancias se producen incrementos abruptos en los niveles de carga, sin embargo, en la primera circunstancia, el incremento abrupto en carga acontece en niveles de infracarga, mientras que, en la



segunda, dicho incremento toma lugar en una situación de sobrecarga mental. Cabe preguntarnos, ¿existirá el mismo nivel de convergencia entre medidas en ambas circunstancias? o, puede que el nivel base de carga cognitiva afecte más a unas medidas que a otras, lo cual incidiría directamente en la convergencia entre las mismas. En otras palabras y, tomando nuevamente el ejemplo anterior, puede que la medida fisiológica (tasa de parpadeo) sea capaz de reflejar el incremento abrupto en carga, pero únicamente en situaciones de infracarga y no de sobrecarga, mientras que el rendimiento y/o la carga subjetiva podrían reflejar adecuadamente los incrementos abruptos en carga en ambas circunstancias, dando esto lugar, nuevamente a divergencia entre medidas, pero, únicamente, en situaciones de infracarga.

En última instancia, en lo concerniente al concepto de **fatiga mental, una de las razones ampliamente discutida acerca de la existencia de disociaciones entre medidas descansa sobre las teorías del control compensatorio**. Dichas teorías postulan la existencia de un mecanismo de control compensatorio que ejerce su función sobre la regulación del rendimiento humano en condiciones de estrés y elevada carga mental. Más concretamente, **la función de dicho mecanismo será suministrar recursos adicionales, de forma dinámica, en función del nivel de estrés y carga mental experimentado, de manera que el rendimiento en la tarea no se verá mermado en condiciones de elevada demanda de recursos mentales gracias a este suministro de recursos extra que reasigna el mecanismo compensatorio, a expensas de un coste psicofisiológico (agotamiento) y acompañado del surgimiento de la fatiga mental subjetiva (ver Figura 14).**

Hockey (1997) propuso el modelo del control compensatorio, el cual establece que “el mantenimiento de la estabilidad en el rendimiento bajo condiciones de alta demanda cognitiva es un proceso activo que se encuentra bajo el control del individuo y requiere la autogestión de los recursos cognitivos a través de la movilización de esfuerzo mental”. Más concretamente, dicho modelo establece la existencia de dos niveles de control, un bucle de control inferior, que gestiona la actividad reguladora rutinaria, y un bucle superior, que entra en funcionamiento cuando se deben afrontar situaciones más demandantes que amenazan al rendimiento en la tarea. Inicialmente, el control ejecutivo prioriza una meta seleccionada (G) sobre otras metas rivales (g1, g2, etc) (descansar, cambiar de tarea, etc.). En el momento en que se detecta tensión en el bucle inferior,

debido a un aumento en la demanda de la tarea, el bucle superior entra en acción. En esta situación, caben dos posibles opciones: 1) incrementar el esfuerzo (reasignando recursos extra) para impedir el deterioro en el rendimiento, lo cual lleva asociado ciertos costes (aparición de fatiga subjetiva y costes psicofisiológicos) o 2) mantener los niveles de esfuerzo actuales (coste mínimo) pese al riesgo de que se produzca un deterioro en el rendimiento (ya sea manteniendo la meta actual con un rendimiento inferior o reemplazando la meta actual con otra de las metas rivales).

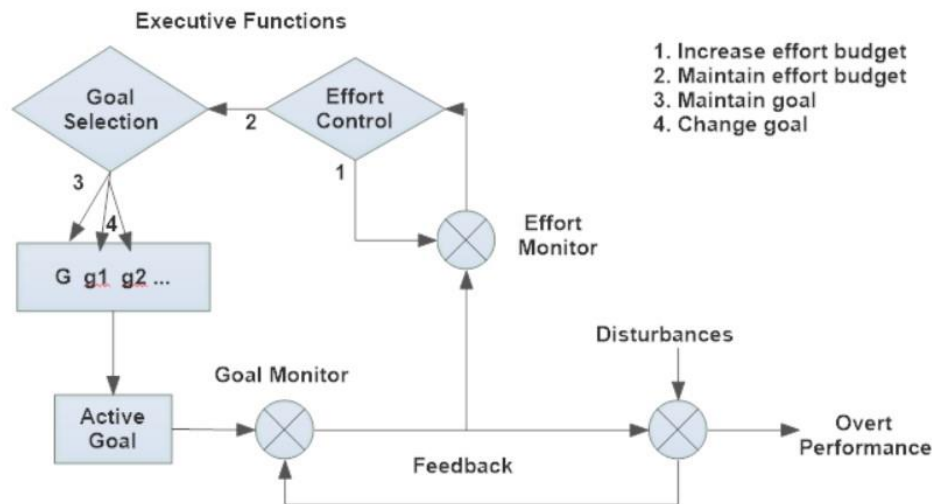


Figura 14. Modelo del Control Compensatorio (Hockey 1997).

Centrándonos en el papel de la fatiga subjetiva en dicho modelo, Hockey (2011) defiende el surgimiento de la fatiga mental subjetiva como el resultado, durante el proceso de monitorización, de detectar la necesidad de realizar un esfuerzo mayor para mantener el rendimiento en niveles adecuados, de tal manera que, **un mayor esfuerzo percibido para hacer frente a una situación más demandante, redundará en una mayor fatiga subjetiva percibida**. Asimismo, no debemos olvidarnos del papel fundamental del tiempo en este sistema, de modo que, **un mayor tiempo dedicado a la tarea demandante, redundará en un decremento progresivo de la cantidad de recursos mentales de que dispone el individuo para hacer frente a la tarea, de modo que, también influirá en el surgimiento de la fatiga subjetiva**. En otras palabras, **la fatiga subjetiva estaría actuando como una señal de alarma del organismo ante el posible agotamiento de los recursos mentales** como consecuencia de la necesidad de reasignar una mayor

cantidad de recursos mentales, especialmente durante un tiempo prolongado, para mantener un rendimiento adecuado en una tarea que ha aumentado en dificultad.

Cabe preguntarnos ahora, de qué manera afecta el control compensatorio en el surgimiento de divergencias entre medidas de fatiga mental. Podríamos pensar que una persona que experimenta fatiga mental mostraría un empeoramiento del rendimiento en la tarea como consecuencia de la mayor dificultad que supone, una menor activación fisiológica como consecuencia del agotamiento de los recursos mentales y una mayor fatiga subjetiva reportada. Sin embargo, si tenemos en cuenta la acción del mecanismo compensatorio el resultado podría ser bien distinto: dicho mecanismo permitiría al individuo mantener un rendimiento adecuado, incluso en condiciones de alta demanda cognitiva, gracias a la reasignación de recursos extra para afrontar dicha tarea. Esto se verá reflejado en un aumento del nivel de activación fisiológica y, todo ese esfuerzo extra, resultará en un incremento de los niveles subjetivos de fatiga mental, que actúa como una señal de alarma en contra del agotamiento de los recursos mentales, de tal modo que, a mayor cantidad de esfuerzo, mayor fatiga subjetiva percibida. A modo de ejemplo, resulta posible que un operario pueda, en un momento determinado de gran dificultad en su trabajo, mantener niveles adecuados de rendimiento y elevados niveles de activación fisiológica y, pese a ello, reportar que se encuentra altamente fatigado mentalmente. Esto es, **cabría encontrar la existencia de divergencias entre medidas de fatiga mental, como consecuencia de la acción de dicho mecanismo compensatorio.**

## 5. Objetivos e hipótesis de la tesis

De lo expuesto hasta ahora en esta tesis, hemos podido apreciar la **acuciante necesidad** existente en el campo de la Ergonomía y en las Ciencias del comportamiento **de poder evaluar adecuadamente los niveles de fatiga y carga mental, tanto en el ambiente de laboratorio** (para incrementar nuestro conocimiento básico sobre dichos constructos) **como, especialmente, en el ambiente aplicado** (para poder monitorizar los niveles de carga y fatiga de los operarios evitando, así, posibles accidentes laborales y mejorando la satisfacción del trabajador y su productividad en la empresa). Dicha necesidad es la principal motivación de la presente tesis doctoral y queda plasmada como el objetivo principal de la misma. Más concretamente, **el objetivo principal de esta tesis**

**doctoral es arrojar luz sobre las causas y circunstancias en las que surgen divergencias entre las diferentes metodologías de evaluación de carga y fatiga mental**, para poder tener una mayor y mejor comprensión sobre dicho fenómeno e incidir, directamente, en el conocimiento que nuestra ciencia tiene sobre éste por un lado y, de otro lado, en un perfeccionamiento de las metodologías existentes de evaluación de fatiga y carga mental de los operarios en el ambiente aplicado. Para ello, las diferentes posibles causas de divergencias entre medidas, expuestas en el apartado anterior, serán testadas mediante una serie de estudios empíricos diseñados explícitamente para dar respuesta a este problema, uno de los mayores problemas a los que se enfrenta nuestra ciencia en la actualidad.

Dicho objetivo principal, se encuentra dividido en una serie de objetivos específicos que se exponen a continuación:

1. **Determinar si ciertas divergencias que acontecen entre medidas de evaluación de carga mental son debidas a la existencia de diferencias (entre medidas) en cuanto a la escala temporal de respuesta, inherente a cada tipo particular de medida.** Algunas medidas podrían tener una menor latencia de respuesta que otras ante cambios en carga y viceversa.
2. **Determinar si ciertas divergencias que acontecen entre medidas de evaluación de carga mental responden a la existencia de una sensibilidad diferencial (entre medidas) a la tasa de cambio en la magnitud de la demanda cognitiva de las tareas (durante las transiciones en carga).** Algunas medidas podrían ser más sensibles que otras al reflejar cambios lineales o abruptos en los niveles de demanda cognitiva de la tarea.
3. **Determinar si ciertas divergencias que acontecen entre medidas de evaluación de carga mental responden a la existencia de una sensibilidad diferencial (entre medidas) al nivel preexistente de carga mental durante las transiciones abruptas en carga.** Algunas medidas, frente a otras, podrían ser capaces de reflejar los cambios abruptos en carga durante ciertas condiciones preexistentes de carga mental.

4. **Determinar si ciertas divergencias que acontecen entre medidas de evaluación de fatiga mental responden a la acción del mecanismo del modelo de control compensatorio de Hockey (1997).**
5. **Explorar nuevas medidas que superen las posibles limitaciones de otro tipo de medidas más intrusivas, particularmente en los ambientes aplicados.**

Los anteriores objetivos específicos serán testados con una serie de estudios empíricos que se exponen en la presente tesis doctoral. Dichos estudios buscarán dar respuesta a las siguientes hipótesis de partida:

1. El primer estudio buscará dar respuesta a dos hipótesis específicas: 1) **existen diferencias en cuanto a la latencia de respuesta de las diferentes metodologías de evaluación de carga mental** y 2) **dichas diferencias en cuanto a la latencia de respuesta podrían explicar parcialmente la aparición de ciertas disociaciones e insensibilidades** que ocurren en las medidas de carga mental, afectando así a la convergencia entre medidas.
2. El segundo estudio empírico incluido da respuesta a la siguiente hipótesis: **se producen más divergencias** entre medidas de carga mental, dado el surgimiento de disociaciones y/o insensibilidades, **cuando la tasa de cambio de las transiciones en carga es de naturaleza abrupta**, en contraposición de cuando dichas transiciones acontecen de forma lineal.
3. El tercer estudio de la presente tesis trata de arrojar luz sobre la siguiente hipótesis: los efectos en las medidas de carga mental serán diferentes en función del nivel de carga mental preexistente cuando se produce el incremento abrupto en carga durante las transiciones en carga. Más concretamente, hipotetizamos que **se producirán mayores divergencias entre las medidas de carga mental cuando el incremento abrupto en carga acontece en situaciones de baja demanda cognitiva**, comparativamente a cuando ocurre en situaciones de alta demanda cognitiva.

4. El cuarto estudio trata de responder a la hipótesis siguiente: **algunas de las disociaciones que se producen entre las diferentes medidas de fatiga mental están relacionadas con la acción del mecanismo de control compensatorio** (Hockey 1997), que supe al organismo con recursos adicionales cuando la demanda de la tarea lo requiere, produciéndose así disociaciones y/o insensibilidades principalmente en los niveles de fatiga subjetiva respecto a los niveles de activación fisiológica y a las medidas de rendimiento.
  
5. En última instancia, el quinto estudio se aleja ligeramente del objetivo principal de la tesis, pero no resulta menos importante dadas las numerosas limitaciones actuales de las metodologías de evaluación de carga y fatiga mental. La hipótesis de partida de este último estudio es la siguiente: **la frecuencia fundamental de la voz (F0), que constituye un parámetro acústico de la misma, aumentará conforme aumenta la fatiga mental**, por lo que podría validarse como un parámetro fiable, válido y pertinente de fatiga mental en los ambientes aplicados, tal y como sugieren algunas investigaciones (Hagmüller Kubin & Rank, 2005; Diepeveen, 2020).

De este modo y, al objeto de cumplir todos los objetivos perseguidos en la presente tesis doctoral y de responder a todas las hipótesis de partida, a continuación, se exponen 5 capítulos que constituyen 5 estudios empíricos que han sido publicados en revistas y libros científicos de elevado impacto.

# Empirical Studies

---

*Estudios Empíricos*





## **Chapter II**

---

*Latency Differences Between Mental Workload  
Measures in Detecting Workload Changes*



## **Latency Differences Between Mental Workload Measures in Detecting Workload Changes**

Enrique Muñoz-de-Escalona<sup>1</sup>

José Juan Cañas<sup>1</sup>

<sup>1</sup> Mind, Brain and Behaviour Research Centre,  
Department of Experimental Psychology, Faculty of Psychology,  
University of Granada

---

This article has been published: Muñoz-de-Escalona, E., & Cañas, J. J. (2018, September). Latency differences between mental workload measures in detecting workload changes. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 131-146). Springer, Cham. [https://doi.org/10.1007/978-3-030-14273-5\\_8](https://doi.org/10.1007/978-3-030-14273-5_8)

## Abstract

Mental workload has traditionally been measured by three different methods corresponding to its primary reflections: performance, subjective and physiological measures. Although we would expect a certain degree of convergence, research has shown that the emergence of dissociations and insensitivities between measures is very frequent. One possible explanation could be related to the differing latencies between each workload assessment method. We tested this explanation by manipulating task complexity through time spent performing a simulated air-traffic control task. In the experimental session, we collected physiological (pupil size), performance and subjective data. Our results showed two periods of bad performance caused by high traffic density and aircraft configurations. Those periods corresponded to higher mental workload as detected by subjective and physiological measures. However, subjective mental workload reacted sooner than physiological mental workload to task demands. These results suggest that the differences in latency could partially explain mental workload dissociations and insensitivities between measures.

**Keywords:** mental workload, latency, workload measures, dissociations, insensitivities.

## 1. Introduction

There is at present no doubt about the importance of measuring mental workload. Modern society needs some sort of cognitive work measure, and it would be very useful to have a valid and reliable standardised method for assessing it; so far, however, this does not appear possible at present. The study of mental workload poses a series of difficulties, since the mechanisms underlying it are not very well known, which makes evaluation of mental workload difficult.

Mental workload is considered a multifaceted construct that cannot be seen directly, but must be inferred from what can be seen or measured [1]. In other words, we do not currently have any objective method to directly measure mental workload as a psychological construct, and we thus have to trust some indirect indicators corresponding to its three primary reflections: 1) performance measures (quantity and quality of performed task); 2) subjective perceptions (questionnaires and scales); and 3) physiological responses (pupil diameter, electrodermal activity, EEG, heart-rate variability, etc.) [2, 3].

We can assume that if the three types of measures reflect the same construct, we would expect a certain degree of convergence between them such that, when a task becomes more complex, the three primary measures would change as follows: 1) lower task performance, 2) subjective perception of higher workload and 3) higher physiological activation responses (taking into account only those physiological variables reflecting physiological activation; other variables such as, e.g. blink rate [4] and heart rate variability [5], have also been shown to decrease during periods of high workload). In other words, we would expect to find associations between task load changes and primary reflections of mental workload, in such a way that we could find positive correlations between measures. However, literature research has shown that this is not always the case, so that the emergence of disassociations, insensitivities and subsequent lack of correlation between measures is more common than we realise [6–8]. An insensitivity occurs when a workload measure does not reflect any change in task load levels, whereas a dissociation takes place when that measure actually changes, but in the opposite direction [6].

Consider, for example, a situation in which a worker is performing a task that becomes more difficult over time. We would expect to find a task performance impairment, higher physiological activation and higher subjective workload perception (associations). We could nevertheless find that the worker is actually showing a stable performance

(insensitivity), but a higher physiological activation (association) and a lower subjective workload perception (dissociation). In other words, in this situation it would not be possible to establish positive correlation between workload measures that are supposed to be measuring the same construct. Such situations are very problematic for our science and practice, and understanding these situations represents one of the greatest challenges to be faced in the mental workload research sphere at present [6]. One might well wonder why this occurs with respect to the occurrence of associations, insensibilities and dissociations between mental workload measures.

One possible explanation of some disassociations and insensitivities that happen between workload measures is related to the timescale considered between measures. We should consider that each different method for measuring mental workload has its own inherent timescale. While some measures could reflect mental workload within seconds, others could show a longer latency between task load changes and mental workload index reflection. We should therefore take into account such timescale latency differences to avoid considering some disassociations and insensitivities as such when, in fact, they are the result of a methodological misconception [6].

This experiment aims to shed some light on the occurrence of disassociations and insensitivities between mental workload measures due to the differences in timescale between primary workload indexes. We hypothesised that these temporal differences between methods timescale are real, and could explain certain disassociations and/or insensitivities that occur between workload measures.

The rest of the paper proceeds as follows. Section 2 outlines related work about general mental workload, focusing on mental workload measurement techniques and measures. Section 3 describes the design and the methodology followed in conducting the experiment. Section 4 presents the obtained results. Section 5 presents a discussion about our findings, as well as limitations and possible new future work. Section 6 concludes the study, summarising the key findings and its impact on the body of knowledge.

## 2. Related Work

Mental workload is a complex construct, which has received multiple definitions over the years, but there is no agreed consensus about its definition [9]. However, what does seem clear is its importance to a modern society, which needs some form of cognitive work assessment. High mental workload can result in physical, psychological and social

problems [10–12], whereas very low mental workload can lead to the risky situation of being “out of the loop” [13]. Stress and fatigue can be the result of performing highly demanding tasks sustained over time, which would ultimately affect both worker health and performance [14–16]. Poor performance can also lead to dangerous and risky situations in fields such as Air Traffic Control [17, 18] and driving [19–21], increasing the likelihood that accidents will occur. Mental workload management thus appears to be a key issue in supporting increased future cognitive demands.

Mental workload is a multi-factorial construct, which depends not only on demanded task resources but also on available resources [22, 23]. For this reason, cognitive models have been developed in order to make it possible to predict mental workload based on a combination of aggregated factors [24–26]. To this effect, computational models have proven appropriate and highly effective in computing the prediction of scientific models in numerous fields [27].

But first, in order to make research possible, validate proposed models and deepen the understanding of mental workload issues, we need reliable methods for assessing it. With this in mind, over the last several decades there has been an exponential increase in the number of studies regarding mental workload measurement techniques and measures [3, 10, 28–30]. There are three main different methodologies for measuring mental workload, which correspond to its three primary reflections:

- 1. Performance measures:** this type of methodology is composed of primary task measures (e.g. n° of errors, reaction time, etc.) and secondary task measures (e.g. choice reaction-time tasks, time estimation, memory-search tasks etc.). The aim is to measure objective performance indexes in order to quantify the quantity and the quality of performed tasks.
- 2. Subjective measures:** this sort of methodology includes all self-reported measures (e.g. NASA Task Load Index, Instantaneous Self-Assessment, Activation Scale, etc.). The aim is to obtain easy and low-cost subjective data about perceived mental workload.
- 3. Physiological measures:** this type of methodology comprises every physiological response to mental workload (e.g. pupil diameter, electrodermal

activity, electroencephalogram, etc.). The aim is to collect objective physiological data, which reacts to changes in mental workload [31].

As stated in the section above, presumably, since the three types of measures reflect the same concept, we would expect them to be interrelated. However, as Hancock [6] reports, literature research has found frequent inconsistencies (dissociations and insensitivities) between measures. Although the occurrence of dissociations and insensitivities may be affected by several factors, this study addresses a gap in the literature about temporal differences between methods' timescale, which may partly explain certain inconsistencies between mental workload measures.

### 3. Design and Methodology

We manipulated two independent variables in this experiment: time on task (TOT) and task complexity, whose effects were tested in an air traffic control (ATC) simulation experiment in which participants were trained and instructed to avoid conflicts between aircraft. This complex and dynamic task would allow us to observe the possible effects of our manipulated independent variables, as well as our measured dependent variables. Participants performed the task for 120 minutes while their performance, pupil diameter and subjective activation were collected.

The hypotheses for the present study are the following:

H1. There exist temporal latency differences between methods.

H2. These temporal latency differences would partly explain certain dissociations and/or insensitivities that occur between mental workload measures.

#### 3.1 Materials and Instruments

**ATC Lab-Advanced Software.** The software used for simulating air traffic control (ATCo) tasks was an air traffic simulator called <sup>ATC</sup>Lab-Advanced, which is available for free public download (see Figure 1) [32].



The <sup>ATC</sup>Lab-Advanced software provided a high level of realism (that is, a high level of similarity to real ATC operational scenarios) as well as simplified and easy handling, which allowed it to be used by all participants in several learning sessions. Additionally, the software allowed strong experimental control of air traffic scenarios parameters because its XML code could be modified to develop scenarios consistent with research needs and objectives.

For scenario development, the static characteristics of the simulation environment (control sector size and possible pathways through which aircraft could travel) were defined first. Next, aircraft quantity (density) and initial aircraft parameters (altitude, assigned altitude, speed, time of appearance on stage and planned route) were defined for each aircraft presented in the scenario.

Once the structural and dynamic scenario parameters were established, a file that could be launched by the simulator was obtained. This file recorded a .log file with performance data for each participant during simulation.

Finally, we note that the <sup>ATC</sup>Lab-Advanced simulator provided participants with all tools needed to carry out the ATCo task, such as the route (the aircraft's fixed route was displayed), distance scale (which allows horizontal aircraft distance measurement) and altitude and speed change tools.

**Scenarios.** Scenarios used in the study varied according to whether the participant was in the training or the experimental stage. During the training sessions, the standard scenarios provided by the software creator software were used, but a specific scenario was programmed by the experimenters to achieve experimental session goals. This specific scenario was programmed with the following features:

- The purpose of the scenario was to subject participants to a variable complexity task situation, that is, a variable mental workload situation to capture changes in different mental workload indexes.
- The initial number of aircrafts presented was 9, and 6 of these were under participant control.
- Overall, a total of 70 aircraft were presented, with 50 coming from external locations A, D, E, F, W, P, N, L, M and Z and 20 from internal locations C and J. More specifically, 5 aircraft came from each of A, D, E, F, W, P, N, L, M and Z; 12 from J (8 going to P and 4 to A); and 8 from C (2 going to A, D, E and F respectively).



taught to use the scale just before beginning the experimental stage. While the method is relatively obtrusive to the primary task, it was considered the least intrusive of the available on-line workload assessment techniques [33, 34].

Participante: \_\_\_\_\_ Edad: \_\_\_\_\_

**ISA SCALE**

Responda según las siguientes opciones:

1 (Escasa carga), 2 (Algo de carga), 3 (Carga moderada), 4 (Bastante Carga), 5 (Máxima carga).

5 MIN----- 1 2 3 4 5

10 MIN----- 1 2 3 4 5

15 MIN----- 1 2 3 4 5

20 MIN----- 1 2 3 4 5

**Fig. 2.** Instantaneous self-assessment scale

### 3.2 Participants

Thirty-two psychology students at the University of Granada participated in the study under the motivation of earning extra credit. Participant ages ranged from 19 to 29, with an average of 22.1 and a median of 22. A total of 23 women and 9 men participated. A further requirement was that none of the participants had any previous experience in ATCo tasks.

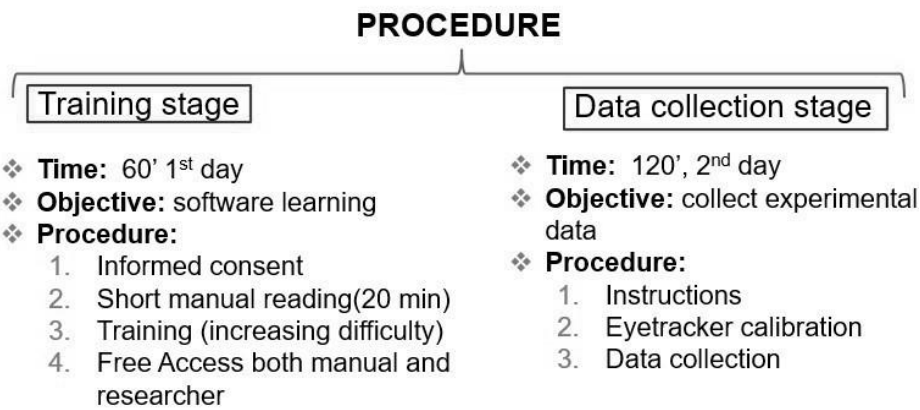
### 3.3 Procedure

Participants had to perform the ATCo tasks with the previously described <sup>ATC</sup>Lab-Advanced software, so they had to learn how to use it before proceeding through the experimental stage in which the performance data were collected. Thus, we established 2 distinct stages (see Figure 3):

1. **Training stage:** training took place for a total of 90 minutes. The main objective of this first stage was for participants to familiarise themselves with the software

so that they could handle it comfortably during the experimental stage. The training stage procedure was as follows: during the first day, once informed consent was given by the participants and their main task goal explained (maintaining air traffic security and preventing potential conflicts between aircraft), they started reading a short manual about the operation of the simulator for about 20 minutes and were asked to call the researcher once they had finished. Then the participants sat in front of the running simulator while the researcher reviewed the manual in detail with the participants to ensure both correct understanding of the task and assimilation of knowledge through content review. Participants then started using the simulator on their own while the researcher executed a total of six different ATC scenarios in order of difficulty. The participants had free access to both the manual and researcher at all times in case of doubts or questions. The researcher also periodically checked the participants' performance to monitor their learning. Once the training period concluded, participants were ready for the experimental session, which took place the following day.

2. **Data collection stage:** the aim of the data collection stage, which lasted a total of 120 minutes, was to collect experimental data from participants while they performed ATCo tasks. Both objective (physiological and execution data) and subjective (mental workload subjective index) data were collected. The participants were told the differences between the training and experimental stages, which were as follows: first, they would perform ATCo tasks in front of an eye-tracker system that had been previously calibrated. Secondly, participants were instructed to minimise head and body movements during the session and to fill in the ISA scale every 5 minutes, when a scheduled alarm sounded. The whole session lasted 2 hours. At the end of the session, the participants were thanked and given extra credit.



**Fig. 3.** Experiment design procedure diagram

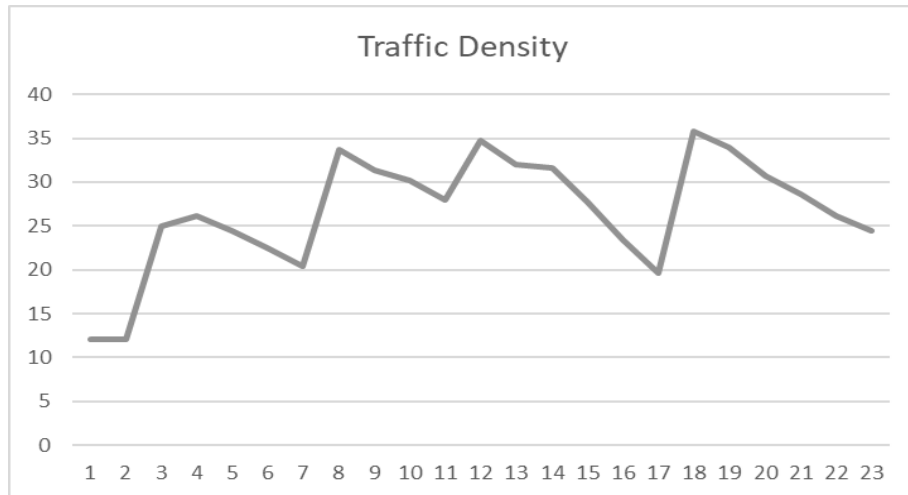
### 3.4 Experimental Room Conditions

Sessions were held in several different rooms, depending on whether the participant was in the training or data collection stage: during the training stage, participants could work in one of three different rooms equipped for training with the simulator, and no special attention to room conditions was needed. However, during the data collection stage, standardising room conditions was essential. Thus, the testing rooms were temperature controlled to 21°C, and lighting conditions (the main extraneous variable in pupil diameter measurement) were kept constant with artificial lighting and no natural light in the rooms. Moreover, participants always sat in the same place, a comfortable chair spaced 60 cm from the eye-tracker system.

### 3.5 Variables

#### **Independent Variables:**

*Task Complexity.* Complexity of the task is directly related to mental workload: the more complex the task becomes, the higher the mental workload. We manipulated task complexity by modifying aircraft traffic density (occupancy) through the 2 hours of the experimental session (see Figure 4).



**Fig. 4.** Aircraft density through intervals

*Time on Task.* TOT was manipulated by setting 5-minute intervals through the 120-minute duration of the experimental session. We thus set a total of 24 intervals from when the first measurement was taken as baseline for pupil diameter, resulting in 23 suitable time intervals.

#### **Dependent Variables:**

*Performance.* Although performance related to ATCo task can include a wide range of indicators, we found it appropriate to consider conflict rate as our performance indicator, taking into account its well-established correlation with our independent variable, traffic density [35]. However, for the analysis we operationalised this dependent variable by dividing the number of conflicts by the total number of aircraft present in the radar at a given time. We thought that using only the number of conflicts as a performance measurement would not be an appropriate performance indicator, since it largely reflects air traffic density.

*Pupil size.* We used pupil diameter as our physiological mental workload indicator, as it effectively reflects mental workload [36–44]. While our eye-tracking system allows continuous sampling rate recording at 120Hz, we set a total of 24 intervals lasting 5 minutes each to facilitate subsequent analyses.

Since expressing pupil size in absolute values has the disadvantage of being affected by slow random fluctuations of pupil size (source of noise), we followed the

recommendations provided by Sebastiaan Mathôt [45] about baseline correction of pupil-size data. To do this, we took the first interval (5 minutes) as a reference of standard individual average pupil size, which was then subtracted from the obtained value in each of the remaining 23 intervals, thereby giving a differential standardised value allowing us to reduce noise in our data. Analyses were carried out both for the left and right pupils. A negative value meant that the pupil was contracting while a positive value meant that it was dilating. Finally, we should bear in mind that some participants have larger pupils than others, but such between-subject differences are taken into account statistically, through a repeated measures ANOVA analysis.

*Subjective mental workload.* In order to make it possible to establish comparisons between the three primary workload measures, it was necessary to obtain the subjective momentary ratings continuously throughout the experimental session. With this goal, we used ISA, which is an on-line subjective workload scale created for this purpose. Ratings were obtained at 5-minute intervals throughout the 2 hours of the experimental stage, obtaining a total of 24 intervals, from which we discarded the first interval (used in the physiological scale as a baseline); this left 23 analysable intervals remaining (see Figure 2).

This study was carried out in accordance with the recommendations of the local ethical guidelines of the committee of the University of Granada institution called Comité de Ética de Investigación Humana. The protocol was approved by the Comité de Ética de Investigación Humana. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## 4. Results

We used a one-way, within-subject ANOVA to analyse the obtained results.

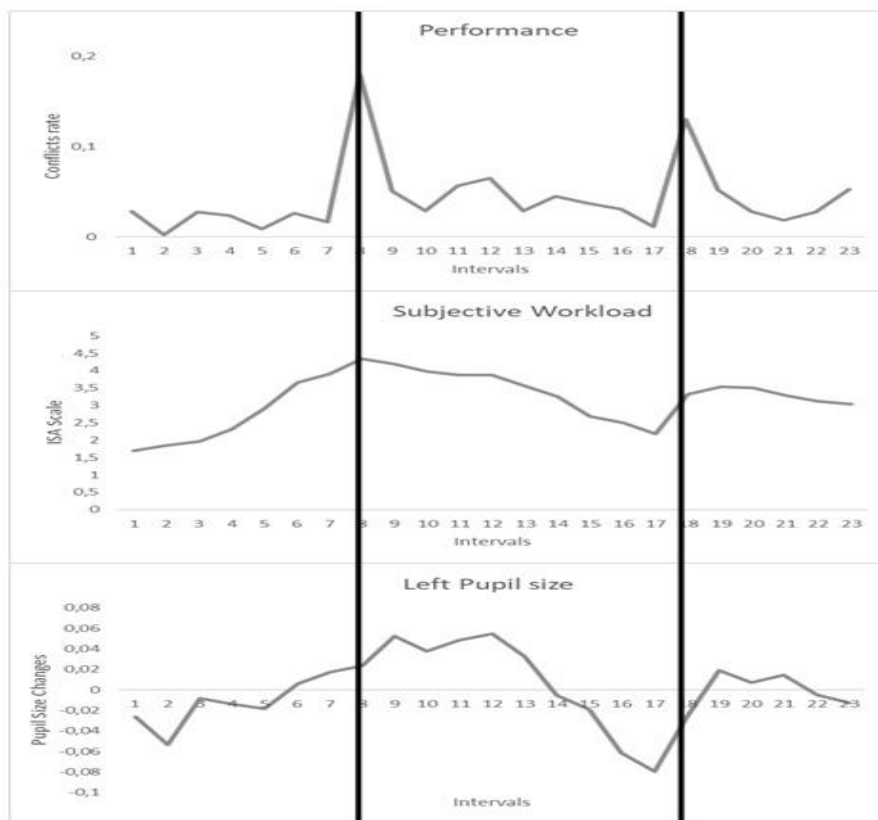
The analyses of participant performance during the experimental scenario confirmed that our task complexity manipulation was successful as, although participants were able to avoid most conflicts, the results showed a significant effect in intervals  $F(22,682) = 42.44$ ,  $MSe = .001$ ,  $p < .001$  on participant performance, due mainly to worse performance in intervals 8 and 18, which correspond to an increase in task demand (see Figure 5).

Left and right pupil diameter also showed the effect of workload at intervals  $F(22,682) = 7.98$ ,  $MSe = .005$ ,  $p < .001$ , and  $F(22,682) = 8.34$ ,  $MSe = .005$ ,  $p < .001$ , respectively.

Based on our results, we can see that pupil size changes over time, reaching maximum values with increases in task demands; in particular, we can find high pupil size peaks at intervals 9–12 and 19–21 (see Figure 5).

With regard to the subjective primary measure, we also found a significant main interval effect at  $F(22,682) = 25.17$ ,  $MSe = .758$ ,  $p < .001$ . Similar to pupil size, we can see how subjective mental workload changes over the intervals, reaching maximum peaks at higher task difficulty and more specifically at intervals 8–12 and 18–21 (see Figure 5).

Finally, Table 1 reveals that, as expected, there were high positive correlations between left and right pupils  $.97$ ,  $p < .01$ . Therefore, taking left pupil as a reference, we found a positive correlation between pupil diameter and subjective workload measure  $.81$ ,  $p < .01$ , whereas we could not find a correlation between pupil diameter and performance  $.27$ ,  $p > .05$ . Furthermore, we found only a slight correlation between performance and subjective workload measure  $.26$ ,  $p < .05$  (see Table 1).





**Fig. 5.** Participant performance, left pupil and subjective mental workload during task development**Table 1.** Correlation chart between measures

		Performance	R. pupil	L. pupil	Subjective
<b>Performance</b>	Pearson	1	.279	.266	.473*
	Sig. (bilateral)		.198	.219	.023
	N	23	23	23	23
<b>R. pupil</b>	Pearson	.279	1	.965**	.870**
	Sig. (bilateral)	.198		.000	.000
	N	23	23	23	23
<b>L. pupil</b>	Pearson	.266	.965**	1	.808**
	Sig. (bilateral)	.219	.000		.000
	N	23	23	23	23
<b>Subjective</b>	Pearson	.473*	.870**	.808**	1
	Sig. (bilateral)	.023	.000	.000	
	N	23	23	23	23

\*p &lt; .05, \*\*p &lt; .01

## 5. Discussion

To synthesise our data, our performance results revealed two lower performance peaks at intervals 8 and 18 (higher task demand), which correspond with physiological and subjective measures reaching maximum values (with different latencies). We should consider that, despite aircraft density reaching 4 high-density peaks at intervals 4, 8, 12 and 18, we only found 2 real high-complexity peaks reflected as bad performance peaks at intervals 8 and 18. In that respect, even though it seems from the literature that traffic density strongly influences complexity and so mental workload [46], complexity of ATCo scenarios are not only dependent on this factor but on others such as traffic configuration and aircraft evolution [46, 47, 25]. In that sense, by analysing the recorded scenario, we can actually verify that task demand does in fact become much higher at intervals 8 and

18, not only because of traffic density but also because of traffic evolution and configuration. That is the reason why there is not a comparable increase in mental workload reflections at interval 4 and 12 as were observed at 8 and 18.

The first important finding from our data is that pupil size and subjective workload measures follow a similar pattern: they increase at high-imposed task load, and they decrease when task demands become lower. Taking into account that both workload responses track to external task load, showing a positive correlation between each other (.87), our data revealed what has been called a double association [6].

A key finding should be emphasised. We found that at highest demand peaks (8 and 18), subjective mental workload measures reacted sooner than physiological response, even taking into account that subjective measures take some time for administration. In other words, subjective mental workload reflection responded with a lower latency than physiological measures, only at high-task load peaks. As we can see in the results, higher pupil size peaks are reached one interval after the highest demand peaks (intervals 9–12 and 19–21), whereas higher subjective workload responses were achieved immediately: at the first high demand peak at interval 8, and at the second high demand peak at intervals 18–19.

Therefore, although these correlations showed associations among our physiological and subjective measures, this collected data also suggested some indications of dissociations in line with the explanations provided by Hancock [6] about disassociations and insensitivities between mental workload measures: “there is strong reason to believe that each of these methods (and each of their component elements) possess their own inherent timescale and that certain, if not many, of the associations, dissociations and insensitivities are contingent upon such temporal differences” (p. 8). In other words, in our results there were some indications of the existence of different timescales for physiological and subjective measures. Further research is therefore necessary to explore this relation between insensitivities, dissociations and timescales.

The performance measure correlated only slightly with the subjective measure (.47), whereas it did not correlate with the physiological measure (.27). We think that these poor correlation results between performance and the rest of the primary indicators may be due to the low variability obtained on the performance variable (ceiling effect): participants were able to solve the vast majority of conflicts, and their performance was nearly perfect, so that variance was not particularly measurable. In any case, we believe that, in the absence of this ceiling effect, this lack of correlation would not have occurred.

It is worth noting an unexpected finding in our data that is not related to our main objective in this study: the effects of task learning on mental workload. At the first high-demand period, we found participants were experiencing more cognitive load, as they were not used to solving conflicts in that particular scenario. However, the more the participants practiced with the experimental scenario, the lower the cognitive demands as new cognitive strategies were being developed. We could therefore say that changing from a more consciously controlled to a more automated task would ultimately decrease mental workload. Our results showed up a mental workload decline at high demand peaks with TOT, which was reflected by the performance, subjective and physiological primary measures. This evidence is consistent with findings by Foroughi et al. (2017), who found that trial completion times and maximum pupil size significantly reduced across trials performing a cognitive task that required individuals to orient themselves in space relative to a target [48]. Furthermore, these results are also in line with cognitive load theory, which postulates that individuals would shift from learned procedures to automatic processes over time practising a new task [49].

In summary, by looking at our results, we could assume that differences in latency may partially explain mental workload dissociations and insensitivities between measures at high mental workload peak experiences. Moreover, mental workload decline over time could be explained by the effects of learning on cognitive efficiency, as literature research has shown.

Several limitations must be addressed. First, we think it would be necessary to deepen the study of this phenomenon with other physiological measures, such as electrodermal activity, heart-rate variability and EEG. Dissociations and insensibilities have been found not only between the three primary measures of workload, but also within different indicators of each kind of primary measure. In that sense, we are convinced that each different physiological indicator would have its own particular timescale, and further research is thus needed to untangle the issue. Another point to note is the granularity of the timescale considered on collected data. Intervals were defined in our experiment as every 5 minutes. We think it would be interesting to consider shorter intervals in order to obtain better granularity on data analysis. It is important to bear in mind, however, the related level of intrusiveness created by shorter data collection intervals: for example, an on-line subjective workload measure collected every 30 seconds would strongly interfere with both primary task development and experienced mental workload. Finally, it also needs to be noted that we conducted this research with students under simulated

conditions, and we think it would be necessary to replicate this experiment under real conditions to improve the validity of our findings.

## 6. Conclusions

Cognitive mental workload assessment is fundamental to the development of modern society and has been one of the greatest challenges for the last several decades. Hundreds of studies have been conducted for measuring mental workload in its three main axes: performance, subjective and physiological measures [3]. Although we would expect a certain degree of convergence between different mental workload measures, the literature has shown that associations do not always occur.

In this study, we explored the occurrence of disassociations and insensitivities across the different methods of measuring mental workload. A possible explanation for this might be related to the inherent timescale of each of these different methods. That is, each of the different methods for assessing a mental workload primary reflection possesses its own timescale, so that some show longer response latency to task demand changes than others [6].

Our findings suggest that dissociations and insensitivities may appear at high mental workload peak experiences due to latency differences between measures: subjective measure showed lower latency response than physiological response (pupil size). One important implication of this finding is that latency response in mental workload measures might vary depending on task demand levels. Thus, mental workload measures may correlate at low- and medium medium-task-demand levels, but when the situation becomes very complex (high cognitive-demand- peaks), dissociations and insensitivities between measures may appear. In other words, a worker experiencing moderate task demands may be reflecting moderate mental workload levels through physiological and subjective responses equally, but if the situations suddenly become much more difficult, we might better rely on subjective measures to reflect mental workload changes, rather than relying on physiological responses (at least those physiological variables reflecting physiological activation). Another interesting finding, in line with literature research, are the effects of task learning on cognitive efficiency, which was reflected as a mental workload decline with TOT.

However, due to methodological limitations, explained in the previous section, these promising findings must be taken cautiously before generalising them. Further research

should be undertaken to map the intrinsic timescale of the measurement problem, which will bring us closer to achieving a better understanding of associations, insensitivities and dissociations among mental workload responses.

## References

1. Crevits, I., Debernard, S., & Denecker, P.: Model building for air-traffic controllers' workload regulation. *European Journal of Operational Research*, **136**(2), 324-332 (2002). DOI: 10.1016/S0377-2217(01)00119-9
2. Moray, N. (ed.): *Mental workload: its theory and measurement*. Plenum Press, New York (1979). DOI: 10.1007/978-1-4757-0884-4\_2
3. Wickens, C. D.: *Mental workload: assessment, prediction and consequences*. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 18-29). Springer, Cham (2017). DOI:10.1007/978-3-319-61061-0\_2
4. De Alwis Edirisinghe, V.: *Estimating Mental Workload of University Students using Eye Parameters* (Master's thesis, NTNU) (2017)
5. Murai, K., Hayashi, Y., Okazaki, T., & Stone, L. C.: Evaluation of ship navigator's mental workload using nasal temperature and heart rate variability. In *2008 IEEE International Conference on Systems, Man and Cybernetics* (pp. 1528–1533). New York, NY: IEEE (2008). DOI: 10.1109/ICSMC.2008.4811503
6. Hancock, P. A.: *Whither workload? mapping a path for its future development*. In *International Symposium on Human Mental Workload: Models and Applications*, pp. 3–17. Springer, Cham. National Center for Biotechnology Information (2017). DOI: 10.1007/978-3-319-61061-0\_1
7. Yeh, Y. H., & Wickens, C. D.: *The dissociation of subjective measures of mental workload and performance (final report)*. (No. NASA-CR-176609; NAS 1.26:176609; EPL-84-2/NASA-84-2) (1984)
8. Casper, P. A. *Dissociations among measures of mental workload: Effects of experimenter-induced inadequacy* (1990)
9. Moray, N. (Ed.): *Mental workload: Its theory and measurement* (Vol. 8). Springer Science & Business Media (2013). DOI: 10.1007/978-1-4757-0884-4
10. Longo, L., & Leva, M. C. (Eds.): *Human Mental Workload: Models and Applications: First International Symposium, H-WORKLOAD 2017, Dublin, Ireland,*

- June 28–30, 2017, Revised Selected Papers (Vol. 726). Springer, Cham (2017). DOI 10.1007/978-3-319-61061-0
11. Dawson, D., Ian Noy, Y., Härmä, M., Åkerstedt, T., & Belenky, G.: Modelling fatigue and the use of fatigue models in work settings. *Accid. Anal. Preven.* **43**(2), 549–564 (2011). DOI: 10.1016/j.aap.2009.12.030
  12. Jorna, P. G.: Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological psychology*, **34**(2), 237–257 (1992). DOI: 10.1016/0301-0511(92)90017-O
  13. Endsley M.: From here to autonomy: lessons learned from human–automation research. *Human Factors*. **59**(1), 5–27 (2017). DOI: 10.1177/0018720816681350
  14. Josten, E. J., Ng-A-Tham, J. E., & Thierry, H.: The effects of extended workdays on fatigue, health, performance and satisfaction in nursing. *J. of advanced nursing*, **44**(6), 643–652 (2003). DOI: 10.1046/j.0309-2402.2003.02854.x
  15. Taylor, A. H., & Dorn, L.: Stress, fatigue, health, and risk of road traffic accidents among professional drivers: the contribution of physical inactivity. *Annu. Rev. Public Health*, **27**, 371–391 (2006). DOI: 10.1146/annurev.publhealth.27.021405.102117
  16. Fan, J., & Smith, A. P.: The impact of workload and fatigue on performance. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 90–105). Springer, Cham (2017). DOI: 10.1007/978-3-319-61061-0\_6
  17. Sawaragi, T., Horiguchi, Y., & Hina, A. (2006). Safety analysis of systemic accidents triggered by performance deviation. *제어로봇시스템학회 국제학술대회 논문집*, 1778–1781 (2006). DOI: 10.1109/SICE.2006.315635
  18. Edwards, T. E., Martin, L., Bienert, N., & Mercer, J.: Workload and Performance in Air Traffic Control: Exploring the Influence of Levels of Automation and Variation in Task Demand (2017). DOI: 10.1007/978-3-319-61061-0\_8
  19. Brookhuis, K. A., & de Waard, D.: Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention*, **42**(3), 898–903 (2010). DOI: 10.1016/j.aap.2009.06.001
  20. da Silva, F. P.: Mental workload, task demand and driving performance: what relation? *Procedia-Social and Behavioral Sciences*, **162**, 310–319 (2014). DOI: 10.1016/j.sbspro.2014.12.212
  21. Paxion, J., Galy, E., & Berthelon, C.: Mental workload and driving. *Frontiers in psychology*, **5**, 1344 (2014). DOI: 10.3389/fpsyg.2014.01344

22. Wickens, C. D.: Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, **3**(2), 159-177 (2002). DOI: 10.1518/001872008X288394. 2008 50: 449
23. Munoz-de-Escalon, E. & Cañas, J.: On-line measuring of available resources. *H-Workload 2017: The first international symposium on human mental workload*, Dublin Institute of Technology, Dublin, Ireland, June 28-30. (2017). DOI:10.21427/D7DK96
24. Cañas, J.J., Ferreira, P.N.P, Puntero, E., López, P., López, E., Gomez-Comendador V.F.: An air traffic controller psychological model with automation. In: 7th EASN International Conference: “Innovation in European Aeronautics Research”. Warsaw, Poland. (2017). DOI:10.3390/s18051586. 4
25. Majumdar, A., & Ochieng, W.: Factors affecting air traffic controller workload: Multivariate analysis based on simulation modeling of controller workload. *Transportation Research Record*. **1788**, 58–69 (2002). DOI: 10.3141/1788-08
26. Wu, C., & Liu, Y.: Queuing network modeling of driver workload and performance. *IEEE Transactions on Intelligent Transportation Systems*, **8**(3), 528-537 (2007) DOI: 10.1109/TITS.2007.903443
27. Sozou, P.D., Lane, P.C., Addis, M., Gobet, F.: Computational scientific discovery. In: Magnani, L., Bertolotti, T. (eds.) *Springer Handbook of Model-Based Science*, pp. 719–734. Springer (2017). DOI: 10.1007/978-3-319-30526-4\_33
28. Moustafa, K., Luz, S., & Longo, L.: Assessment of mental workload: a comparison of machine learning methods and subjective assessment techniques. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 30-50). Springer, Cham (2017). DOI 10.1007/978-3-319-61061-0\_3
29. Rizzo, L., Longo, L.: Representing and inferring mental workload via defeasible reasoning: a comparison with the NASA Task Load Index and the Workload Profile. *1st Workshop on Advances In Argumentation In Artificial Intelligence*, Bari, Italy, 2017 (2017)
30. Rizzo, L., Dondio, P., Delany, S. J., & Longo, L.: Modeling mental workload via rule-based expert system: a comparison with NASA-TLX and workload profile. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 215-229). Springer, Cham (2016). DOI: 10.1007/978-3-319-44944-9\_19
31. Marinescu, A. C., Sharples, S., Ritchie, A. C., Sánchez López, T., McDowell, M., & Morvan, H. P.: Physiological parameter response to variation of mental workload. *Human factors*, **60**(1), 31-56 (2018). DOI: 10.1177/0018720817733101

32. Fothergill, S., Loft, S., & Neal, A.: ATC-labAdvanced: An air traffic control simulator with realism and control. *Behav Res. Methods.* **41**(1), 118–127 (2009). DOI:10.3758/BRM.41.1.118
33. Brennan, S.D: An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. DRA Technical Memorandum (CAD5) 92017, DRA Maritime Command and Control Division, Portsmouth (1992)
34. Jordan.C.S. Experimental study of the effect of an instantaneous self assessment workload recorder on task performance. DRA Technical Memorandum (CAD5) 92011. DRA Maritime Command Control Division, Portsmouth (1992)
35. Prandini, M., Piroddi, L., Puechmorel, S., & Brázdilová, S. L.: Toward air traffic complexity assessment in new generation air traffic management systems. *IEEE transactions on intelligent transportation systems*, **12**(3), 809-818 (2011). DOI: 10.1109/TITS.2011.2113175
36. Matthews, G., Middleton, W., Gilmartin, B. y Bullimore, M. A.: Pupillary diameter and cognitive and cognitive load. *J. Psychophysiol.* **5**, 265–271 (1991)
37. Bacs, R. W. y Walrath, L. C.: Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Appl. Ergon.* **23**, 243–254 (1992). DOI: 10.1016/0003-6870(92)90152-L
38. Hyönä, J., Tommola, J. y Alaja, A.: Pupil dilation as a measure of processing load in simultaneous interpreting and other language tasks. *Q. J. Exp. Psychol.* **48**, 598–612 (1995). DOI: 10.1080/14640749508401407
39. Granholm, E., Asarnow, R. F., Sarkin, A. J. y Dykes, K. L.: Pupillary responses index cognitive resource limitations. *Psychophysiology*, **33**, 457–461 (1996). DOI: 10.1111/j.1469-8986.1996.tb01071.x
40. Iqbal, S. T., Zheng, X. S. y Bailey, B. P.; Task evoked pupillary response to mental workload in human-computer interaction. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1477–1480. ACM, New York (2004). DOI: 10.1145/985921.986094
41. Verney S. P., Granholm, E. y Marshall S. P.: Pupillary responses on the visual backward masking task reflect general cognitive ability. *Int. J. Psychophysiol.* **52**, 23–36 (2004). DOI: 10.1016/j.ijpsycho.2003.12.003
42. Porter, G., Troscianko, T. y Gilchrist, I. D.: Effort during visual search and counting: insights from pupillometry. *Q. J. Exp. Psychol.* **60**, 211–229 (2007). DOI: 10.1080/17470210600673818



43. Privitera, C. M., Renninger, L. W., Carney, T., Klein, S. y Aguilar, M.: Pupil dilation during visual target detection. *J, Vision*. **10**, 1–14 (2010). DOI:10.1167/10.10.3
44. Reiner, M., & Gelfeld, T. M.: Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *Int. J. Psychophysiol*. **93**(1), 38–44 (2014). DOI: 10.1016/j.ijpsycho.2013.11.002
45. Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S.: Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav. Res. Methods*. **50**(1), 94–106 (2018). DOI: 10.3758/s13428-017-1007-2
46. Mogford, R. H., Guttman, J. A., Morrow, S. L., & Kopardekar, P.: *The Complexity Construct in Air Traffic Control: A Review and Synthesis of the Literature*. CTA INC MCKEE CITY NJ (1995)
47. Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., & Collet, C.: ATC complexity and controller workload: Trying to bridge the gap. In *Proceedings of the International Conference on HCI in Aeronautics*, pp. 56–60. AAAI Press, Cambridge, MA (2002)
48. Foroughi, C. K., Sibley, C., & Coyne, J. T.: Pupil size as a measure of within-task learning. *Psychophysiology*. **54**(10), 1436–1443 (2017). DOI: 10.1111/psyp.12896
49. Sweller, J.: Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*. **4**(4), 295–312 (1994). DOI: 10.1016/0959-4752(94)90003-5



## **Chapter III**

---

*Task Demand Transition Rates of Change*

*Effects on Mental Workload Measures*

*Divergence*



## **Task demand transition rates of change effects on mental workload measures divergence**

Enrique Muñoz-de-Escalona<sup>1</sup>

José Juan Cañas<sup>1</sup>

Jair van Nes<sup>2</sup>

<sup>1</sup> Mind, Brain and Behaviour Research Centre,  
Department of Experimental Psychology, Faculty of Psychology,  
University of Granada

<sup>2</sup> Faculty of Psychology and Neuroscience, University of Maastricht, Maastricht, Netherlands  
vannes@[student.maastrichtuniversity.nl](mailto:student.maastrichtuniversity.nl)

---

This article has been published: Muñoz-de-Escalona, E., Cañas, J. J., & van Nes, J. (2019, November). Task Demand Transition Rates of Change Effects on Mental Workload Measures Divergence. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 48-65). Springer, Cham. [https://doi.org/10.1007/978-3-030-32423-0\\_4](https://doi.org/10.1007/978-3-030-32423-0_4)

This article received an award for being the best student paper presented in the “The 3<sup>th</sup> International Symposium on Human Mental Workload: models and applications (H-Workload 2019)”, held in Rome, Italy, 2019.

## Abstract

Mental workload is a complex construct that may be indirectly inferred from physiological responses, as well as subjective and performance ratings. Since the three measures should reflect changes in task-load, one would expect convergence, yet divergence between the measures has been reported. A potential explanation could be related to the differential sensitivity of mental workload measures to rates of change in task-load transitions: some measures might be more sensitive to change than the absolute level of task demand. The present study aims to investigate whether this fact could explain certain divergences between mental workload measures. This was tested by manipulating task-load transitions and its rate of change over time during a monitoring experiment and by collecting data on physiological, subjective, and performance measures. The results showed higher pupil size and performance measure sensitivity to abrupt task-load increases: sensitivity to rates of change could partially explain mental workload dissociations and insensitivities between measures.

**Keywords:** mental workload, workload measures, convergence, divergence dissociations, insensitivities, task demand transitions, rates of change.

## 1. Introduction

The study of mental workload is of crucial importance in many fields such as Emergency Room healthcare or air traffic control (ATC), in which the lives of countless people are at stake and dependent on human performance. Human related accidents address an ongoing problem in social sciences and cognitive ergonomics: how can we minimize and avoid human error?

A high mental workload generally leads to poor performance [3, 4], and extreme cases of overload may result in errors which can in turn lead to fatal accidents. On the other hand, mental underload remains as undesirable as mental overload, likewise leading to poor performance and errors [5]. Automation, which does have its benefits, is significantly and repeatedly associated with mental underload [6], and can cause a major loss of situation awareness, making it difficult for people to detect flaws and intervene adequately and timely. Therefore, optimizing levels of mental workload is vital to maintaining effective performance, avoiding both overload and underload during task performance. Furthermore, high levels of mental workload can have detrimental effects on people's psychosocial and physical health. High mental workload has been associated with high work-related fatigue, high-stress complaints, and/or burnout [7], as well as high scores on health complaint questionnaires [8]. Therefore, by using mental workload measures, we can learn its limits and dimensions and appropriately discover how to improve performance and minimize human error within organizations or within people's own personal work practices [4].

Task demand is dynamic in many fields such as ATC and aircraft pilots: workers may experience changes in mental workload as, for example, traffic load gets higher or a sudden unexpected storm appears when flying an aircraft, respectively. These changes in task-load may also be gradual or abrupt and affect individuals' mental workload, as well as the way its measures reflect these changes.

To better understand and assess the construct of mental workload, the present study aims to investigate whether differential sensitivity to rates of change in task demand transitions would affect the convergence of mental workload measures. However, we will begin by defining mental workload as it in itself is a very poorly defined construct. Furthermore, we will define convergence and divergence phenomena between mental workload measures (associations, dissociations, and insensitivities), as well as the above-

mentioned sensitivity to rates of change, to better understand the current mental workload assessment literature.

The rest of the paper proceeds as follows. Section 2 outlines related work about general mental workload, focusing on defining and measuring mental workload and on introducing sensitivity to rates of change in task demand transitions. Section 3 describes the design and the methodology followed in conducting the experiment. Section 4 presents the obtained results, while Section 5 presents a discussion about our findings, as well as limitations and possible future work. Finally, Section 6 concludes the study, summarizing its key findings and implications on the extant body of knowledge.

## 2. Related Work

### 2.1 Defining mental workload

Mental workload is a complex construct without a clear consensus regarding a definition [9, 10, 11]. It has been defined as the product of the immediate demands of the environment and an individual's maximum mental capacity [12, 13], hence mental workload is a multi-factorial construct, which depends not only on demanded task resources but also on those available [14, 15, 16]. When the demands of the environment exceed a person's maximum mental capacity, mental overload occurs and performance deteriorates as a consequence of our limited capacity [17]. This is because when at the very limit of our human mental resources, we are unable to reallocate these resources in an adaptable way. On the contrary, when the environment demands too little, such as in work situations that are heavily automated, mental underload occurs and performance similarly deteriorates. Why low environmental demands are detrimental is still poorly understood, however, some have suggested this may be due to shrinkages in our maximum mental capacity in response to environmental demand reduction [5, 18]; this in turn can influence several factors, including vigilance, workload, attention, and situation awareness [6].

For the purpose of our research, mental workload may be considered as the amount of mental effort involved in performing a task [3, 19]. In other words, the amount of mental resources in use during the performance of a task given the demands of the environment. The term task-load refers to environmental demands and it is used to manipulate the amount of experienced mental workload. In general, we say that measures reflecting



mental workload are valid when they reflect a change in task-load (the demands of the immediate environment).

## 2.2 Measuring mental workload

The measurement of mental workload is one of the biggest challenges facing psychology and social sciences at present. There is a widespread need to assess cognitive work as, on the one hand, it is fundamental to the development of modern society and, on the other, it has been identified as one of the main causes of work-related accidents. Despite mental workload not being a directly measurable construct, it can be assessed with three types of individual “primary measures” reflecting mental workload [9, 11, 16, 20]: (a) physiological responses (Electroencephalography (other brain imaging techniques, heart-rate variability (HRV), pupil diameter, etc.); (b) perceived or subjective perception of mental workload (questionnaire or scale response); and (c) task performance (response speed and accuracy). According to Hancock (2017), if these three measures of mental workload mean to assess the same construct then they should demonstrate convergence. In other words, if task-load were to increase then we would expect the following associations between task-load changes and primary measures: (a) higher physiological activation responses, (b) higher perceived mental workload, and (c) a decrease in task performance. Thus, there should exist a convergence between these measures of mental workload, given the expected association of task-load with each respective measure. However, current research has demonstrated that this is not always the case: dissociations and insensitivities between mental workload measures have been repeatedly reported [16, 20, 21].

Insensitivities occur when a certain task has distinct levels of task-load, but measures of mental workload fail to reflect a change regardless of task-load levels. For example, when piloting an airplane, pilots may have to deal with dynamic changes within the immediate environment. Task-load may increase in line with the increasing demands of a complex situation, yet our measures reflect static levels of reported mental workload. For instance, task-load may increase due to air turbulence or even a failure in automation, yet physiological measures (such as pupil dilation or HRV) reflect no change. One must note, however, that it is possible for an insensitivity to occur for one measure of mental workload but not for the other. Given the current example, it may be that pilots’ physiological measures reflect no change in their actual mental workload, yet they can

report an increase (or even decrease) in their perceived mental workload. Furthermore, dissociations occur when we have contradictory results: task-load increases but subjects report a lower perceived mental workload, whereas normally one would expect an increase in perceived mental workload if task-load were to increase. Using our pilot example again, the situation at hand may increase task-load but pilots report (a) lower physiological activation responses (dissociation), (b) higher perceived mental workload (association), and (c) static levels in task performance (insensitivity).

There are several factors that might affect the occurrence of dissociation and insensitivities between task-load and primary mental workload measures [20]. One possible explanation might be related to the timescale considered between measures. Muñoz-de-Escalona & Cañas (2018) identified that dissociations and insensitivities may appear at high mental workload peak experiences due to latency differences between measures: subjective measures showed lower levels of latency response than physiological response (pupil size) [16]. Despite this, there are also several other factors that might contribute to the emergence of divergence between mental workload measures, including sensitivity to rates of change [20].

### **2.3 Task demand transitions and sensitivity to rates of change**

Task demand transitions research has been very limited and largely focused on its effects in task performance and mental workload perceptions [22]. The current literature has revealed that a change in task-load levels affects mental workload, fatigue, and performance ratings [22, 23, 24, 25]. However, the authors could not find any research focusing on how sensitivity to rates of change in task demand transitions affects the convergence and divergence of mental workload measures. Behavioral sciences has repeatedly shown that humans are more sensitive to change rather than the absolute level of a stimulus [20]. If we translate this into the study of mental workload convergence and divergence between measures, it may be possible that there exists a differential sensitivity to task-load rate of change in mental workload measures. Some measures might be more sensitive to change than to the absolute level of task demand, while others might be sensitive only to absolute levels of task demand. These differences would result in dissociations and insensitivities which would ultimately affect convergence between mental workload measures.

This study aims to shed light on the effects of sensitivity to rates of change in task demand transitions on the convergence of mental workload measures. In the present study, we manipulated two independent variables: (1) task-load rate of change and (2) task-load change direction, whose effects were tested in a task-battery experiment in which participants were trained and instructed to perform to the best of their abilities. Participants performed the task-battery for 20 minutes, whilst data on the dependent variables, task performance, pupil diameter, and perceived mental workload were obtained. We hypothesized that there would be higher divergence (dissociations and/or insensitivities) between mental workload measures with abrupt rate of change conditions rather than linear ones.

### 3. Design and Methodology

#### 3.1 Materials and Instruments

**MATB-II Software.** Measurements of task performance were collected through the use of the second version of the Multi-Attribute Task Battery (MATB-II), a computer program designed to evaluate operator performance and workload through means of different tasks similar to those carried out by flight crews, with a user-friendly interface as to allow non-pilot participants to utilize it [25]. MATB-II comes with default event files which can easily be altered to adapt to the needs or objectives of an experiment. The program records events presented to participants, as well as participants' responses. The MATB-II contains the following four tasks: the system monitoring task (SYSMON), the tracking task (TRACK), the communications task (COMM), and the resource management task (RESMAN) (see Figure 1).

1) The SYSMON task is divided into two sub-tasks: lights and scales. For the lights sub-task, participants are required to respond as fast as possible to a green light that turns off and a red light that turns on, and to turn them back on and off, respectively. For the scale sub-task, participants are required to detect when the lights on four moving scales deviate from their normal position and respond accordingly by clicking on the deviated scale.

2) In the TRACK task, during manual mode, participants are required to keep a circular target in the center of an inner box displayed on the program by using a joystick with their

left hand (the dominant hand was needed for the use of the mouse). During automatic mode, the circular target will remain in the inner box by itself.

3) In the COMM task, an audio message is played with a specific call sign and the participant is required to respond by selecting the appropriate radio and adjusting for the correct frequency, but only if the call sign matches their own (call sign: “NASA504”). No response is required of the participant for messages from other call signs.

4) In the RESMAN task, participants are required to maintain the level of fuel in tanks A and B, within  $\pm 500$  units of the initial condition of 2500 units each. In order to maintain this objective, participants must transfer fuel from supply tanks to A and B or transfer fuel between the two tanks.

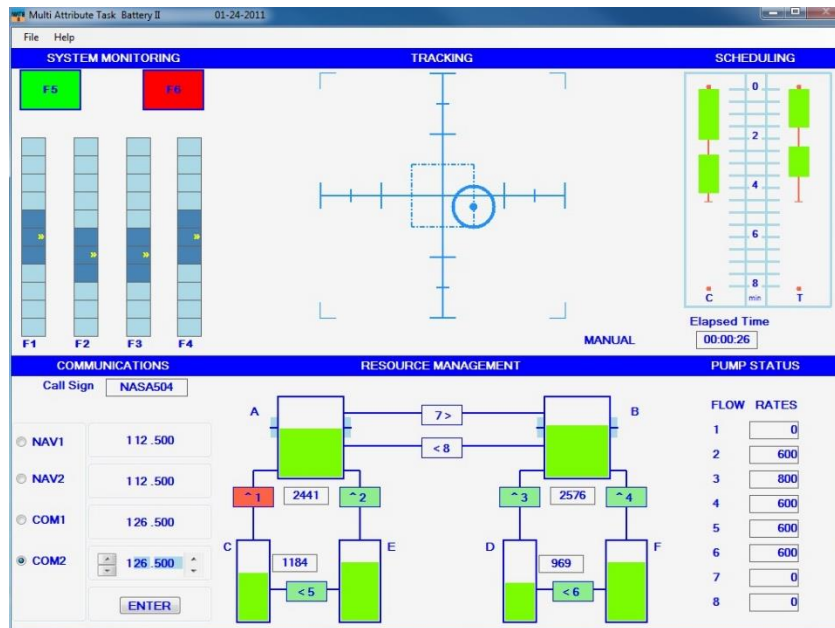


Figure 1: MATB-II task display. Taken from <https://matb.larc.nasa.gov/>

**Tobii T120 Eyetracker.** Pupil diameter measurements were obtained using an infrared-based eye tracker system, the Tobii T120 model marketed by Tobii Video System (see Figure 2). This system is characterized by its high sampling frequency (120 Hz).

This equipment is completely non-intrusive, has no visible eye movement monitoring system, and provides high precision and an excellent head compensatory movement mechanism, which ensures high-quality data collection. In addition, a calibration procedure is completed within seconds, and the freedom of movement it offers participants allows them to act naturally in front of the screen, as though it were an

ordinary computer display. Thus, the equipment allows for natural conditions in which to measure eye-tracking data [26].



**Figure 2:** Tobii T120 Eyetracker system

**Instantaneous Self-assessment Scale.** We employed an easy and intuitive instant subjective workload scale called instantaneous self-assessment (ISA), which provides momentary subjective ratings of perceived mental workload during task performance (see Figure 3). ISA has been used extensively in numerous domains, including during ATC tasks. Participants write down how much mental workload they currently experience on a scale ranging from 1 (no mental workload) to 5 (maximum mental workload), presented from left to right in ascending order of mental workload experienced. Participants were taught to use the scale just before beginning the experimental stage. While the method is relatively obtrusive, it was considered the least intrusive of the available on-line workload assessment techniques [27, 28].

Participante: \_\_\_\_\_ Edad: \_\_\_\_\_

**ISA SCALE**

Responda según las siguientes opciones:

1 (Escasa de carga), 2 (Algo de carga), 3 (Carga moderada), 4 (Bastante carga), 5 (Máxima carga)

2.5 MIN \_\_\_\_\_ 1 2 3 4 5

5 MIN \_\_\_\_\_ 1 2 3 4 5

7.5 MIN \_\_\_\_\_ 1 2 3 4 5

10 MIN \_\_\_\_\_ 1 2 3 4 5

12.5 MIN \_\_\_\_\_ 1 2 3 4 5

15 MIN \_\_\_\_\_ 1 2 3 4 5

17.5 MIN \_\_\_\_\_ 1 2 3 4 5

20 MIN \_\_\_\_\_ 1 2 3 4 5

**Fig. 3.** Instantaneous self-assessment scale

### 3.2 Participants

Fifty-six psychology students from the University of Granada participated in the study. Participants’ ages ranged from 18 to 30, with an average of 22.7 and a standard deviation of 4. A total of 39 women and 17 men participated. It should be noted that there is a greater number of female participants due to the fact that psychology students at the University of Granada are mostly women. Recruitment was achieved through the dispersion of posters and flyers around the university, as well as an advertisement for the study on the university’s on-line platform for experiments (<http://experimentos.psiexpugr.es/>). The requirements for participation included (1) not being familiar with the MATB-II program, (2) Spanish as a native language, and (3) visual acuity or correction of visual impairment with contact lenses, as glasses impair the utilized eye-tracking device from collecting data. Participants’ participation was rewarded with two experimental vouchers for which they received extra credit.

### 3.3 Procedure

3. **Training stage:** training took place for no longer than 30 minutes. The objective of this stage was for participants to familiarize themselves with the program so that they could carry out the tasks securely during the data collection stage. The

procedure was conducted as follows: upon entering the lab and after filling out the informed consent form, the participant was instructed to read the MATB-II instruction manual and inform the researcher once they had finished. The researcher then sat down with the participant to allow for questions and resolve any doubts on how to use the program. Afterward, on a computer monitor, participants were presented each MATB-II task separately and were first given a demonstration as to how to execute the task and given time to perform the task themselves. The participants were always free to consult the manual and ask the researcher questions during the training stage in case of doubts or uncertainties. Once the participants had completed all four tasks and resolved all doubts, they were ready for the data collection stage, which followed immediately afterwards. During the training stage, participants could work in one of three different rooms equipped for training with the MATB-II software, and no special attention to room conditions was needed.

4. **Data collection stage:** the data collection stage lasted approximately 20 minutes and involved participants completing 1 of the 4 randomly assigned experimental conditions, while task performance, perceived mental workload, and pupil diameter were recorded. The participants were instructed to fill in the ISA scale every 2 and a half minutes when a scheduled alarm sounded. Prior to the start of the task-battery, the eye-tracker system was calibrated, and the participants were told to keep head and body movements to a minimum. During the data collection stage, standardizing room conditions was essential. Thus, the testing rooms were temperature controlled to 21°C, and lighting conditions (the main extraneous variable in pupil diameter measurement) were kept constant with artificial lighting; there was no natural light in the rooms. Moreover, participants always sat in the same place, a comfortable chair spaced 60 cm from the eye-tracker system.

This study was carried out in accordance with the recommendations of the local ethical guidelines of the committee of the University of Granada institution: Comité de Ética de Investigación Humana. The protocol was approved by the Comité de Ética de Investigación Humana under the code: 779/CEIH/2019. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

### 3.4 Variables

#### Independent Variables:

In the present study we manipulated 2 independent variables:

- *Task-load rate of change*: this is the intensity in which task demand changes over time. We manipulated the task-load rate of change by modifying the number and the combination of active tasks that participants had to perform over time; this occurred during the data collection stage. We established 2 levels: (1) the variable rate of change and (2) the linear rate of change. Possible task combinations are illustrated in Table 1.

**Table 1:** Possible Task Combination in MATB-II Software

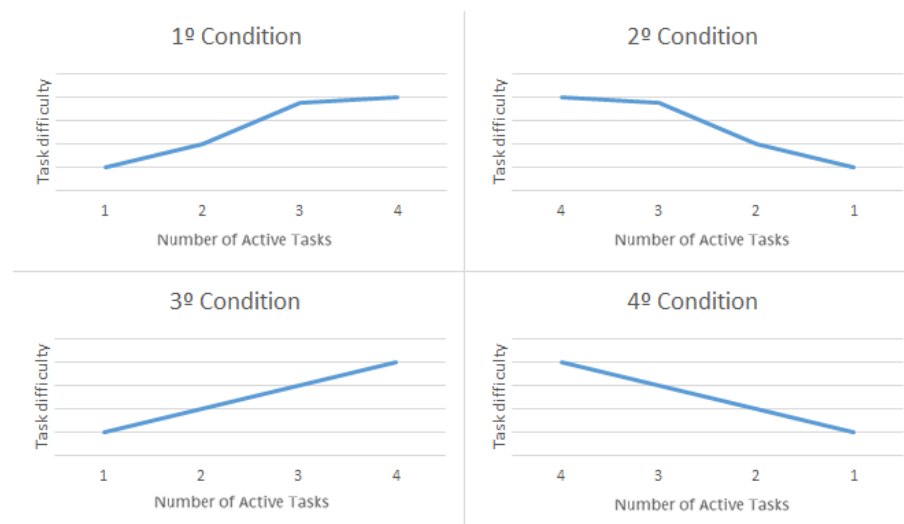
<b>Task Combination</b>	<b>MATB-II Active tasks</b>
1	SYSMON
2	SYSMON + TRACK
3	SYSMON + TRACK + COMM
4	SYSMON + COMM + RESMAN
5	SYSMON + TRACK + COMM + RESSMAN

- *Task-load change direction*: this is the direction in which task-load changes over time. Since task demand transitions can occur in two directions, we manipulated this variable on 2 levels: (1) increasing task-load change and (2) decreasing task-load change.

As a result of the manipulation of these two variables, we obtained 4 experimental conditions (see Figure 4), namely:



1. **Condition: Increasing task-load with a variable rate of change.** Task-load increased every 5 minutes, but with a variable rate of change. Participants had to perform 1, 2, 4, and 5 sets of task combinations: there is an initial rate of change from task combination 1 to 2, then there is an abrupt increase in the rate of change from task combination 2 to 4 and then a decrease in the rate of change from task combination 4 to 5.
2. **Condition: Decreasing task-load with a variable rate of change.** Participants performed the same set of tasks from condition (1) but in descending order, resulting in 5, 4, 2, and 1 task combinations.
3. **Condition: Increasing task-load with a linear rate of change.** Task-load increased every 5 minutes with a linear rate of change. Participants had to perform task combinations 1, 2, 3, and 5.
4. **Condition: Decreasing task-load with a linear rate of change.** Participants performed the same set of tasks from condition (3) but in descending order, resulting in task combinations 5, 3, 2, and 1.



**Figure 4:** Task-load rate of change evolution for experimental conditions.

Note that the difference between the increasing variable and linear rate of change conditions demonstrates that the former condition involves a sharp increase in task-

demands from task combination 2 to 4. The reason for this lies with the elimination of an already practiced task (TRACKING) and the addition of 2 new non-practiced tasks (COMM & RESMAN), whereas the latter conditions involve a linear increase in task-demands from task combination 2 to 3 since only a single new task is added (COMM) and vice-versa regarding a decreasing variable and a linear rate of change conditions.

### **Dependent Variables:**

*Performance.* MATB-II provides us with many indicators of participants' performance: e.g. root mean square deviation (RMSD) for the TRACK task, number of correct and incorrect responses for the SYSMON and COMM tasks, and the arithmetic mean of tanks "A-2500" and "B-2500" in absolute values for the RESMAN task. However, for the purposes of this experiment we will only consider the SYSMON performance indicator, as it is the only task present during all 4 of the task-load levels in the 4 experimental conditions, allowing us to compare participants' performances between conditions. The SYSMON performance indicator will be considered as the number of correct responses divided into the number of possible responses. The result is a number between 1 (best possible performance) and 0 (worse possible performance).

*Pupil size.* Mental workload can be reflected by several physiological indexes such as EEG, HVR, and several ocular metrics. We decided to use pupil diameter as our physiological mental workload indicator, as it effectively reflects mental workload [29–37] and minimize intrusiveness. While our eye-tracking system allows continuous sampling rate recording at 120Hz, we set a total of 8 intervals lasting 2.5 minutes each in order to obtain 2 measures per task-load level.

Since expressing pupil size in absolute values has the disadvantage of being affected by slow random fluctuations in pupil size (source of noise), we followed the recommendations provided by Sebastiaan Mathôt [38] regarding the baseline correction of pupil-size data. To do this, for every participant, we took his/her average pupil size during the session as a whole as a reference, which was then subtracted from the obtained value in each of the 8 intervals, thereby giving a differential standardized value allowing us to reduce noise in our data. Analyses were carried out for the average of both the left

and right pupils. A negative value meant that the pupil was contracting while a positive value meant that it was dilating.

*Subjective mental workload.* Traditional off-line subjective workload assessment tools, such as the NASA Taskload Index (NASA-TLX), do not allow researchers to obtain continuous subjective ratings from participants. In order to facilitate and establish comparisons between mental workload measures, it was necessary to obtain the subjective momentary ratings continuously throughout the experimental session. With this goal, we used the ISA, which is an on-line subjective workload scale created for this purpose. Ratings were obtained at 2.5-minute intervals throughout the 20 minutes of the experimental stage, obtaining a total of 8 subjective mental workload ratings (2 measures per task demand level).

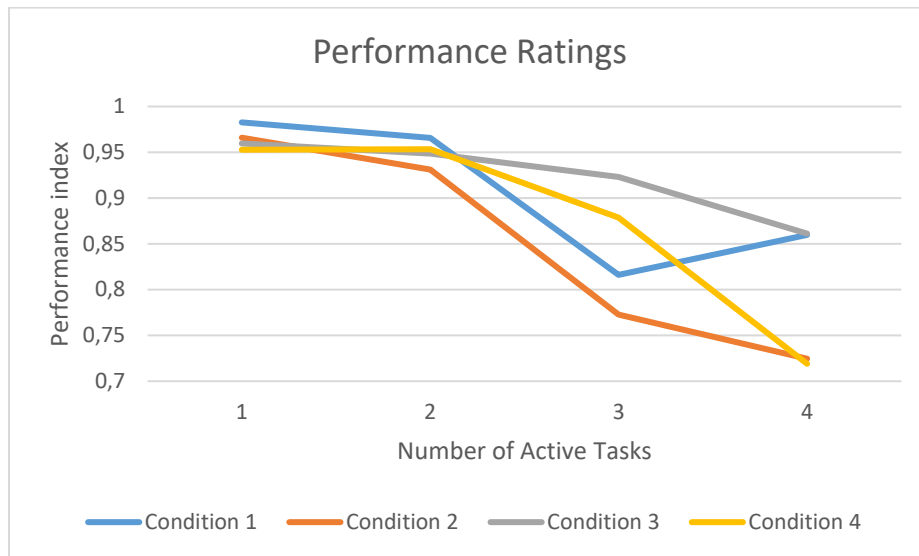
*Synchronization of measures.* Performance, pupil size, and subjective measures were obtained continuously throughout the experimental session. Synchronization between measures was simple, as the eyetracker and MATB-II performance log files began to record data simultaneously at the start of the experimental session. The scheduled alarm (every 2.5 minutes) was also synchronized by the experimenter, as it was simultaneously activated with the MATB-II software. This would also allow the ISA scale to be synchronized with the performance and pupil size measures.

## 4. Results

We used three one-way, within-subjects ANOVA to analyze the obtained results, one for each mental workload measurement.

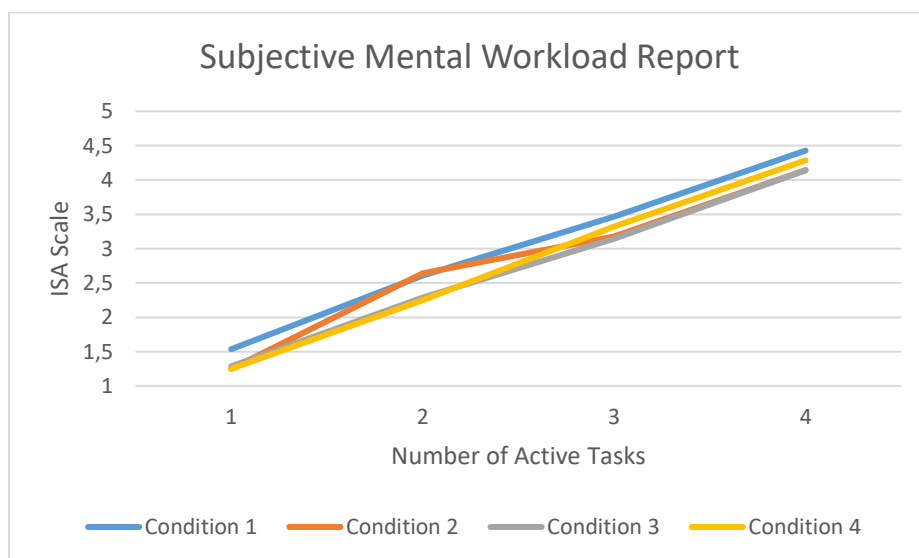
First, the analyses of participants' performance showed that our task-load level manipulation was successful. The ANOVA analyses identified a very significant main task-load level effect  $F(3,156) = 74.34$ ,  $MSe = .005$ ,  $p < .01$ , which reflects that participants' performances decreased as task demand increased. The main effect of the experimental condition was found to be significant  $F(3,52) = 3.24$ ,  $MSe = .02$ ,  $p < .01$ . Moreover, an interaction effect of the task-load level x experimental condition was also found to be significant  $F(9,156) = 6.14$ ,  $MSe = .005$ ,  $p < .01$ . This demonstrates that performance variations evolve differently throughout the different task-load levels depending on the considered experimental condition: there is a higher decrease in

performance in the variable rate of change conditions compared to the linear rate of change conditions from task-load levels 2 to 3 (see Figure 5).



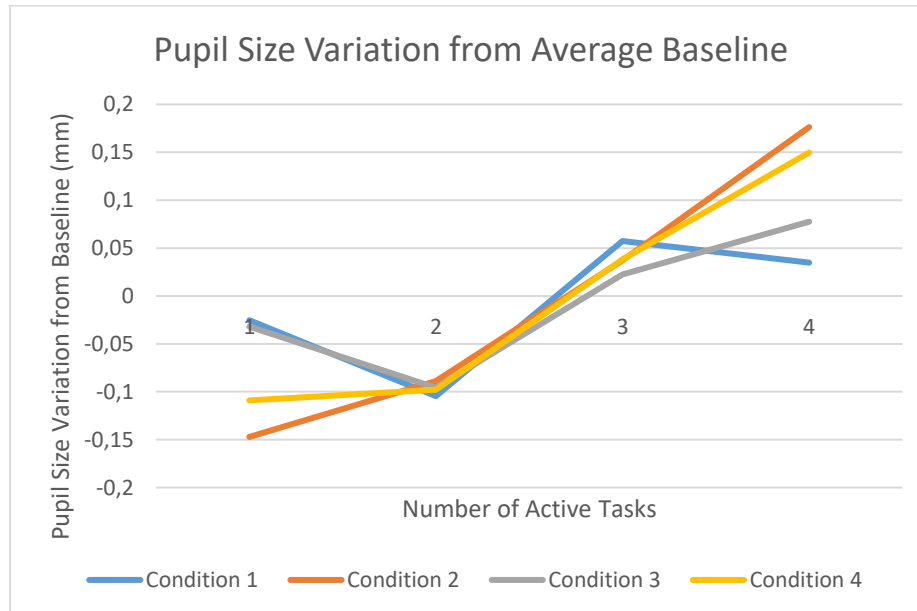
**Figure 5:** Participants’ performance during task development.

Regarding subjective perceptions, a linear increase in participants’ subjective mental workload ratings occurs between task-load levels in every experimental condition: in the ANOVA, the main effect of task-load levels turned out to be very significant indeed,  $F(3,156) = 358.9$ ;  $MSe = .240$ ,  $p < .01$ ; whereas the main effect of the experimental condition  $F(3,52) = 1.09$ ;  $MSe = .84$ ,  $p > .05$  and the interaction of the task-load level x experimental condition  $F(9,156) = .72$ ;  $MSe = .240$ ,  $p > .05$  were not significant (see Figure 6).



**Figure 6:** Participants’ subjective mental workload ratings during task development.

For pupil size, our physiological measurement, we also found a very significant main effect of task-load level  $F(3,156) = 70.94, p < .01$  which supported participants' pupil size increasing (higher activation) as task demand rose. Despite the main effect of the experimental condition not being significant  $F(3,52) = 1.03, MSe = .00, p > .05$ , a significant interaction effect of the task-load level x experimental condition was found  $F(9,156) = 4.93; MSe = .008, p < .01$ . This implies that pupil size variation evolves differently through task-load levels depending on the considered experimental condition: in decreasing task-load conditions (2 & 4), pupil size increases linearly with higher task demand, regardless of the task-load rate of change; whereas in increasing task-load conditions (1 & 3), pupil size starts at a higher state of dilation (task-load level 1), then decreases in the following task-load level (task-load level 2). From task-load level 2 to 3, there is a higher dilation in the variable rate of change condition (1) regarding the linear condition (3). Finally, from task-load level 3 to 4, pupil dilation decreases in the variable rate of change condition (1), whereas it continues increasing in the linear rate of change condition (3) (see Figure 7).



**Figure 7:** Participants' pupil size variation ratings from the average baseline during task development.

## 5. Discussion

To synthesize our data, we can see how subjective ratings increased linearly as task-load rose in every experimental condition, whereas the performance and physiological measures reacted differently depending on the experimental condition in question. Performance linearly decreased as task-load increased in both (increasing and decreasing) linear rate of change conditions (3 & 4), as well as in the decreasing variable rate of change condition (2). Conversely, in the increasing variable rate of change condition (1), performance decreased until task-load level 3 and then improved again from task-load levels 3 to 4. Pupil size increased linearly in both decreasing task-load conditions (2 & 4) as task demand increased; while in both increasing task-load conditions there was a decrease in pupil size from task-load levels 1 to 2, followed by an increase from task-load levels 2 to 3. Finally, from task-load levels 3 to 4, pupil size continued increasing in the linear rate of change condition but decreased in its variable rate of change counterpart.

Hence, according to our data and in terms of association, dissociation, and insensitivities:

- Regarding *subjective response*, we found associations between task-load and subjective perceptions in every experimental condition: there is a direct mapping between mental workload subjective ratings and task demand. We can observe how subjective ratings have not been influenced either by task-load rate of change or task-load change direction.
- Regarding *performance response*, our results presented associations or dissociations depending on the considered experimental condition: we found associations between task-load and performance responses in both (increasing and decreasing) the linear rate of change conditions and in the decreasing variable rate of change condition; performance became more linearly impaired as task-load increased. However, in the increasing variable rate of change condition, we found associations between task demand and performance responses until task-load level 3, but an improvement in performance was observed from task-load levels 3 to 4 (dissociation) which contradicted the task-load increase.

- Concerning our *physiological response*, we also found associations or dissociations depending on the considered experimental condition: our results discovered associations between pupil size variations and task demand in both decreasing task-load conditions: pupil size increased linearly with higher task demand. However, considering increasing task-load conditions we found dissociations from task-load levels 1 to 2 (pupil size decreases) and associations from task-load levels 2 to 3 (pupil size increases) in both variable and lineal rate of change conditions. To contrast, from levels 3 to 4, we again found a dissociation in the variable rate of change condition (pupil size decreases), but an association with the linear rate of change condition (pupil size increases).

Therefore, this data partially confirmed our hypothesis that there exists a higher divergence between mental workload measures with variable rate of change conditions: we found performance dissociation only in the increasing task-load with a variable rate of change condition (condition 1). Furthermore, we found physiological dissociations in both variable rate of change conditions from task-levels 1 to 2, but from task-load levels 3 to 4 we found a dissociation only in the increasing variable rate of change condition. Hence, taking into account our results, we could say that there exists a higher divergence between mental workload measures with a variable rate of change condition, particularly in increasing task-load variable rate of change conditions. These results could be explained under the explanation provided by Hancock (2017, page 12) in which he claims that ...[one of the more well-established principles that we do have in the behavioural sciences is that human frequently prove more sensitive to change rather than the absolute level of a stimulus array]...[20]. In other words, some mental workload reflections could be more sensitive to change rather than the absolute level of task demand, whereas others might be more sensitive to the absolute levels of task demand. Therefore, by analyzing our results, subjective perception ratings appeared to be less sensitive to task demand rate of change: they linearly increased in every experimental condition showing no statistical differences among them. Nevertheless, performance measures appeared to be more sensitive to an abrupt increase in task-load level, as the worst performance peak in the increasing variable rate of change condition was achieved in the abrupt transition from task-load levels 2 to 3, even though task demand was higher in task-load level 4. Moreover, physiological measures also appeared to be more sensitive to abrupt changes in task demand only in the increasing task-load condition. Similarly to what happened

with performance measures, a higher pupil size peak was achieved in the abrupt transition from task-load levels 2 to 3, despite task-load being higher in task-load level 4. As pupil dilation reflects activation (among other factors which were controlled), when there is an abrupt increase in task demand, we seem to overreact in order to prepare ourselves to face environmental threats but, in line with resources theories, due to the fact that mental resources are limited and can be depleted, when an abrupt increase is followed by a soft increase in task-load, our organism detects that there is no need to continue activating and it deactivates in order to save resources. On the other hand, higher pupil size in task-load level 1 for both increasing task-load conditions (compared to task-load level 2 in the same condition, and compared to task-load level 1 in the decreasing task-load conditions) could be explained by the fact that participants' activation were higher at the beginning of the experimental session because of the natural nervousness experienced by participants. While in both the decrement task-load conditions, this nervousness activation is added to the activation produced by task-demand level 4, as is reflected in our data: pupil dilation for decreasing task-load conditions were higher in task-load level 4 than for increasing task-load conditions.

These findings should be viewed in light of some study limitations. Although we were able to overcome other studies' limitations, such as the examination of a single direction task demand transition [39], because of the high number of possible combinations, it was not suitable to analyse other interesting experimental conditions (as in for example, low-high-low/high-low-high transitions or changing the intervals in which the abrupt change in task-load demand occurs: beginning, middle, and end of the scenario). Moreover, we think it would be highly interesting to introduce other physiological measures in this study, such as EEG and/or HRV. We must bear in mind that divergences have been found not only between the three primary mental workload measures (performance, physiological, and subjective), but within different indexes of each primary indicator. Lastly, this study has been conducted with students under simulated conditions and we think that it would be appropriate to validate these findings under real conditions in order to improve ecological validity.

Further research is needed to untangle mental workload divergence between measures. Future research could address the aforementioned methodological limitations. For example, it would be interesting to analyse how sensitivity to the rate of change effects varies depending on when the abrupt change takes place or how low-high-low/high-low-high affects mental workload measures' convergence.



## 6. Conclusions

Mental workload is a complex construct which can be measured by its three primary measures: performance, physiological, and subjective. Despite expecting to find convergence between them as they reflect the same construct, dissociations and insensitivities have been repeatedly reported in the literature.

A potential explanation for these divergences could be related to the differential sensitivity of mental workload measures to task-load transitions rate of change: some measures might be more sensitive to change than the absolute level of task demand, while others might be more sensitive to absolute levels of task demand. These differences would result in dissociations and insensitivities, which would ultimately affect convergence between mental workload measures.

Our results suggest that dissociations in performance and physiological pupil size measures may appear after an abrupt change takes place, albeit mostly during increasing task-load conditions. However, subjective ratings may not be affected by the task-load rate of change but by the absolute level of task demand. In other words, our results partially confirmed our hypothesis, as we found higher divergence (dissociations and/or insensitivities) between mental workload measures with abrupt rates of change but only during the increasing task-load condition.

An important implication of this finding is that we should give more weight to one of the mental workload reflections depending on environmental rate of change demands. In other words, if task-load transitions are linear, then we could rely on every primary mental workload measure; but when there is an abrupt task-demand transition from low to high mental workload, we may prefer to rely on subjective ratings rather than physiological or performance measures, as the subsequent decrement in physiological activation (saving resources) would not necessarily mean a decrement in an operator experienced mental workload.

## References

1. Pape, A. M., Wiegmann, D. A., & Shappell, S. A. Air traffic control (ATC) related accidents and incidents: A human factors analysis. (2001).
2. Reuters. What We Know About the Deadly Aeroflot Superjet Crash Landing. (2019, May 6th). Retrieved from: <https://www.themoscowtimes.com/2019/05/06/what-we-know-about-the-deadly-aeroflot-superjet-crash-landing-a65495>

3. Byrne, A. J., Sellen, A. J., & Jones, J. G. Errors on anaesthetic record charts as a measure of anaesthetic performance during simulated critical incidents. *British Journal of Anaesthesia*, **80**(1), 58-62 (1998).
4. Byrne, A. Measurement of mental workload in clinical medicine: a review study. *Anesthesiology and pain medicine*, **1**(2), **90** (2011). doi: 10.5812/kowsar.22287523.2045
5. Young, M. S., & Stanton, N. A. Attention and automation: new perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, **3**(2), 178-194 (2002). <https://doi-org.ezproxy.ub.unimaas.nl/10.1080/14639220210123789>
6. Endsley, M. R. From here to autonomy: lessons learned from human–automation research. *Human factors*, **59**(1), 5-27 (2017). <https://doi-org.ezproxy.ub.unimaas.nl/10.1177%2F0018720816681350>
7. Zoer, I., Ruitenbunrg, M. M., Botje, D., Frings-Dresen, M. H. W., & Sluiter, J. K. The associations between psychosocial workload and mental health complaints in different age groups. *Ergonomics*, **54**(10), 943-952 (2011). doi: 10.1080/00140139.2011.606920.
8. Kawada, T., & Ooya, M. Workload and health complaints in overtime workers: a survey. *Archives of medical research*, **36**(5), 594-597 (2005). <https://doi-org.ezproxy.ub.unimaas.nl/10.1016/j.arcmed.2005.03.048>
9. Cain, B. A review of the mental workload literature. Defence Research and Development Toronto (Canada) (2007).
10. Meshkati, N., & Hancock, P. A. (Eds.). *Human mental workload* (Vol. 52). Elsevier.
11. Moray, N. (Ed.). (2013). *Mental workload: Its theory and measurement* (Vol. 8). Springer Science & Business Media (2011).
12. Kantowitz, B. H. Attention and mental workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 44, No. 21, pp. 3-456) (2000, July). <https://doi-org.ezproxy.ub.unimaas.nl/10.1177%2F154193120004402121>
13. Wickens, C. D. Multiple resources and mental workload. *Human factors*, **50**(3), 449-455 (2008). <https://doi.org/10.1518%2F001872008X288394>
14. Wickens, C. D.: Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, **3**(2), 159-177 (2002). DOI: 10.1518/001872008X288394. 2008 50: 449

15. Munoz-de-Escalon, E. & Cañas, J.: On-line measuring of available resources.H-Workload 2017: The first international symposium on human mental workload, Dublin Institute of Technology, Dublin, Ireland, June 28-30. (2017). DOI:10.21427/D7DK96
16. Muñoz-de-Escalona, E., & Cañas, J. J.. Latency Differences Between Mental Workload Measures in Detecting Workload Changes. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 131-146). Springer, Cham (2018, September).
17. Durantin, G., Gagnon, J. F., Tremblay, S., & Dehais, F. Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural brain research*, **259**, 16-23 (2014). <https://doi.org/10.1016/j.bbr.2013.10.042>
18. Young, M. S., & Stanton, N. A. Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human factors*, **44**(3), 365-375 (2002). <https://doi.org/10.1518%2F0018720024497709>
19. Byrne, A. J., Oliver, M., Bodger, O., Barnett, W. A., Williams, D., Jones, H., & Murphy, A. Novel method of measuring the mental workload of anaesthetists during clinical practice. *British journal of anaesthesia*, **105**(6), 767-771. <https://doi.org/10.1093/bja/aep268>
20. Hancock, P. A. (2017, June). Whither workload? Mapping a path for its future development. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 3-17). Springer, Cham (2010). [https://doi.org/10.1007/978-3-319-61061-0\\_1](https://doi.org/10.1007/978-3-319-61061-0_1)
21. Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, **30**(1), 111-120. <https://doi-org.ezproxy.ub.unimaas.nl/10.1177%2F001872088803000110>
22. Edwards, T., Martin, L., Bienert, N., & Mercer, J. The relationship between workload and performance in air traffic control: exploring the influence of levels of automation and variation in task demand. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 120-139). Springer, Cham (2017, June).
23. Helton, W.S., Shaw, T., Warm, J.S., Matthews, G., Hancock, P.: Effects of warned and unwarned demand transitions on vigilance performance and stress. *Anxiety Stress Coping* 21, 173–184 (2008)
24. Cox-Fuenzalida, L.E.: Effect of workload history on task performance. *Hum. Factors* 49, 277–291 (2007)

25. Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock Jr, J. R. The multi-attribute task battery ii (matb-ii) software for human performance and workload research: A user's guide (2011)
26. Lee, J., & Ahn, J. H. Attention to banner ads and their effectiveness: an eye-tracking approach. *International Journal of Electronic Commerce*, 17(1), 119-137 (2012). <https://doi.org/10.2753/JEC1086-4415170105>
27. Brennan, S.D: An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. DRA Technical Memorandum (CAD5) 92017, DRA Maritime Command and Control Division, Portsmouth (1992)
28. Jordan.C.S. Experimental study of the effect of an instantaneous self assessment workload recorder on task performance. DRA Technical Memorandum (CAD5) 92011. DRA Maritime Command Control Division, Portsmouth (1992)
29. Matthews, G., Middleton, W., Gilmartin, B. y Bullimore, M. A.: Pupillary diameter and cognitive and cognitive load. *J. Psychophysiol.* **5**, 265–271 (1991)
30. Backs, R. W. y Walrath, L. C.: Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Appl. Ergon.* **23**, 243–254 (1992). DOI: 10.1016/0003-6870(92)90152-L
31. Hyönä, J., Tommola, J. y Alaja, A.: Pupil dilation as a measure of processing load in simultaneous interpreting and other language tasks. *Q. J. Exp. Psychol.* **48**, 598–612 (1995). DOI: 10.1080/14640749508401407
32. Granholm, E., Asarnow, R. F., Sarkin, A. J. y Dykes, K. L.: Pupillary responses index cognitive resource limitations. *Psychophysiology*, 33, 457–461 (1996). DOI: 10.1111/j.1469-8986.1996.tb01071.x
33. Iqbal, S. T., Zheng, X. S. y Bailey, B. P.; Task evoked pupillary response to mental workload in human-computer interaction. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1477–1480. ACM, New York (2004). DOI: 10.1145/985921.986094
34. Verney S. P., Granholm, E. y Marshall S. P.: Pupillary responses on the visual backward masking task reflect general cognitive ability. *Int. J. Psychophysiol.* **52**, 23–36 (2004). DOI: 10.1016/j.ijpsycho.2003.12.003
35. Porter, G., Troscianko, T. y Gilchrist, I. D.: Effort during visual search and counting: insights from pupillometry. *Q. J. Exp. Psychol.* **60**, 211–229 (2007). DOI: 10.1080/17470210600673818

36. Privitera, C. M., Renninger, L. W., Carney, T., Klein, S. y Aguilar, M.: Pupil dilation during visual target detection. *J, Vision.* **10**, 1–14 (2010). DOI:10.1167/10.10.3
37. Reiner, M., & Gelfeld, T. M. Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *International Journal of Psychophysiology*, **93**(1), 38-44 (2014).
38. Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S.: Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav. Res. Methods.* **50**(1), 94–106 (2018). DOI: 10.3758/s13428-017-1007-2
39. Morgan, J. F., & Hancock, P. A. The effect of prior task loading on mental workload: An example of hysteresis in driving. *Human factors*, **53**(1), 75-86 (2011).

## **Chapter IV**

---

*Task Demand Transition Peak Point Effects on  
Mental Workload Measures Divergence*



**Task demand transition peak point effects on mental workload  
measures divergence**

Enrique Muñoz-de-Escalona<sup>1</sup>

José Juan Cañas<sup>1</sup>

Chiara Leva<sup>2</sup>

Luca Longo<sup>3</sup>

<sup>1</sup> Mind, Brain and Behaviour Research Centre,  
Department of Experimental Psychology, Faculty of Psychology,  
University of Granada

<sup>2</sup> School of Environmental Health, Technological University Dublin, Kevin Street, Dublin 2,  
Republic of Ireland

<sup>3</sup> ADAPT, The Global Centre of Excellence for Digital Content Technology, Dublin,  
Republic of Ireland

---

This article has been published: Muñoz-de-Escalona, E., Cañas, J. J., Leva, C., & Longo, L. (2020, December). Task Demand Transition Peak Point Effects on Mental Workload Measures Divergence. *In International Symposium on Human Mental Workload: Models and Applications* (pp. 207-226). Springer, Cham. [https://doi.org/10.1007/978-3-030-62302-9\\_13](https://doi.org/10.1007/978-3-030-62302-9_13)

This article received an award for being the best paper presented in the “The 4<sup>th</sup> International Symposium on Human Mental Workload: models and applications (H-Workload 2020)”, held in Granada, Spain, 2020.



## Abstract

The capacity to assess and manage mental workload is becoming more and more relevant in the current work environments as it helps to prevent work related accidents and achieve better efficiency and productivity. Mental workload is often measured indirectly by inferring its effects on performance, mental states, and psychophysiological indexes. Since these three main axes should reflect changes in task demands, convergence between measures is expected, however research has found that this convergence is not to be taken for granted as it is not often present. This study aims to explore how the task demand transition peak point may affect in mental workload divergence between measures during task-load changes: some measures might be more sensitive to abrupt changes in task demand than others and this could also be mediated by the task-load baseline. This was tested by manipulating task-load transitions and the point at which the change in the task load magnitude reaches its highest relative peak over time during a monitoring experiment, while psychophysiological (pupil size) and subjective perceptions were collected as indicators of subjects' workload alongside objective indicators of task performance from the simulation. The results showed that performance measure proved to be sensitive to abrupt increases in task demand in every condition whereas our physiological measure was only sensitive to a sudden increase in task-load during low mental workload baseline circumstances. Furthermore, contrary to what expected, subjective ratings of mental workload did not react to abrupt transitions in task-load in every condition but only to an absolute measure of the overall level of task demand.

**Keywords:** mental workload, workload measures, convergence, divergence, dissociations, insensitivities, task demand transitions, rates of change, peak point.

## 1. Introduction

Mental workload has been the subject of extensive research in many fields over the last several decades. Despite multiple efforts were made for reaching a consensus regarding its definition, to date there is no solid agreement on defining a commonly accepted construct [1-4]. However, despite the difficulties in defining mental workload, the need to assess it in many safety critical workplaces is increasing, particularly in the workplace, for interaction [45,46] and instructional design [47]. Mental workload is closely linked to mental fatigue and performance and so, with human error and work-related accidents [5-10]. On the one hand, a high mental workload situation sustained over time, which is defined as mental overload, increases the chance of making mistakes at work. But also, on the other hand, a very low mental workload can lead to the dangerous and unwanted phenomena defined as “out of the loop” [11, 12]. Hence, assessing mental workload could not only help to predict overload and underload situations and to prevent incidents and fatalities, but also having a good management of mental workload levels would also be advantageous to increase productivity at work [13-15]. There exist several different methodologies for measuring mental workload, but literature research has shown that not all those methodologies do necessarily coincide in providing the same results [16-20]. Empirical research has shown that divergences between measurements exist and, for that reason, the potential causes of these divergences has been the subject of intensive research in recent years [21-26]. There are several different explanations for this lack of convergence between measures. One of these explanations could be the differential effects of “task demand transitions” on mental workload measurements. While a person is performing a task, task demand may increase or decrease. This change in the magnitude of task demand over the duration of a task is what we call a “task demand transition”, which could be smooth or abrupt.

The aim of this paper is to shed some light about mental workload divergences that might be produced by task demand transitions. Particularly, we want to explore how the task demand transition peak point (the moment in which the abrupt increase in task demand occurs) may affect in mental workload divergence between measures during task-load changes. In other words, it may be possible that task demand transition peak point may affect more and in different ways some mental workload measures than others so that it would facilitate the emergence of divergence between them. In Section 2 we will briefly review the most important explanations that have been suggested for this lack of

convergence between measures, as well as certain key concepts about measuring mental workload. Then, in Section 3 we present the design and methodology used to undertake our study. In Section 4 we describe the results obtained from the data collection campaign, while in Section 5 we discuss our findings and suggest possible future directions in this research area. Finally, in Section 6 we provide an overview of the key conclusions derived from the study, as well as its major implications.

## 2. Related Work

### 2.1 Assessing Mental Workload

Assessing mental workload is not a simple task since it is not something that can be measured directly and it has to be inferred from its different effects on behavior, mental states, and psychophysiological indexes. Mental workload has three well-established components, two of which are observable (performance and physiological) and one that is a non-observable subjective feeling of workload. In this sense there are three different categories of measures which are commonly used to assess mental workload:

- 1. Performance measures:** this category is composed of primary task measures (n° of errors, reaction time) and secondary task measures (choice reaction-time tasks, time estimation, memory-search tasks). The goal is to measure objective task performance indexes with a view to assess the quantity and the quality of performed tasks.
- 2. Physiological measures:** it includes physiological responses to mental activation (pupil diameter, heart rate variability (HRV), electroencephalogram (EEG)). The aim is to collect objective physiological data, which reacts to mental workload.
- 3. Subjective measures:** comprises self-reported measures (NASA-TLX) which aim is to collect easy and low-cost subjective individual-related mental workload data.

Each different category has its own advantages and disadvantages. Performance measures, for example, have a high diagnostic value, which allows to investigate which cognitive process is mainly involved in performing certain tasks; but it comes with the

disadvantage of a high level of intrusiveness, as operators have to perform simultaneous tasks and that can be dangerous in certain contexts [27]. Physiological measures, on the other hand, are very sensitive to mental workload phasic changes and have high internal validity but are also highly intrusive and require special equipment and specific expert knowledge [28]. Finally, subjective self-reporting measures are very low cost and easy to implement but are vulnerable to cognitive biases [29]. In general it is safe to say that the ideal way of assessing mental workload would involve a triangulation of at least one of each different type of measures, so as to obtain a performance measure, a physiological measure and a subjective measure in such a way that each would contribute extra information and reinforce the overall evaluation. In line with this assumption, we could expect that each different measure of mental workload should converge, but the literature shows multiple examples of situation where this assumption could not be verified [16-19,21,23].

## **2.2 Convergences and Divergences Between Mental Workload Measurements**

Far from what could be expected, the different indexes of mental workload, gathered by using different categories of measures, do not necessarily coincide in providing the same results. Literature research has increasingly shown studies that offer evidence about this lack of correlation among measurements (divergences) [16-19,21,23], which could be due to the occurrence of dissociations and/or insensitivities in mental workload measures. Associations occur when mental workload measures tracks task-load change, dissociations take place when a mental workload reflection contradict a change in task demands whereas insensitivities occurs when that workload reflection does not change with a change in task demands. Research has shown that this lack of convergence among different measurements due to dissociations or insensitivities do not only occurs between the three different categories of methodologies, but also between different particular methods within each category. For example, numerous studies have shown proof of divergences among well-established physiological measures which are supposed to reflect the level mental workload experienced and its changes in the same way [20-21]. For example, pupil size and blink rate have been proved to be ocular parameters sensitive to task demand changes, but sometimes both can be affected by certain non-workload factors (such as brightness and relative humidity), that would impact on the emergence of dissociations among measures. One can only wonder why do we find divergences

between measurements that are supposed to be reflecting the same construct? Researchers have identified multiple possible causes about this phenomenon, which can be related to a lack of specificity or to a lack of diagnosticity. The former would explain divergences by the existence of non-workload factors affecting measurements, while the latter might point to the fact that different measures could be reflecting different aspects of mental workload. Mental workload is a multidimensional construct and these lack of specificity and diagnosticity would be consistent with multi-resources theories [30]. Hancock (2017) has offered many reasons about the occurrence of dissociations and insensitivities among mental workload measures, which he defined as the AID's of workload [22]. Timescale considered between measures is one of the possible causes: it may be possible that some measures reflect task demand changes within seconds while other may show a longer latency. Muñoz-de-Escalona & Cañas (2018) found that divergences among measures are more susceptible to appear after an abrupt change in task demands occurs, due to the emergence of higher latency differences between measurements. They found that the subjective measure of mental workload reacted sooner than pupil size (physiological response) to a high peak in task demands. In other words, divergences between subjective and physiological measures were higher because subjective measure may have reached its maximum value before the physiological measures did. [23]

### **2.3 Sensitivity to Rates of Change and Peak Point During Task Demand Transitions**

Hancock (2017) points out in his study that another possible factor underpinning dissociations and insensitivities among assessment methods is the differential sensitivity of mental workload measures to rates of change during task-load transitions [22]. In other words, some measures might be more sensitive to change in task demand than others, so that when there abrupt shifts in task-load are experienced, divergences will also appear across different measures due to their different sensitivities towards the rate of change. That is precisely what was found in a previous study showing higher divergences between mental workload measures during variable rate of change conditions than during lineal rate of change conditions [23]. According to the study by Muñoz-de-Escalona & Cañas (2019) [23] dissociations in performance and pupil size (physiological measure) appeared after an abrupt change in task demands during increasing task-load conditions, while the subjective measure was not affected by the rate of change in the task-load but rather by the absolute level of task demand experienced. These interesting results revealed that

divergences between measures might be higher during increasing variable rate of change conditions. However, this study did not address the possible effects that the task demand transition peak point may have in the emergence of divergence between mental workload measures: it may be possible that some measures are more sensitive to abrupt changes in task-load levels than others and, furthermore, it may also be possible that some measures might be more sensitive to task-load change under certain circumstances. Therefore, this motivated the purpose of this study, to explore the possible effects that the peak point during task demand transitions may have in mental workload divergences between measures.

### 3. Design and Methodology

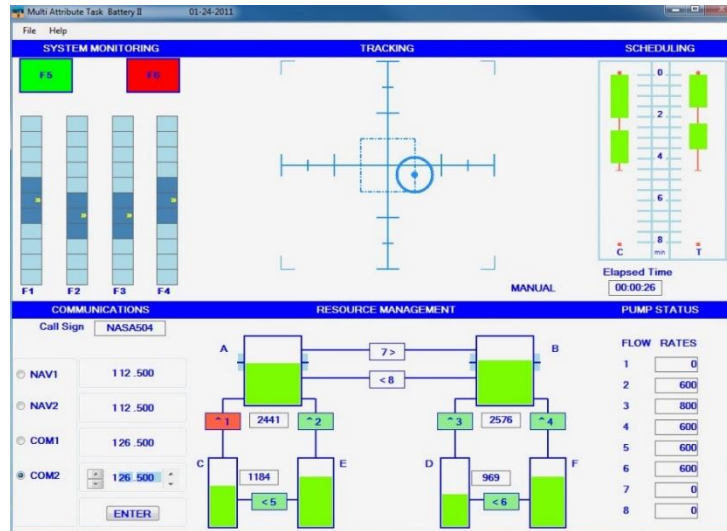
The hypothesis of this research is that the effects on mental workload measures will be different depending on when the high shift in task-load arises. When task demands are low, there are many resources available to be used so, if there is a sudden shift in task-load, the amount of mental resources mobilized to cope with the task will be higher than if we depart from a higher task demand situation, since in this latter case, human resources are limited. Thus, in more detail, we hypothesized that we will find higher divergences between mental workload measurements when the high shift in task-load arises in a low task demand situation because physiological measures should react sooner than performance and subjective measures, as they reflect that extra activation needed to cope with the task. In this study we manipulated 2 variables: (1) task-load change and (2) task-load change peak point. This manipulation was tested in a task-battery experiment which participants performed for 20 minutes to collect (a) perceived mental workload, (b) task performance and (c) pupil size as our three types of complementary mental workload primary measures.

#### 3.1 Materials and Instruments

**The MATB-II Software.** Measurements of task performance were collected through the use of the second version of the Multi-Attribute Task Battery (MATB-II), a computer program designed to evaluate operator performance and workload through means of different tasks similar to those carried out by flight crews, with a user-friendly interface as to allow non-pilot participants to utilize it [31]. MATB-II comes with default event

files, which can easily be altered to adapt to the needs or objectives of an experiment. The program records events presented to participants, as well as participants' responses. The MATB-II contains the following four tasks: the system monitoring task (SYSMON), the tracking task (TRACK), the communications task (COMM), and the resource management task (RESMAN) (see Figure 1).

- 1) The SYSMON task is divided into two sub-tasks: lights and scales. For the lights sub-task, participants are required to respond as fast as possible to a green light that turns off and a red light that turns on, and to turn them back on and off, respectively. For the scale sub-task, participants are required to detect when the lights on four moving scales deviate from their normal position and respond accordingly by clicking on the deviated scale.
- 2) In the TRACK task, during manual mode, participants are required to keep a circular target in the center of an inner box displayed on the program by using a joystick with their left hand (the dominant hand was needed for the use of the mouse). During automatic mode, the circular target will remain in the inner box by itself.
- 3) In the COMM task, an audio message is played with a specific call sign and the participant is required to respond by selecting the appropriate radio and adjusting for the correct frequency, but only if the call sign matches their own (call sign: "NASA504"). No response is required of the participant for messages from other call signs.
- 4) In the RESMAN task, participants are required to maintain the level of fuel in tanks A and B, within  $\pm 500$  units of the initial condition of 2500 units each. In order to maintain this objective, participants must transfer fuel from supply tanks to A and B or transfer fuel between the two tanks.



**Figure 1:** MATB-II task display. Taken from <https://matb.larc.nasa.gov/>

**The use of Tobii T120 Eyetracker.** Pupil diameter measurements were obtained using an infrared-based eye tracker system, the Tobii T120 model marketed by Tobii Video System (see Figure 2). This system is characterized by its high sampling frequency (120 Hz). This equipment is completely non-intrusive, has no visible eye movement monitoring system, and provides high precision and an excellent head compensatory movement mechanism, which ensures high-quality data collection. In addition, a calibration procedure is completed within seconds, and the freedom of movement it offers participants allows them to act naturally in front of the screen, as though it were an ordinary computer display. Thus, the equipment allows for natural conditions in which to measure eye-tracking data [32].



**Figure 2:** Tobii T120 Eyetracker system





### 3.3 Procedure

Participants went through an experimental session consisting of two phases:

1. **Training stage:** a 30 minutes training session allowed each participant to familiarize themselves with the program so that they could carry out the tasks securely during the data collection stage. The procedure was conducted as follows: upon entering the lab and after filling out the informed consent form, the participant was asked to read the MATB-II instruction manual. The researcher then sat down with the participant to allow for questions and resolve any doubts on how to use the program. Afterward, on a computer monitor, participants were presented with each MATB-II task separately and were first given a demonstration as to how to execute the task. Following this they were given some time (3 minutes or more if needed) to try and perform the task themselves. Participants would then evaluate task difficulty on a scale ranging from 1 (very easy) to 9 (very difficult). The average difficulties evaluated by the participants for each task of the MATB-II are reported in Table 1.

**Table 1:** Task Difficulty Level Average Assessed by Participants.

Task	Difficulty Average
TRACK	4.16 ( $\sigma$ 1.69)
SYSMON	2.71 ( $\sigma$ 1.39)
COMM	3.33 ( $\sigma$ 1.97)
RESMAN	5.11 ( $\sigma$ 1.95)

The participants were always free to consult the manual and ask the researcher questions during the training stage in case of doubts or uncertainties. Once each participant had completed all four tasks and resolved all doubts, they were ready for the data collection stage, which followed immediately afterwards. During the training stage, participants could work in one room equipped for training with the MATB-II software, and no performance shaping factors was recorded in relation to the environmental conditions as we ensure to have similar average luminosity and comfort for temperature and seating arrangements.

2. **Data collection stage:** the data collection stage lasted a period of 20 minutes for each participant and it was divided into four intervals of 5 minutes. In each interval participants had to perform one combination of tasks according to the assigned experimental conditions, while task performance, perceived mental workload, and pupil diameter were recorded. The participants were instructed to fill in the ISA scale every 2,5 minutes when a scheduled alarm sounded. Prior to the start of the MATB-II tasks, the eye-tracker system was calibrated for each participant while he or she was instructed to keep head and body movements to a minimum. During the data collection stage ensuring conformity with standard room conditions was essential to avoid external performance shaping factors to interfere with the results of the experiment. Thus, the testing rooms were temperature controlled to 21°C, and lighting conditions (the main extraneous variable in pupil diameter measurement) were kept constant with artificial lighting; there was no natural light in the rooms. Moreover, participants always sat in the same place, a comfortable chair spaced 60 cm from the eye-tracker system.

This study was carried out in accordance with the recommendations of the local ethical guidelines of the committee of the University of Granada institution: Comité de Ética de Investigación Humana. The protocol was approved by the Comité de Ética de Investigación Humana under the code: 779/CEIH/2019. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

### 3.4 Variables

**Independent Variables** - We manipulated 2 independent variables:

1. **Task-load change:** this variable describes how task demand changes over time. Experimental conditions followed an increasing task demand pattern with a variable rate of change over time.
2. **Task-load rate of change peak point:** this variable describes when the participants experienced the most abrupt change in task demands.

First, we manipulated task-load by modifying the number and the combination of MATB-II active tasks that participants had to perform over time. We created six combination of tasks. Each combination of task made up a condition of increasing difficulty. Table 2 reports the task combinations considered. Task-load increases as the number of active task increases, but also as the relative difficulty of active tasks increases as well. Therefore, since task-load in each individual task was different, the overall task-load in each task combination allowed us to manipulate the task-load change peak point during the experimental session. For example, task combinations 2 and 3 both include 2 tasks, but task-load is higher in task combination 3 as the RESMAN task is more difficult than SYSMON task. Similarly, this is also true for task combinations 4 and 5 (both includes three tasks but task combination 5 is more difficult because it included the RESMAN tasks instead of the COMM task).

**Table 2:** Possible Sets of Task Combinations.

<b>Task Combination</b>	<b>MATB-II Active tasks</b>
1	TRACK
2	TRACK + SYSMON
3	TRACK + RESMAN
4	TRACK + SYSMON + COMM
5	TRACK + SYSMON + RESMAN
6	TRACK + SYSMON + COMM + RESSMAN

Then, to increase the task-load along the four intervals of five minutes during the experimental session, we created three experimental conditions each one corresponding to a different choice for the temporary moment in which the magnitude of task-load changes reaching its highest relative peak over time:

- in combination (1) the steepest change in task load is given during early stages of the overall task-load condition

- in combination (2) the steepest change in task load is given during medium stages of the overall task-load condition
- in combination (3) the steepest change in task load is given during the later stages of the overall task-load condition.

The three experimental conditions were:

**1. Condition 1: Highest rate of change during early stages of overall task load.**

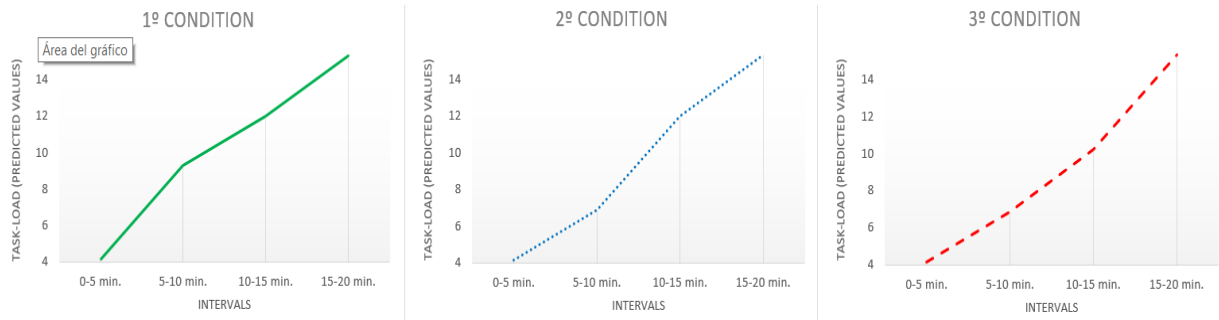
Task-load increased every 5 minutes, with a variable rate of change. Participants had to perform the following task combination: 1, 3, 5, and 6. The initial task-load in the first interval (task combination 1) was followed by an abrupt increase in the second interval (task combination 3) and from there, there was a decrease in the rate of change from task combination 3 moving onto 5 and then 6.

**2. Condition 2: Highest rate of change during medium stages of overall task load.**

Task-load increased every 5 minutes, with a variable rate of change. Participants had to perform the following task combination: 1, 2, 5, and 6. The initial task-load of the first interval (task set 1) is followed by a slight task-load increase in the second interval (task set 2), then we find an abrupt increase in task-load rate of change where the participants were moved to task combination 5 and finally a slight decrease in the rate of change for task combination 6.

**3. Condition 3: Highest rate of change during later stages of overall task load.**

Task-load increased every 5 minutes, with a variable rate of change. Participants had to perform the following task combination: 1, 2, 4, and 6. The initial task-load of the first interval (task set 1) is followed by a slight task-load increase moving into task set 2, followed by another slight change (moving into task combination 4) and finally we find an abrupt increase in task-load rate of change moving into task combination 6.



**Figure 4:** Task-load rate of change evolution (current task-load) for the three experimental conditions.

By comparing these three conditions, we were able to measure the effect of task-load rate of change. A 3x4 mixed factorial experimental design was devised with 2 independent variables. One variable was task-load rate of change with 3 different levels that were manipulated between groups. The other variable was the interval when task-load changed, this variable had 4 levels and was manipulated within-subjects.

### Dependent Variables -

*Performance.* MATB-II records many indicators of participants' performance: root mean square deviation (RMSD) for the TRACK task, number of correct and incorrect responses for the SYSMON and COMM tasks, and the arithmetic mean of tanks "A-2500" and "B-2500" in absolute values for the RESMAN task. However, for the purposes of this experiment we will only consider the RMSD performance indicator, as it is the only task present during all task combinations in every experimental condition, allowing us to compare participants' performances through and between conditions. The SRMSD performance indicator reflects the distance of the circle to the target point, so that a higher score on this variable reflects a negative performance.

*Pupil size.* Mental workload can be revealed by several physiological indexes such as EEG, HVR, and several ocular metrics. We decided to use pupil diameter as our physiological mental workload indicator, as it has been verified in the literature as an objective indicator for mental workload [35-43] and minimizes intrusiveness. While the eye-tracking system used allows a continuous sampling rate recording at 120Hz, a total of 8 intervals lasting 2.5 minutes was set, so as to obtain 2 measures in each of the three

intervals of different overall task-load levels. Since expressing pupil size in absolute values has the disadvantage of being affected by slow random fluctuations in pupil size (source of noise), we followed the recommendations provided by Sebastiaan Mathôt [44] regarding the baseline correction of pupil-size data. To do this, for every participant, we took his/her average pupil size during the session as a whole as a reference, which was then subtracted from the obtained value in each of the 8 intervals, thereby giving a differential standardized value that allowed us to reduce noise in the collected data. Analyses were carried out for the average of both the left and right pupils. A negative value meant that the pupil was contracting while a positive value meant that it was dilating.

*Subjective mental workload.* Traditional off-line subjective workload assessment tools, such as the NASA Task load Index (NASA-TLX), do not allow researchers to obtain continuous subjective ratings from participants. In order to facilitate and establish comparisons between mental workload measures, it was necessary to obtain the subjective momentary ratings continuously throughout the experimental session. With this goal, we used the ISA, which is an on-line subjective workload scale created for this purpose. Ratings were obtained at 2.5-minute intervals throughout the 20 minutes of the experiment duration, obtaining a total of 8 subjective mental workload ratings (2 measures in each of the three intervals of different overall task-load levels).

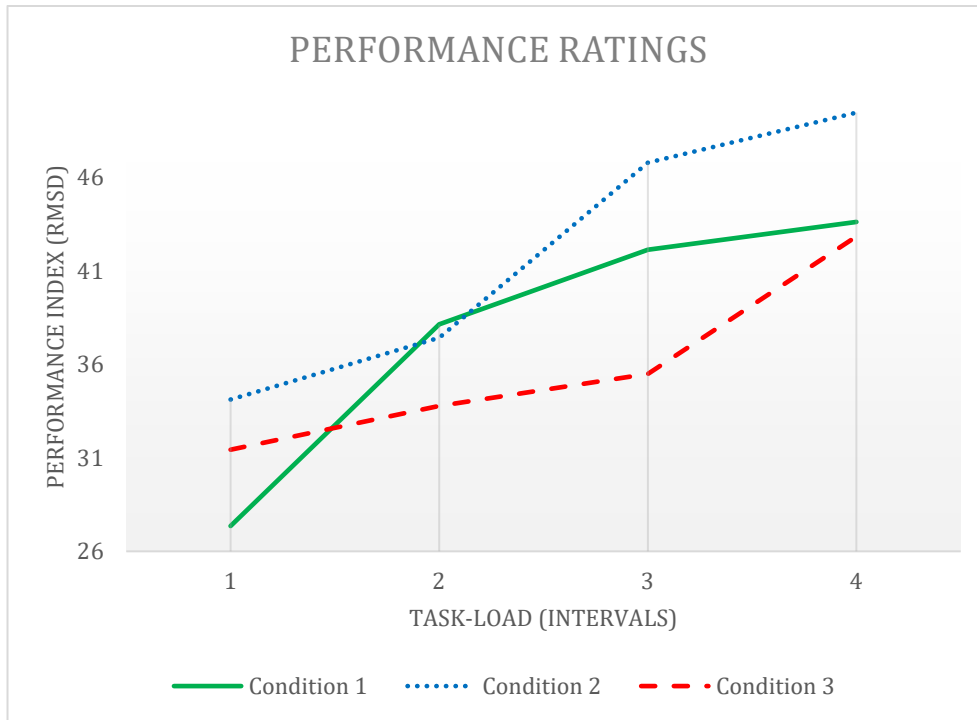
*Synchronization of measures.* Performance, pupil size, and subjective measures were obtained continuously throughout the experimental session. Synchronization between measures was simple, as the eyetracker and MATB-II performance log files began to record data simultaneously at the start of the experimental session. The scheduled alarm (every 2.5 minutes) was also synchronized by the experimenter, as it was simultaneously activated with the MATB-II software. This allowed the ISA scale to be synchronized with the performance and pupil size measures as well.

## 4. Results

To analyze collected results, we performed three one-way within-subjects ANOVA, one for each different workload measure. First, the ANOVA analysis for our performance measure revealed a significant task-load change main effect  $F(3,126) = 42.53$ ,  $MSe =$

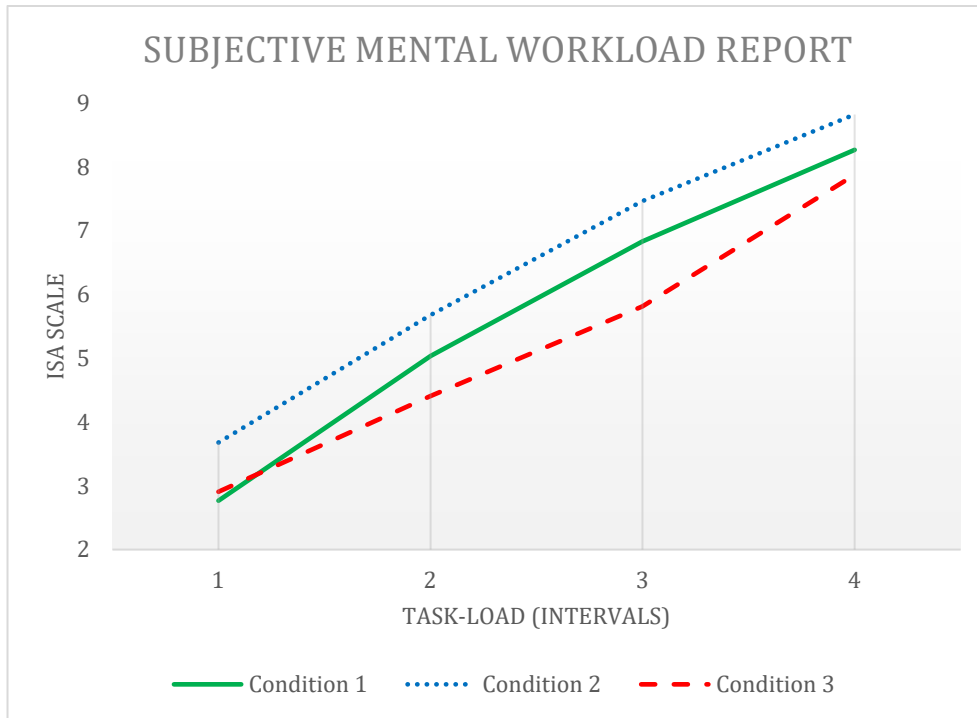
40.89,  $p < .01$ , which means that participants' performance was impaired with the increase in task-load through intervals. On the other hand, although the main effect of the experimental condition was not significant  $F(2,42) = 1.33$ ,  $MSe = 427.5$ ,  $p > .05$ , the interaction effect task-load level by experimental condition was found to be significant  $F(6,126) = 2.99$ ,  $MSe = 40.89$ ,  $p < .01$ . This interaction was due to the different evolution of performance impairment through intervals depending on different experimental condition: in experimental condition 1, the greatest difficulties in performance normally occurred between intervals 1 to 2; in experimental condition 2 between intervals 2 to 3 and in experimental condition 3 between intervals 3 to 4. In order to confirm the above statements, we performed 3 mixed factorial ANOVAs, with experimental condition as between group factor, and a one pair of intervals (1 versus 2, 2 versus 3, and 3 versus 4) as the within-subject factor. The first partial ANOVA (int. 1 versus 2) showed a significant task-load change main effect  $F(1,42) = 16.14$ ,  $MSe = 41.87$ ,  $p < .01$ , a non-significant experimental condition main effect  $F(2,42) = .65$ ,  $MSe = 142.34$ ,  $p > .05$ , but also a significant interaction effect of task-load by experimental condition  $F(2,42) = 3.87$ ,  $MSe = 41.87$ ,  $p < .05$ . This interaction effect was explained by the abrupt impairment in task performance in experimental condition 1 that occurred in the second interval. This impairment evolved in a smoother way in experimental conditions 2 and 3. The second partial ANOVA (int. 2 versus 3), revealed a significant task-load main effect  $F(1,42) = 16.19$ ,  $MSe = 34.93$ ,  $p < .01$ , as well as a non-significant experimental condition main effect  $F(2,42) = 1.70$ ,  $MSe = 270.50$ ,  $p > .05$ , but a significant interaction effect task-load change by experimental condition  $F(2,42) = 3.26$ ,  $MSe = 34.93$ ,  $p < .05$ . These results were probably caused by the issues experienced by participants with performance recorded for experimental condition 2. Participants were better able to manage their performance in experimental conditions 1 and 3. In the third partial ANOVA (int. 3 versus 4), task-load change main effect was found to be significant  $F(1,42) = 13.73$ ,  $MSe = 24.13$ ,  $p < .01$ , experimental condition main effect was not found to be significant  $F(2,42) = 3.10$ ,  $MSe = 24.13$ ,  $p > .05$ , but again we found an interaction effect task-load change by experimental condition  $F(2,42) = 3.10$ ,  $MSe = 24.13$ ,  $p = .05$ . In this case participants experienced more difficulty with task load change within experimental condition 3, while in conditions 1 and 2 the impairment was smoother. Therefore, these data confirmed that the greatest performance impairment occurred when there was an abrupt increment in task-load.





**Figure 5:** Participants' performance during task development.

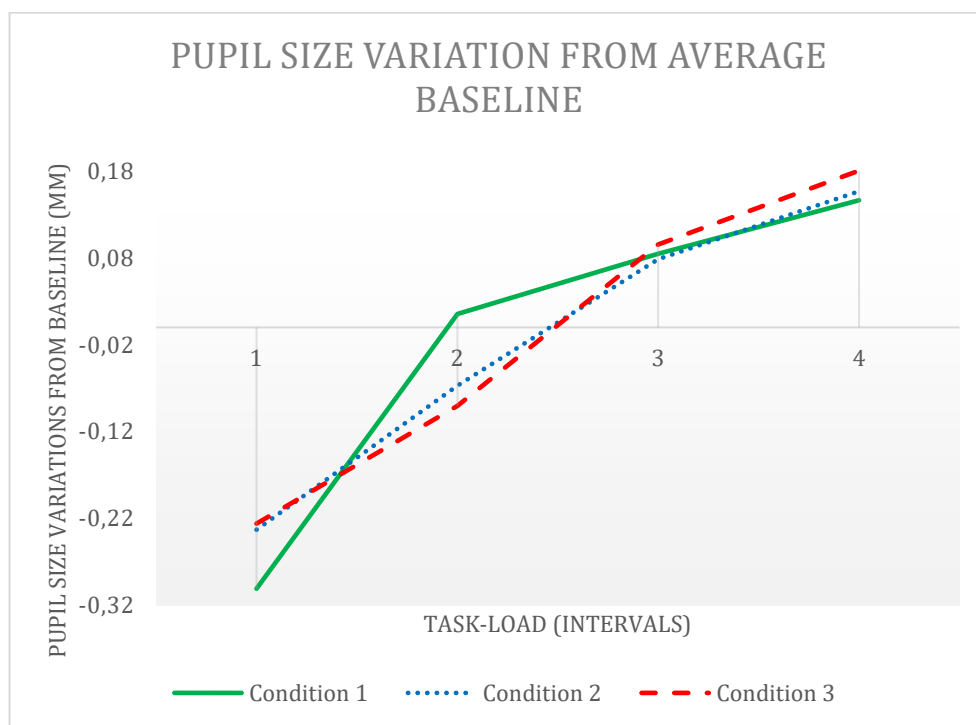
With respect to subjective perceptions, we found a linear increase in the subjective measure of mental workload in every experimental condition. ANOVAs' analysis revealed a significant main effect of task-load  $F(3,126) = 242.5$ ;  $MSe = .92$ ,  $p < .01$ . However, the overall task load effect of the experimental condition was not found to be significant  $F(2,42) = 3.19$ ;  $MSe = 6.33$ ,  $p > .05$ , nor the interaction effect of task-load level by experimental condition,  $F(6,126) = 1.15$ ;  $MSe = .92$ ,  $p > .05$ . Trend analysis confirmed this linear trend of task-load change effect  $F(1,42) = 339.26$ ;  $MSe = 1.97$ ,  $p < .01$  which means that there existed the same lineal increase in mental workload subjective perceptions through intervals in every experimental condition.



**Figure 6:** Participants' subjective mental workload ratings during task development.

Finally, regarding our physiological variable, we found a significant effect for the overall task load  $F(3,126) = 198.81$ ,  $MSe = .007$ ,  $p < .01$ . And, although the main effect of the experimental condition was not found to be significant  $F(2,42) = .90$ ,  $MSe = .001$ ,  $p > .05$ , the interaction effect of task-load change by experimental condition was found to be significant,  $F(6,126) = 3.09$ ,  $MSe = .007$ ,  $p < .01$ . As we can see in the graph, pupil size increased throughout the various intervals in every experimental condition. However, whereas in experimental condition 1 we can observe an abrupt increase in pupil size from interval 1 to 2, and then a slight increase from interval 2 to 4, in experimental conditions 2 and 3 we did not find any sudden abrupt dilation change within intervals. These observations were supported by partial ANOVA analyses. A 3-mixed factorial ANOVAs was performed, with experimental condition as between group factor, and one pair of intervals as the within-subject factor. The first partial ANOVA (int. 1 versus 2) revealed a significant task-load change main effect  $F(1,42) = 176.61$ ,  $MSe = .005$ ,  $p < .01$ , a non-significant experimental condition main effect  $F(2,42) = .19$ ,  $MSe = .009$ ,  $p > .05$ , but also a significant interaction effect of task-load by experimental condition  $F(2,42) = 11.77$ ,  $MSe = .005$ ,  $p < .01$ . This interaction effect could be explained by the abrupt pupil size increase in experimental condition 1, while in the other groups, pupil size increased in a smoother and parallel way. In the second partial ANOVA (int. 2 versus 3), we found a

significant task-load main effect  $F(1,42) = 59.29$ ,  $MSe = .007$ ,  $p < .01$ , a significant experimental condition main effect  $F(2,42) = 3.46$ ,  $MSe = .007$ ,  $p < .05$  and a significant interaction effect task-load x experimental condition  $F(2,42) = 3.46$ ,  $MSe = .007$ ,  $p < .05$ . This interaction effect was due again to the different course of pupil size in experimental condition 1 with regard to experimental conditions 2 and 3, that is, pupil size increment is much slower in condition 1 than in conditions 2 and 3 which both follow a similar course again. Finally, the third partial ANOVA (int. 3 versus 4) showed a significant task-load change main effect  $F(1,42) = 71.17$ ,  $MSe = .002$ ,  $p < .01$ , a non-significant experimental condition main effect  $F(2,42) = .55$ ,  $MSe = .002$ ,  $p > .05$ , but in this case a non-significant interaction effect of task-load by experimental condition  $F(2,42) = .55$ ,  $MSe = .002$ ,  $p > .05$ . This non-significant interaction could be explained by the absence of significant differences between experimental conditions regarding pupil dilation.



**Figure 7:** Participants' pupil size variation ratings from the average baseline during task development.

## 5. Discussion

In summary the results of the experiment showed that performance was negatively affected by overall task-load increased in every experimental condition, but we found that the participants experienced their greatest difficulty with performance in correspondence

with the highest rate of change for the task load in each experimental condition. On the other hand, our physiological index, pupil size, also increased as task-load increased in every experimental condition but was not significantly affected by the rate of change in task load with the exception of the results recorded for experimental condition 1, where the highest dilation change matches the task-load change peak point. Finally, regarding subjective perception of mental workload, we did not find any differences between experimental conditions, as it increases linearly through intervals in every condition.

Hence, considering our results, we found a triple association, which means that the three measures of mental workload reflected this construct successfully. However, it is also true that we found differences in the way each measure evolved through intervals:

- Our **performance index** reflected successfully task demand increase through intervals, but it also appeared to be sensitive to abrupt changes in task demands independently of the timing when this abrupt change took place. In other words, our performance measure seemed to be sensitive to abrupt changes at all times, that is, regardless of the current level of task demand.
- Our **subjective index**, on the other hand, seemed to be sensitive to task-load change, but was not sensitive to task-load rate of change regardless of the moment in which the abrupt change in task demand took place. We can appreciate in our results that subjective perceptions of mental workload increased linearly in every experimental condition with no abrupt changes through intervals. In other words, our subjective measure of mental workload seemed to be more sensitive to the absolute level of task demand than to task-load rate of change.
- Regarding our **physiological index**, our results showed that it was sensitive to task-load change as pupil size dilated through intervals. However, the key finding of this research is that pupil size index was also sensitive to abrupt changes in task demand but only when the abrupt change took place in a low task demand situation (experimental condition 1): we can see that during the “early task-load change peak condition” pupil size greatly increased from interval 1 to 2, that is, when the abrupt change in task-load occurs.

Considering the evolution of the different measures explained above, and in line with what Muñoz-de-Escalona & Cañas (2019) found in their previous research, we have seen that some measures are more sensitive to the absolute level of task demands regardless of task-load rate of change and some others, are more sensitive to abrupt changes in task-

load levels. However, it was also found that pupil dilation might be able to reflect abrupt changes in task demand under low task-load circumstances. Therefore, this results confirmed our hypothesis that there would be higher divergences between mental workload measures when an abrupt increase in task-load occurs during low mental workload situations: pupil size measure overreacted to that sudden change in task-load, while performance impairment also increased but not that sharply and subjective perceptions did it in a much smoother way. However, when the abrupt increase in task demand occurred with medium and high mental workload values, pupil size did not increase abruptly, while performance kept showing the effect of the abrupt change in task demands. And subjective perception of mental workload increased as the absolute value of task load increased but was not affected by any abrupt change in task demand. These results can be explained under resource theories. When mental workload is low, there are plenty of mental resources available to provide to cope with an abrupt increase of task demand. However, when workload level is higher, the absolute pool of mental resources could be running out and there would not be many more resources available to be assigned to the task to cope with the abrupt change in task demands. Therefore, human organism can mobilize much more resources in low task demand situations than in higher task demand situations in which the organism will try to save resources as they are more limited and can be depleted. This effect of the abrupt change in task demands is well reflected by the performance of the participants in the task. However, an additional factor we must consider is the fact that the pupil is a muscle and it has a dilation range, so that, when it approximates its limit it will not be able to continue increasing and that would also contribute to the way it is able or not to reflect changes in mental workload. Hence, this could be a contributing reason why pupil size only was able to reflect the abrupt change in task demand during low mental workload condition.

Finally, we again observed that the subjective feeling of mental workload is independent of some factors that affect performance and psychophysiological parameters of mental resources. People are more sensitive to absolute increases of task demand, as shown in other studies of mental workload and fatigue [24, 25]. This experiment however has some limitations that should be noted as the study was performed with a restricted number of students and we consider that this results should be validated with a more representative and numerous sample of the population in order to increase external validation. On the other hand, it has been repeatedly reported in the literature that

divergences have been found not only between the different types of methodologies for measuring mental workload, but also within different measures of each methodology. It would have been interesting, for example, to register multiple physiological measures, which reflects mental activation, in order to analyze convergences and divergences within measures. In addition, it would have been interesting to gather more neurophysiological signals to support obtained findings about activation changes through conditions. Mental workload divergence between measures is a problem which must be addressed with further research, as the factors affecting this phenomenon are multiple and difficult to interpret. Future research could also consider age effects in workload measure divergences which given the nature of the current sample was not significantly observed.

## 6. Conclusions

Mental workload assessment has been the subject of numerous studies in the last several years. Quantifying mental workload not only helps to prevent work-related accidents, but also increases productivity at work, so there is an increasing need in our modern society to measure a construct that even, to date, is very difficult to be defined. There exist three main axes for measuring mental workload: performance, physiological and subjective primary measures. Although it is natural to think that mental workload measures of the same person in the same situation should converge, previous studies have proven that divergences between mental workload measures are very frequent. This study aimed to shed some light about one possible cause for mental workload divergences. Particularly, we wanted to explore how abrupt increases in task demand would affect mental workload measures divergences, depending on the baseline task-load situation. Our findings suggest that although the three considered measures of mental workload were sensitive to the construct, there are differences that we must consider between the way each different measure reacts to sudden increases in mental workload. Firstly, performance measure proved to be sensitive to abrupt increases in task demand in every condition whereas our physiological measure (pupil size) was only sensitive to a sudden increase in task-load during low mental workload baseline circumstances. While subjective ratings of mental workload did not react to abrupt transitions in task-load in any experimental condition but only to the absolute increase in task load.

An important implication about this finding is that we must be cautious in selecting the methodology for assessing mental workload: the decision should be based on our particular goals, as each measure may be more suitable to represent different aspects of mental workload. For example, if we need to detect sudden changes in mental workload in every condition, it could be a better idea to trust performance index, however, if the goal is to detect sudden increases in mental workload during low mental workload sustained situations (certain train drivers) it would be a better idea to trust pupil size, as it seems to be very sensitive to abrupt increase in task-load in this particular situation. Finally, subjective perceptions of mental workload is sufficiently accurate to give an overall idea about the state in which the person finds himself/herself. In any case it is to be assumed that the best way to assess mental workload involves a triangulation of different types of measures, one for each different kind of methodologies, as every different index have its pros and cons and they all contribute relevant information. Last but not least we must be cautious about these findings, as it was obtained for a limited sample therefore further research is needed to continue clarifying the topic of mental workload assessment.

## References

1. Babiloni, F. (2019, November). Mental Workload Monitoring: New Perspectives from Neuroscience. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 3-19). Springer, Cham
2. Cain, B. A review of the mental workload literature. Defence Research and Development Toronto (Canada) (2007)
3. Meshkati, N., & Hancock, P. A. (Eds.). Human mental workload (Vol. 52). Elsevier
4. Moray, N. (Ed.). (2013). Mental workload: Its theory and measurement (Vol. 8). Springer Science & Business Media (2011)
5. Josten, E. J., Ng-A-Tham, J. E., & Thierry, H.: The effects of extended workdays on fatigue, health, performance and satisfaction in nursing. *J. of advanced nursing*, **44**(6), 643-652 (2003). DOI: 10.1046/j.0309-2402.2003.02854.x
6. Taylor, A. H., & Dorn, L.: Stress, fatigue, health, and risk of road traffic accidents among professional drivers: the contribution of physical inactivity. *Annu. Rev. Public Health*, **27**, 371-391 (2006). DOI: 10.1146/annurev.publhealth.27.021405.102117

7. Lilley, R., Feyer, A. M., Kirk, P., & Gander, P. (2002). A survey of forest workers in New Zealand: Do hours of work, rest, and recovery play a role in accidents and injury?. *Journal of safety research*, 33(1), 53-71
8. Sarsangi, V., Salehiniya, H., Hannani, M., Marzaleh, M. A., Abadi, Y. S., Honarjoo, F. & Derakhshanjazari, M. (2017). Assessment of workload effect on nursing occupational accidents in hospitals of Kashan, Iran. *Biomedical Research and Therapy*, 4(8), 1527-1540
9. Rodgers, S. H. (2000). An Ergonomic Approach to Analyzing Workplace Accidents. *Applied occupational and environmental hygiene*, 15(7), 529-534
10. Kirschenbaum, A., Oigenblick, L., & Goldberg, A. I. (2000). Well being, work environment and work accidents. *Social Science & Medicine*, 50(5), 631-639
11. Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human factors*, 37(2), 381-394
12. Young, M. S., & Stanton, N. A. Attention and automation: new perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, 3(2), 178-194 (2002). <https://doi-org.ezproxy.ub.unimaas.nl/10.1080/14639220210123789>
13. Priya, D. S., Johnson, P., Padmavathi, R., Subhashini, A. S., Ayyappan, R., & Ayyappan, M. (2010). Evaluation of the relationship between workload and work capacity in petrochemical and tannery workers-A pilot study. *Life Sciences and Medicine Research*, 19(1), 2-12
14. Smith, A. P., & Smith, H. N. (2017, June). Workload, fatigue and performance in the rail industry. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 251-263). Springer, Cham
15. Fan, J., & Smith, A. P.: The impact of workload and fatigue on performance. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 90-105). Springer, Cham (2017). DOI: 10.1007/978-3-319-61061-0\_6
16. Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, 30(1), 111-120.
17. Yeh, Y. H., & Wickens, C. D.: The dissociation of subjective measures of mental workload and performance (final report). (No. NASA-CR-176609; NAS 1.26:176609; EPL-84-2/NASA-84-2) (1984)
18. Casper, P. A. Dissociations among measures of mental workload: Effects of experimenter-induced inadequacy (1990)



19. Horrey, W. J., Lesch, M. F., & Garabet, A. (2009). Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. *Journal of safety research*, 40(1), 7-12
20. Kamzanova, A. T., Kustubayeva, A. M., & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human factors*, 56(6), 1136-1149
21. Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich IV, J. (2015). The psychometrics of mental workload: Multiple measures are sensitive but divergent. *Human Factors*, 57(1), 125-143
22. Hancock, P. A. (2017, June). Whither workload? Mapping a path for its future development. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 3-17). Springer, Cham (2010). [https://doi.org/10.1007/978-3-319-61061-0\\_1](https://doi.org/10.1007/978-3-319-61061-0_1)
23. Muñoz-de-Escalona, E., & Cañas, J. J. (2018, September). Latency differences between mental workload measures in detecting workload changes. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 131-146). Springer, Cham
24. Muñoz-de-Escalona, E., Cañas, J. J., & van Nes, J. (2019, November). Task Demand Transition Rates of Change Effects on Mental Workload Measures Divergence. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 48-65). Springer, Cham
25. Muñoz-de-Escalona, E., Cañas, J. J., & Noriega, P. (2020). Inconsistencies between mental fatigue measures under compensatory control theories. *Psicológica Journal*, 1(ahead-of-print)
26. Hancock, P. A., & Matthews, G. (2019). Workload and performance: Associations, insensitivities, and dissociations. *Human factors*, 61(3), 374-392
27. Cañas, J., Quesada, J., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics*, 46(5), 482-501
28. Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International journal of industrial ergonomics*, 36(7), 623-636
29. Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *American Economic Review*, 91(2), 67-72

30. Wickens, C. D. Multiple resources and mental workload. *Human factors*, **50**(3), 449-455 (2008). <https://doi.org/10.1518%2F001872008X288394>
31. Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock Jr, J. R. The multi-attribute task battery ii (matb-ii) software for human performance and workload research: A user's guide (2011)
32. Lee, J., & Ahn, J. H. Attention to banner ads and their effectiveness: an eye-tracking approach. *International Journal of Electronic Commerce*, **17**(1), 119-137 (2012). <https://doi.org/10.2753/JEC1086-4415170105>
33. Brennan, S.D: An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. DRA Technical Memorandum (CAD5) 92017, DRA Maritime Command and Control Division, Portsmouth (1992)
34. Jordan.C.S. Experimental study of the effect of an instantaneous self assessment workload recorder on task performance. DRA Technical Memorandum (CAD5) 92011. DRA Maritime Command Control Division, Portsmouth (1992)
35. Matthews, G., Middleton, W., Gilmartin, B. y Bullimore, M. A.: Pupillary diameter and cognitive and cognitive load. *J. Psychophysiol.* **5**, 265–271 (1991)
36. Bacs, R. W. y Walrath, L. C.: Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Appl. Ergon.* **23**, 243–254 (1992). DOI: 10.1016/0003-6870(92)90152-L
37. Hyönä, J., Tommola, J. y Alaja, A.: Pupil dilation as a measure of processing load in simultaneous interpreting and other language tasks. *Q. J. Exp. Psychol.* **48**, 598–612 (1995). DOI: 10.1080/14640749508401407
38. Granholm, E., Asarnow, R. F., Sarkin, A. J. y Dykes, K. L.: Pupillary responses index cognitive resource limitations. *Psychophysiology*, **33**, 457–461 (1996). DOI: 10.1111/j.1469-8986.1996.tb01071.x
39. Iqbal, S. T., Zheng, X. S. y Bailey, B. P.; Task evoked pupillary response to mental workload in human-computer interaction. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1477–1480. ACM, New York (2004). DOI: 10.1145/985921.986094
40. Verney S. P., Granholm, E. y Marshall S. P.: Pupillary responses on the visual backward masking task reflect general cognitive ability. *Int. J. Psychophysiol.* **52**, 23–36 (2004). DOI: 10.1016/j.ijpsycho.2003.12.003

41. Porter, G., Troscianko, T. y Gilchrist, I. D.: Effort during visual search and counting: insights from pupillometry. *Q. J. Exp. Psychol.* **60**, 211–229 (2007). DOI: 10.1080/17470210600673818
42. Privitera, C. M., Renninger, L. W., Carney, T., Klein, S. y Aguilar, M.: Pupil dilation during visual target detection. *J. Vision.* **10**, 1–14 (2010). DOI:10.1167/10.10.3
43. Reiner, M., & Gelfeld, T. M. Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *International Journal of Psychophysiology*, **93**(1), 38-44 (2014)
44. Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S.: Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav. Res. Methods.* **50**(1), 94–106 (2018). DOI: 10.3758/s13428-017-1007-2
45. Longo, L. Subjective usability, mental workload assessments and their impact on objective human performance. In *IFIP Conference on Human-Computer Interaction. 2017* (pp. 202-223). Springer, Cham.
46. Longo L. Experienced mental workload, perception of usability, their interaction and impact on task performance. *PloS one.* 2018 Aug 1;13(8):e0199661.
47. Longo L. On the Reliability, Validity and Sensitivity of Three Mental Workload Assessment Techniques for the Evaluation of Instructional Designs: A Case Study in a Third-level Course. In *International Conference of Computer Supported Education CSEDU (2) 2018* (pp. 166-178).

## **Chapter V**

---

*Inconsistencies Between Mental Fatigue  
Measures under Compensatory Control  
Theories*



## **Inconsistencies between mental fatigue measures under compensatory control theories**

Enrique Muñoz-de-Escalona<sup>1</sup>

José Juan Cañas<sup>1</sup>

Paulo Noriega<sup>2,3</sup>

<sup>1</sup> Mind, Brain and Behaviour Research Centre,  
Department of Experimental Psychology, Faculty of Psychology,  
University of Granada

<sup>2</sup> Ergonomics Laboratory, Faculdade de Motricidade Humana, Universidade de Lisboa,  
Portugal

<sup>3</sup> CIAUD, Faculdade de Arquitetura, Universidade de Lisboa, Portugal

---

This article has been published: Muñoz-de-Escalona, E., Cañas, J. J., & Noriega, P. (2020). Inconsistencies between mental fatigue measures under compensatory control theories. *Psicológica Journal*, 41, 103-126. <https://doi.org/10.2478/psicolj-2020-0006>

## Abstract

Mental fatigue has traditionally been defined as a condition of reduced cognitive efficiency and performance, accompanied by a subjective feeling of fatigue. Even though we could expect to find associations between the three defining characteristic of mental fatigue (performance impairment, physiological deactivation and subjective fatigue), research has shown that the emergence of inconsistencies between measures is more frequent than one might expect: people proved capable of maintaining adequate performance levels even after having declared themselves fatigued. This could be explained under the compensatory control mechanism models, which state that humans are able to provide additional resources under demanding conditions, but only at the expense of psychophysiological cost and subjective fatigue. We tested this explanation by manipulating task complexity and time performing a simulated air-traffic control task. We collected psychophysiological, performance and subjective data. A decrease in pupil size was seen in the low-aircraft-density condition, while pupil size remained constant in the high-aircraft-density condition. Participants' task performance was optimal in both conditions, though they showed an increase in subjective feelings of fatigue, especially in the high-complexity task condition. Thus, complexity seemed to trigger compensatory mechanisms, which reallocated extra resources that physiologically activated participants in order to deal with a higher complexity task, whereas subjective fatigue could be acting as a signal to the organism of impending resource depletion. Our findings support compensatory control theories and offer an explanation of inconsistencies between fatigue measures. Further research on compensatory mechanisms is needed to enable better management of fatigue effects to prevent work-related accidents.

**Abbreviations.** TOT: Time On Task, ATCo: Air Traffic Controller.

## 1. Introduction

Human behaviour and mental activity require energy. In the same way that a car needs petrol and oxygen to move through streets, human beings need energy from food to perform mental and motor activities. In a sense, we could say that in the life sciences, research has followed a mechanistic paradigm according to which human machinery function depends on supplied energy. Thus, for instance, it is assumed that performance on a task will improve or deteriorate depending, among other things, on the quantity and quality of the energy (resources) supplied (Kahneman, 1973).

An important feature of the mechanistic paradigm is the assumption that resources are limited and can be depleted. As a result, performance could be affected by the emergence of mental fatigue (Gopher and Donchin, 1986; Kahneman, 1973). Mental fatigue lowers mental alertness levels, and thus performance deteriorates (Grandjean, 1989) after long periods of ‘demanding cognitive activity which requires sustained mental efficiency’ (Lorist et al., 2005). Therefore, mental fatigue has been considered responsible for performance decrements as well as the main cause of work-related accidents (Grandjean, 1989; Dawson, et al., 2011; Hopstaken et al., 2016).

Many studies have tried to define and establish neuropsychological, psychosocial and even bio-mathematical models of fatigue (Dawson et al., 2011). However, almost all models agree that fatigue has three well-established components, two of which are observable (performance and physiological) and one of which is a non-observable subjective feeling. Thus, based on conventional definition of fatigue, it can be defined as the deterioration of task performance with Time on Task (TOT), accompanied by a decrease in physiological activation and the emergence of subjective feelings of fatigue. For this reason, mental fatigue has been approached with a methodology that includes three kinds of measures corresponding to the three defining characteristics of mental fatigue (Hockey, 2013):

- **Performance measures:** this type of methodology is composed of primary task measures (e.g. n° of errors, reaction time, etc.) and secondary task measures (e.g. choice reaction-time tasks, time estimation, memory-search tasks etc.). The aim is to measure objective performance indexes in order to quantify the quantity and the quality of performed tasks. Mental fatigue would result in task performance impairment (Chalder et al., 1993; Smets et al., 1995; Van der Linden, et al., 2003; Barker and Nussbaum, 2011; Vrijkkotte, et al., 2018).



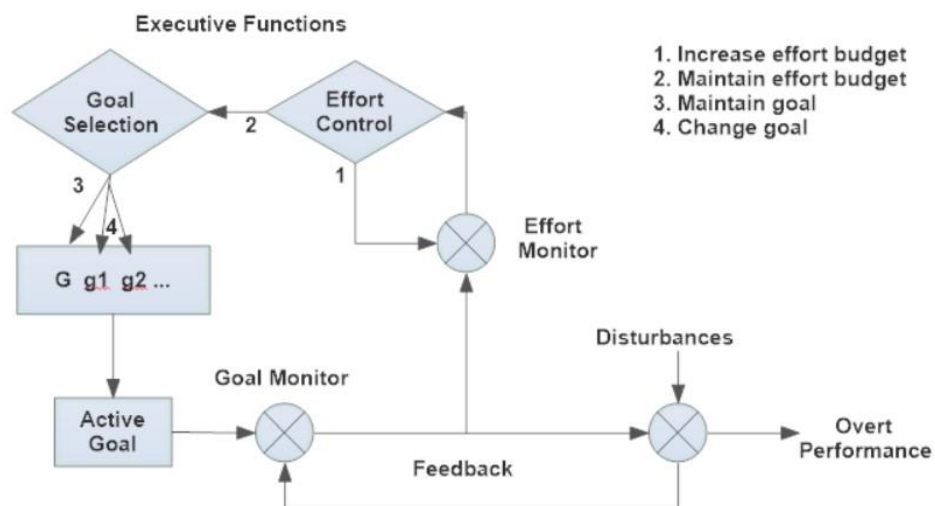
- **Subjective measures:** this sort of methodology includes all self-reported measures (e.g. NASA Task Load Index or Multidimensional Fatigue Inventory (MFI) (Smets et al., 1995)). The aim is to obtain easy and low-cost subjective data about perceived mental fatigue. Mental fatigue would induce a subjective feeling of fatigue (DeLuca, 2005; Kanfer, 2011).
- **Physiological measures:** this type of methodology comprises every physiological response to mental activation (e.g. pupil diameter, heart rate variability (HRV), electroencephalogram (EEG), etc.). The aim is to collect objective physiological data, which reacts to mental fatigue. Literature research has shown that physiological changes arise when a person is experiencing mental fatigue (Kohl, 2016). EEG measurement techniques are considered one of the most effective methods for assessing mental fatigue, as reflected by numerous studies. For example, Zhang et al. (2006) found that Approximate Entropy of EEG can serve as an index for detecting mental fatigue level (Zhang et al., 2008). Moreover, it has also been found that relative wavelet packet energy in  $\beta$ ,  $\beta/\alpha$ ,  $(\alpha+\theta)/\beta$  frequency bands and wavelet packet entropy of EEG are strongly correlated with mental fatigue (Nie, 2005; Zhang and Yu, 2010). Shen et al. (2008) tested an EEG-based mental fatigue monitoring system using a probabilistic-based support vector-machines (SVM) method, which provided a valuable estimation of confidence in the prediction of mental fatigue. On the other hand, eye parameters such as pupil dilation, blinks, fixations and saccades have also been widely used for estimating mental workload (Hess & Polt, 1964; Beatty, 1982; Nakayama et al., 2002; Bailey & Iqbal, 2008; Di Stassi et al., 2010; Jainta and Baccino, 2010; Benedetto et al. 2011; Lemonnier, Roland and Baccino, 2014; De Alwis, 2017; Yokoyama et al. 2018) as well as for mental fatigue (Roy et al., 2014; Horiuchi, 2017). Researchers have found high positive correlations between pupil diameter changes indicative of decreased physiological activation and the amount of resources used to perform certain tasks (Kahneman and Beatty, 1966; Beatty and Lucero-Wagoner, 2000; LeDuc, et al., 2005; Van Orden et al., 2001; Ahlstrom and Friedman-Berg, 2006). For example, Hopstaken et al. (2015) found that increasing mental fatigue coincided with diminished stimulus-evoked pupil dilation; similarly, Hopstaken et al (2016) found that baseline pupil diameter significantly decreased with time on task

throughout the experiment as fatigue increased, which was in line with the former findings.

Presumably, since the three types of measures reflect the same concept, we would expect them to be interrelated. In other words, we would expect correlations within measurements (associations), in such a manner that we could be able to detect fatigue with each kind of measure independently. However, research has found inconsistencies (disassociations and insensitivities) between measures more frequently than one might expect (Johnson, et al. 1997; Johnson, et al.1998; Paul et al., 1998; Krup and Elkins, 2000; Christodoulou, 2005; DeLuca, 2005; Bailey, et al. 2007; Bruce et al., 2010; Walker et al., 2012; Ishii et al., 2013; Hockey, 2013, Hancock, 2017). An insensitivity occurs when a workload measure does not reflect any change with a change in task load, whereas a dissociation takes place when that measure contradict the change in task load (Hancock, 2017). For example, it has frequently been found that people performing prolonged difficult tasks have exhibited no discernible performance decrements even after having declared that they are fatigued (for a review see Hockey, 2013). Additionally, as Gilbert (2009) has stated, most inconsistencies have involved the subjective component of the individual. At this point, it is unclear why such discrepancies between the results for different types of measures which are supposed to be measuring the same construct are so common, and the issue of how and why in certain situations fatigue does not affect performance or physiological activation when a person claims feeling fatigued requires explanation and further investigation.

With that in mind, some researchers have proposed new resource theories which postulate the existence of a compensatory control mechanism that regulates performance under stress and high mental workload conditions. Specifically, the function of this mechanism is to provide additional resources dynamically, depending on the level of stress and mental workload experienced. Thus, task performance will not necessarily be impaired in demanding and fatiguing conditions (e.g., a high-complexity task) due to this provision of extra resources needed to perform the task successfully, but at the expense of psychophysiological cost and subjective fatigue. Hockey (1997) proposed the compensatory control model of fatigue, which asserts that 'the maintenance of performance stability under demanding conditions is an active process under the control of the individual, requiring the management of cognitive resources through the mobilisation of mental effort' (p.78).

The compensatory control model of fatigue makes a distinction between two levels of control (Figure 1): a lower control loop, which manages routine regulatory activity, and an upper loop, which is needed to deal with harder challenges to performance. Even initially, executive control is exerted to the extent that the achievement of a selected task goal (G) is prioritized over many other possible contender goals. However, the other possible goals (g1, g2, etc) remain relevant to motivational needs of the individual (rest, change task, etc.). But if strain is detected in the lower control loop, the upper control loop kicks in. A choice is made between two possible options: 1) increasing effort (reallocating extra resources) to protect performance from impairment, with an increase in costs or 2) maintaining the current level of effort with minimal costs, despite the risk of performance decline (either maintaining the current goal with a lower performance standard or displacing the goal with one of the competing goals). In either case, the feeling of fatigue is expected to disappear as strain becomes extinct (Hockey, 2011).



**Figure 1.** Compensatory control model of fatigue (after Hockey 1997)

Therefore, as Hockey (2011) suggests, ‘the subjective fatigue state may be identified broadly with the outcome of the monitoring process, which detects control problems and the need for greater effort’(p.18), hence ‘within the context of an effort-fatigue compensatory loop a sensed need for greater effort reflects the same affective state as a sensed increase in discomfort of fatigue’(p.18).

Even though there are some studies whose results are consistent with Hockey's compensatory control model (Sauer et al 2003; Venables and Fairclough 2009), we think further research is needed to strengthen compensatory control theories, particularly in relation to inconsistencies between fatigue measures. To address that need, and to provide greater knowledge about the phenomenon of mental fatigue, we conducted the experiment reported herein.

The main objective of this study is to test whether measurements of the aforementioned three aspects of fatigue correlate or not and, if the latter happens, to gather evidence regarding how and why such inconsistencies between measures occur, using the compensatory mechanism approach as our conceptual framework.

We manipulated two independent variables in this experiment. The first independent variable was 'Time on Task' (TOT), whereas the second independent variable was task complexity. An intuitive approach would lead us to believe that both of them are factors that can cause mental fatigue (Stern, 1997): Resources will run out as time is spent on task performance, so the more time is spent, the fewer resources will be available and therefore the level of mental fatigue will be greater. Moreover, it could also be assumed that task complexity determines the quantity of resources dedicated to a task, so that more complex tasks require more resources. Hence, the more complex the task is, the greater resource depletion will be with TOT and thus, the greater the level of mental fatigue. For example, we may think that the end of the working day would be a more likely time for work-related accidents than the beginning of the day, due to fatigue caused by resource depletion, especially with high complexity tasks.

Our hypothesis is that, in line with recent empirical findings (see a review in Hockey, 2013), we will find dissociations between the three kinds of fatigue measures when testing the effects of these two factors, TOT and complexity, that can be explained in terms of compensatory control mechanisms. Moreover, we predict that particularly subjective fatigue will not correlate with the physiological (pupil size) or performance indices because subjective fatigue functions as an alarm signal regarding resource depletion that kicks in even while there are still enough resources to perform the task properly, as high physiological and performance index scores would be reflecting.

The effects of TOT and complexity were tested in an air-traffic control simulation experiment in which participants were trained and instructed to avoid conflicts between aircraft. This complex and dynamic task allowed us to observe the possible effects of our

manipulated independent variables, as well as our measured dependent variables. Participants performed the task for 120 minutes while their performance, pupil diameter and feelings of fatigue were measured.

## 2. Materials and method

### 2.1 ATCLab-Advanced software

The software used for simulating ATCo tasks was an air traffic simulator called ATCLab-Advanced, which is available for free public download. (Fothergill et al., 2009) (see Figure 2). It provided a high level of realism as well as a simplified and easy handling, which allowed it to be used by all participants in several learning sessions. Additionally, it allowed strong experimental control of air traffic scenarios parameters.

For scenario development, first, the static characteristics of the simulation environment (control sector size and possible pathways through which aircraft could travel) were defined. Next, aircraft quantity (density) and initial aircraft parameters (altitude, assigned altitude, speed, time of appearance on stage and planned route) were defined for each aircraft presented in the scenario.

Finally, we note that the <sup>ATC</sup>Lab-Advanced simulator provided participants with every needed tool to carry out the ATCo task, such as the route (the aircrafts' fixed route was displayed), distance scale (which allows horizontal aircraft distance measurement) and altitude and speed change tools. Once the structural and dynamic scenario parameters were established, a file that could be launched by the simulator was obtained. This file recorded a '.log' file with performance data (tools used during the task and aircrafts' parameters) for each participant during simulation.

**Scenarios.** Scenarios used in the study varied according to whether the participant was in the training or the experimental stage. Thus, during the training sessions, the standard scenarios provided by the software creator software were used, but a specific scenario was programmed by the experimenters to achieve experimental session goals. This specific scenario was programmed with two different experimental conditions.

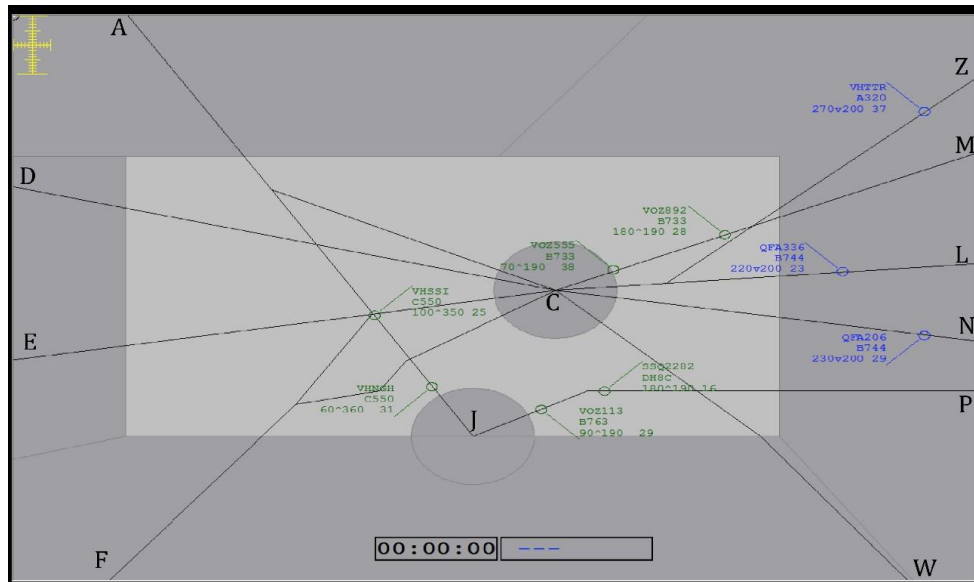
The structural features of both experimental conditions were identical, that is, identical control sectors and routes, as well as the same number of total initial number of aircraft

(9) and number of aircraft under participant control (6). For a better understanding, refer to Figure 2, which represents the initial simulator screen presented to participants; the capital black letters (starting route spots) do not appear on the radar screen.

Several differences in dynamic scenario parameters were established according to the scenario used in each condition:

- 1. Low aircraft density scenario (low-complexity task condition):** A total of 35 aircraft were presented, with 25 coming from external locations A,D,E,F,W,P,N,L,M and Z and 10 from inner locations C and J. More specifically, 3 aircraft came from A,D,E,F and W; 2 from P,N,L,M and Z; 6 from J (4 coming to P and 2 to A); and 4 from C (coming to A,D,E and F respectively).
- 2. High aircraft density scenario (high-complexity task condition):** A total of 70 aircraft were presented, thus double the density of the low aircraft density condition. 50 came from external locations A,D,E,F,W,P,N,L,M and Z and 20 from inner locations. Specifically, 6 each came from A,D,E,F and W; 4 from P,N,L,M and Z; 12 from J (8 of them coming to P and 4 to A); and 8 from C (coming to A,D,E and F respectively).

It should be noted that, traffic density is always variable in the sense that the number of flights cannot be kept exact during the whole 2 hours session, as this would not be realistic. In this sense, there is a small intragroup variation, which was designed to be equivalent in both conditions.



**Figure 2.** <sup>ATC</sup>Lab-Advanced software initial screenshot presented to participants

## 2.2 Procedure

Participants had to perform the ATCo tasks with the previously described <sup>ATC</sup>Lab-Advanced software, so they had to learn how to use it before proceeding through the experimental stage in which the performance data were collected. Thus, we established 2 distinct stages:

1. Training stage: This took place for a total of 2 hours over 2 consecutive days, 1 hour of training each day. The main objective of this first stage was for participants to familiarise themselves with the software so that they could handle it comfortably during the experimental stage. The training stage procedure was as follows: during the first day, researcher explained the participants their main task goal (maintaining air traffic security and preventing potential conflicts between aircraft), they started reading a short manual about the operation of the simulator for about 20 minutes and were asked to call the researcher once they had finished. Later, the participants sat in front of the running simulator while the researcher reviewed the manual in detail with the participants to ensure both correct understanding of the task and the assimilation of knowledge through content review. Participants then started using the simulator on their own while the researcher executed different air traffic control scenarios in order of difficulty (3 scenarios). The participants had free access to both the manual and researcher

at all times in case of doubts or questions. The researcher also periodically checked the participants' performance to monitor their learning. Through the second day, participants would continue working through scenarios (4 in total) in order of increasing difficulty, under the same conditions as the previous day. Once the training period concluded, participants were ready for the experimental session, which took place the next day.

2. Data collection stage: The aim of the data collection stage, which lasted a total of 2 total hours, was to collect experimental performance data from participants while they performed ATCo tasks. Both objective (physiological and execution data) and subjective (mental fatigue subjective index) data were collected. The participants were told the differences between the training and experimental stages, which were as follows: First, they would perform ATCo tasks in front of an eye-tracker system that had been previously calibrated. Secondly, participants were instructed to fill in the mental fatigue scale every 5 minutes, when a scheduled alarm sounded. The whole session lasted 2 hours. Once the differences were explained, participants began the task, with participants working in the low-aircraft-density condition (35 aircraft) or in the high-aircraft-density condition (70 aircraft). At the end of the session, the participants were thanked and given extra study credits.

### **2.3 Tobii T120 eyetracker**

To obtain the pupil diameter variable, an infrared-based eye tracker system, the Tobii T120 Video System was used. This system achieves a minimum accuracy of .01 mm, is completely non-intrusive and provides high precision and an excellent head compensatory movement mechanism, which ensures high-quality data collection. In addition, the freedom of movement it offers participants allows them to act naturally in front of the screen, as if it were an ordinary computer display.



## 2.4 Subjective mental fatigue scale

We needed an on-line subjective mental fatigue measure, thus we developed an easy and intuitive subjective instant self-assessment fatigue scale with which participants could write down how much fatigue they were currently experiencing. The scale was designed to elicit a response at every 5-minute interval throughout the 2 hours of the experimental stage. The scale ranged from 1 ‘no mental fatigue’ to 5 ‘maximum mental fatigue’, presented from left to right in ascending amount of fatigue experienced. Participants were taught to use the scale just before experimental stage begin and were instructed about mental fatigue concept in order to avoid possible confusion with similar concepts such as workload or frustration (See Figure 3).

### Escala online de Fatiga Mental Subjetiva

Responda según las siguientes opciones:  
 1 (Nada Fatigado), 2 (Poco Fatigado), 3 (Moderadamente Fatigado), 4 (Bastante Fatigado), 5 (Totalmente Fatigado).

- 5 min \_\_\_\_\_ 1 2 3 4 5
- 10 min \_\_\_\_\_ 1 2 3 4 5

**Figure 3.** Instantaneous self-assessment subjective fatigue scale.

## 2.5 Participants

Seventy psychology students at the University of Granada participated in the study for extra course credit. Participants were divided into 2 groups, 34 participants in the low complexity condition and 36 participants in the high complexity condition. Participants’ ages ranged from 18 to 27, with an average of 21.6 and a median of 20. A total of 56 women (80%) and 14 men (20%) participated. This gender imbalance is likely the result of the fact that psychology students at University of Granada are mostly women. A further requirement was that none of the participants had any previous experience in ATCo tasks. Informed consent was obtained from each participant.

This study was carried out in accordance with the recommendations of the local ethical guidelines of the committee of the University of Granada institution called “Comité de Ética de Investigación Humana”. The protocol was approved by the “Comité de Ética de Investigación Humana”. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## **2.6 Experimental room conditions**

Sessions were held in several different rooms, depending on whether the participant was in the training or data collection stage: During the training stage, participants could work in one of three different rooms equipped for training with the simulator, and no special attention to room conditions was needed. However, during the data collection stage, standardizing room conditions was essential. Pupillary responsiveness is related to interaction of sympathetic and parasympathetic components and depends on three kind of reflexes: light, proximity and psychosensory reflexes (Beatty & Lucero-Wagoner, 2000). The parasympathetic component is highly sensitive to lighting conditions, thus lighting conditions were kept constant (55 lx measured at the eye by a calibrated luxmeter), with constant artificial lighting and no natural light in the rooms. Moreover, participants always sat into the same place, a comfortable chair spaced 60 cm from the eye-tracker system and the testing room was temperature controlled to 21°C.

## **2.7 Experimental design**

The experimental design was a 22 X 2 Mixed Factorial. One of the independent variables, TOT, was a within-subjects factor. This variable had 24 intervals of 5 minutes, though the first 5-minute interval was taken as a baseline measure of the pupil diameter of each subject. In addition, we discarded the last interval from our analysis due to a high level of experimental noise during the last 5 minutes, which coincided with a researcher being present in the experimental room to oversee and end the session and would unintentionally increase participants' arousal. The second independent variable was Task Complexity, which was measured by the number of aircraft (traffic density) entering into the sector. This variable had two levels, Low Complexity task (low aircraft density

scenario) and High Complexity task (high aircraft density scenario). The dependent variables were task performance, pupil diameter, and subjective feeling of fatigue.

## **2.8 Dependent measures**

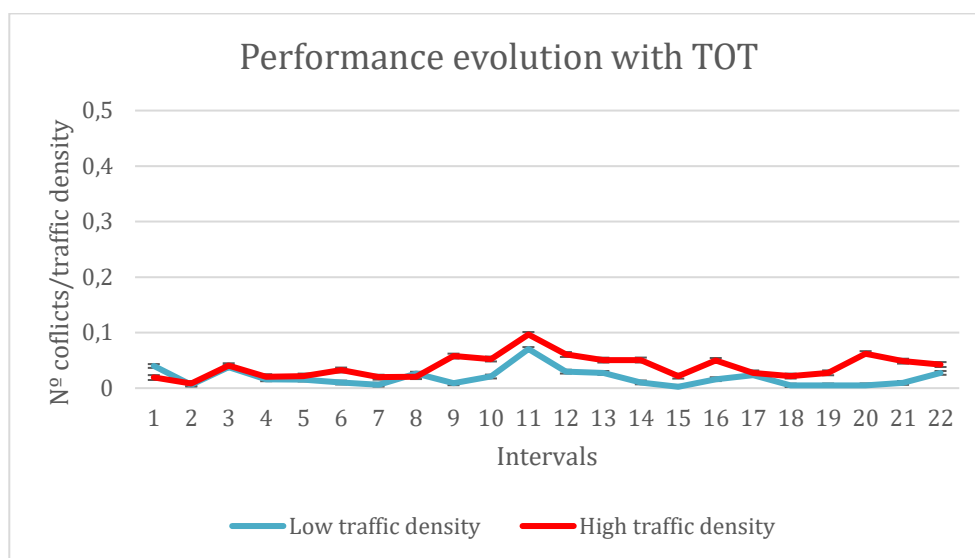
We operationalized the dependent variable ‘performance’ by dividing the number of conflicts by total present aircraft number at a given time, since we thought that using only the number of conflicts as a performance measurement wouldn’t be an appropriate performance indicator, since it largely reflects air traffic density, so that way, it wouldn’t be possible to establish comparison between the two traffic density conditions. It is worth noting that a conflict between 2 aircraft takes place when the safe distance between both aircrafts, either horizontal (5 nautical miles) or vertical (1 thousand feet), is violated. In order to quantify the number of conflicts per interval, a specific software was designed to analyze .log files provided by the ATC simulator.

For the physiological measure (pupil diameter), while our eye tracking system allows continuous sampling rate recording at 120Hz, we set a total of 24 intervals lasting 5 minutes each to facilitate subsequent analyses. However, given high inter-individual variability relative to average pupil size, we followed the recommendations provided by Sebastiaan Mathôt et al (2018) about base-line correction of pupil-size data: we took the first interval (5 minutes) as a reference of standard individual average pupil size, which was then subtracted from the obtained value in each of the remaining 23 intervals, thereby giving a differential standardized value that allowed us to compare participants. Analyses were carried out both for the left and right pupils. A negative value meant that the pupil was contracting while a positive value meant that it was dilating.

Subjective mental fatigue was obtained on-line by the instant self-assessment fatigue scale explained above.

### 3. Results

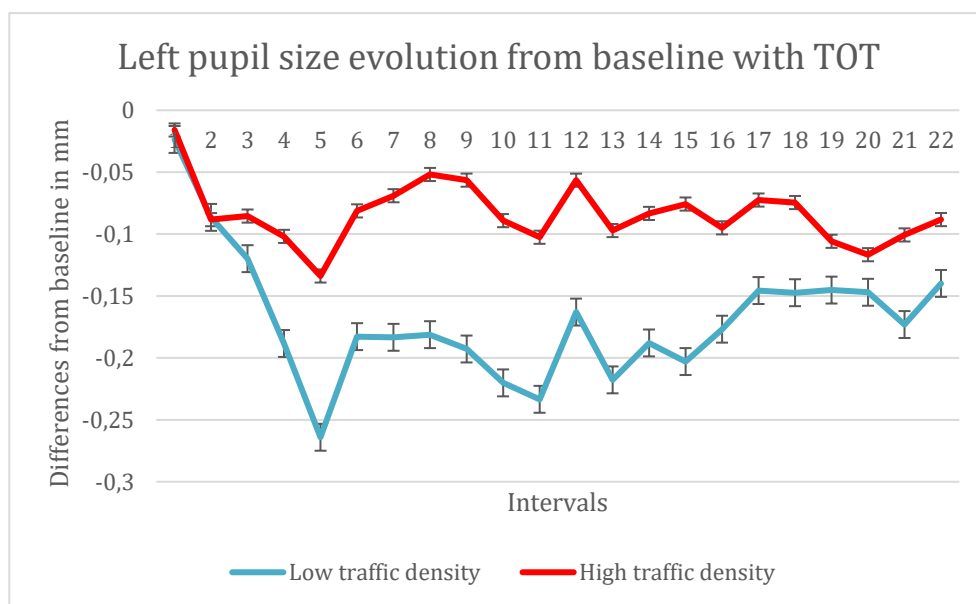
We used a two-way, mixed factorial ANOVA to analyse obtained results. First, we can see in Figure 4 that air traffic control task performance was optimal; thus, participants were able to avoid the vast majority of conflicts. However, in the ANOVA, the main effect of traffic density turned out to be significant ( $F(1,68) = 89.5$ ,  $MSe = .002$ ,  $p < .05$ ;  $\eta^2 = .57$ ). As expected, average performance was better in the low-aircraft density condition than in the high-aircraft density condition, which shows that our independent variable manipulation of task complexity proved to be effective. Furthermore, the main effect of interval also proved to be significant:  $F(21,1428) = 12.51$ ,  $MSe = .001$ ,  $p < .05$ ;  $\eta^2 = .16$ ). Thus, we noted a lower performance peak, which coincides with the maximum air traffic intensity peak on both conditions. This effect could not be considered a fatigue effect since it happens at approximately half of the total performance time but not at the end as would be expected if fatigue were the cause. We also found that the interaction of traffic density x TOT was significant:  $F(21,1428) = 4.22$ ,  $MSe = .006$ ,  $p < .05$ ;  $\eta^2 = .05$ . However, we consider this interaction effect very small and difficult to interpret. Baseline measurements between groups were very similar (low complexity: .11; high complexity: .14) and presented no significant differences ( $t(1,68) = -.51$ ,  $p > .05$ ).



**Figure 4.** Participants' performance during task development

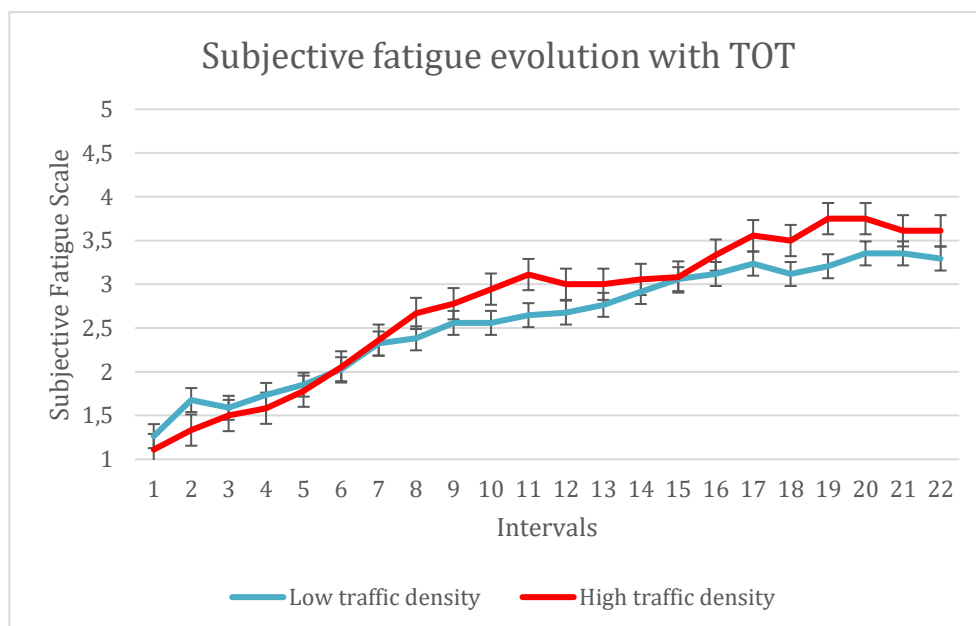
For pupil diameter, considering that all participants presented no ocular abnormalities, and left and right pupil measurements were highly correlated both for low ( $r = .92$ ;  $p < .01$ )

and high-complexity ( $r = .79$ ;  $p < .01$ ) conditions, only data recorded from the left pupil were used for further analyses. The main effect of traffic density was found to be significant for pupil diameter ( $F(1,68) = 5.95$ ,  $MSe = 0.47$ ,  $p < .05$ ;  $\eta^2 = .08$ ). In other words, average pupil size was smaller in the low complexity condition than in the high complexity condition. We also found a significant main interval effect for pupils ( $F(21,1428) = 6.22$ ,  $MSe = .13$ ,  $p < .05$ ;  $\eta^2 = .08$ ). Figure 5 shows a decrease in the pupil diameter variable until the fifth interval, from which it increases from interval 5 to 6, stabilizing from then until the last interval. However, this interval effect must be considered in light of its interaction with the traffic density variable, which also was significant ( $F(21,1428) = 2.49$ ,  $MSe = .013$ ,  $p < .05$ ;  $\eta^2 = .04$ ). We can see how changes occur in the variable depending on air traffic density, that is, depending on task complexity level. As shown in Figure 5, an identical decrease in pupil diameter occurred in both experimental conditions until interval 2. Nevertheless, from then on, pupil diameter continued to decrease in the low aircraft density condition, stabilizing in the sixth interval and then slightly increasing again, while it remained constant in the high aircraft density condition. Trend analysis revealed a quadratic trend in pupils ( $F(1,33) = 27.25$ ,  $MSe = .91$ ,  $p < .01$ ;  $\eta^2 = .45$ ) in the low-aircraft density condition. Baseline measurements between groups were very similar (low complexity: 3.55mm; high complexity: 3.53mm) and presented no significant differences ( $t(1,68) = .176$ ,  $p > .05$ ).



**Figure 5.** Participants' left pupil diameter differences from baseline in mm, during task development

As shown in Figure 6, a linear increase in subjective experienced fatigue occurs over time, although we found a greater increase in the high-complexity task condition. The main effect of traffic density on task was not significant ( $F(1,68) = .94$ ,  $MSe = 11.96$ ,  $p > .05$ ;  $\eta^2 = .01$ ). However, the interval main effect turned out to be significant ( $F(21,1428) = 80.5$ ,  $MSe .47$ ,  $p < .05$ ;  $\eta^2 = .54$ ). Thus, a significant variation in subjective fatigue occurred over the intervals, assuming an upward linear and quadratic trend, given a linear ( $F(1,68) = 230.64$ ,  $MSe = 3.24$ ,  $p < .05$ ;  $\eta^2 = .78$ ) and a quadratic ( $F(1,68) = 21.3$ ,  $MSe = 1.9$ ,  $p < .05$ ;  $\eta^2 = .24$ ) contrast test. Lastly, the interaction between the two variables was significant, ( $F(21,1428) = 1.99$ ,  $MSe .47$ ,  $p < .05$ ;  $\eta^2 = .03$ ), maintaining linear and quadratic trends on both low-aircraft ( $F(1,33) = 77.07$ ,  $MSe = 279.14$ ,  $p < .01$ ;  $\eta^2 = .7$  &  $F(1,33) = 11.37$ ,  $MSe = 6.41$ ,  $p < .05$ ;  $\eta^2 = .16$ ) and high-aircraft ( $F(1,35) = 168.55$ ,  $MSe = 485.36$ ,  $p < .01$ ;  $\eta^2 = .83$  &  $F(1,35) = 15.91$ ,  $MSe = 31.72$ ,  $p < .01$ ;  $\eta^2 = .31$ ) density conditions, respectively. The quadratic trend could be interpreted as a ceiling effect: participants reached the maximum level of fatigue. Regarding subjective fatigue baseline measures, there exist a difference of .18 between groups (low complexity: 1.24; high complexity: 1.06), which represents less than 1% of difference. Nevertheless, we found this difference statistically significant ( $t(1,68) = 2.19$ ,  $p < .05$ ) and despite it could be due to the low variance error in the ANOVA, this partly limits any inference on this variable.



**Figure 6.** Participants' subjective mental fatigue during task development

Finally, we performed a bayesian correlation analysis in order to assess the probability of the null hypothesis being true (it contrasts the likelihood of the data fitting under the null hypothesis with the likelihood of fitting under the alternative hypothesis). In Bayesian statistics, a Bayesian Factor  $B_{01}=1$  means no evidence in favour of either the null hypothesis or the alternative. Bayesian Factors  $< 1$  denote evidence inclined toward the alternative hypothesis, while Bayesian Factors  $> 1$  denote evidence inclined toward the null hypothesis (Jarosz and Wiley, 2014). As depicted in Table 1, we could not find correlations between pupil diameter and subjective fatigue measures for either low ( $r = -.13$ ;  $p>.05$ ;  $B_{01}=3.28$ ) or high-complexity ( $r = -.12$ ;  $p>.05$ ;  $B_{01}=3$ ) conditions. Similarly, we could not find correlations between subjective fatigue and performance measures for either low ( $r = -.26$ ;  $p>.05$ ;  $B_{01}=1.95$ ) or high-complexity ( $r = .38$ ,  $p>.05$ ;  $B_{01}=1.19$ ) conditions.

**Table 1.** Correlation chart between measures

		Low-Complexity Task			High-Complexity Task		
		Pupil	Subjective Fatigue	Performance	Pupil	Subjective Fatigue	Performance
Pupil	Pearson	1	-,130 $B_{01}=3.28$	-,019 $B_{01}=3.86$	1	-,122 $B_{01}=3$	-,307 $B_{01}=.67$
	Sig. (bilateral)		,553	,931		,581	,154
	N	23	23	23	23	23	23
Subjective Fatigue	Pearson	-,130 $B_{01}=3.28$	1	-,261 $B_{01}=1.95$	-,122 $B_{01}=3$	1	,382 $B_{01}=1.19$
	Sig. (bilateral)	,553		,229	,581		,072
	N	23	23	23	23	23	23
Performance	Pearson	-,019 $B_{01}=3.86$	-,261 $B_{01}=1.95$	1	-,307 $B_{01}=.67$	,382 $B_{01}=1.19$	1
	Sig. (bilateral)	,931	,229		,154	,072	
	N	23	23	23	23	23	23

## 4. Discussion

Mental fatigue is defined as a condition of reduced cognitive efficiency and performance caused by excessive mental workload sustained over time, accompanied by a subjective feeling of being exhausted and physical discomfort. Hence, **we could expect that a person experiencing mental fatigue would show task performance impairment, lower physiological activation and higher subjective mental fatigue.** In other words, we could expect to find associations between performance, physiological activation and subjective feeling of fatigue.

However, many authors have found that in certain situations where a decrease in performance was expected due to a highly demanding and/or lengthy task, no such decrease appeared (performance insensitivity). In these cases, people proved capable of maintaining adequate performance levels, even after having declared themselves fatigued (for a review see Hockey, 2013).

These results have led some authors to propose new resource theories that defend the existence of a **compensatory control mechanism responsible for the ability to maintain high levels of performance and physiological activation, even when the individual feels fatigued.**

The results from our study supported the existence of such a compensatory control mechanism. Our findings revealed relative stable performance in both traffic density conditions, as participant's performance was optimal; however, we did find differences in activity levels: Participants performing the low-complexity task experienced higher pupil diameter decrement, specifically during the first 5 intervals, than participants in the high-complexity task. From then on, in the high-complexity condition, pupil size remained stable, whereas a slight increase can be noted in the low-complexity condition until the end of the task (quadratic trend). On the other hand, we observed that subjective fatigue linearly increased in both groups (as trend analysis showed up) through the intervals, although we found a greater increase with TOT in the high-complexity task condition. Hence, our results showed:

- **Performance and physiological insensibilities:** we would expect performance to get worse and physiological activation to get lower with TOT, especially in the high-complexity task. However, **both remained stable at different levels** throughout the whole task, and the latter even slightly increased in the low-complexity condition.



- **Dissociations between subjective and physiological/performance measures:** taking into account the stability found on performance and physiological measures (at different levels), we would not expect subjective fatigue to appear with TOT. Nevertheless, our results revealed a steady subjective fatigue increase in both groups with TOT, being greater in the high-complexity condition. In other words, **we did not find correlations between subjective fatigue measures and physiological/performance measures.**

We think these insensitivities and dissociations can be explained by the compensatory control mechanism approach. **The compensatory mechanism modulates human performance, by reallocating resources when necessary, in order to successfully cope with the task:** participants in the high-complexity task maintained a steady higher level of activation over time to deal with a higher complexity task. In contrast, participants in the low-complexity task did not require that high level of activation and their levels decreased during the first intervals until they reached the appropriate activation levels for that lower task difficulty, slightly increasing from then until the last interval. Although we do not have a clear explanation about that slight increase in activation, we believe that the participant might be fighting against task disengagement and boredom by slightly increasing their activation level throughout the rest of the task. We must bear in mind that a 2-hour low-complexity monitoring task can be very exasperating. Although good performance and stable physiological activation levels would not be in principle indicators of mental fatigue, subjective fatigue linearly increased in both conditions with TOT, especially in the high-complexity condition. **This subjective fatigue increase would be higher in the high-complexity condition because a higher level of resources are being reallocated (effort) by the compensatory control mechanism in order to cope with a higher difficulty task.** Hence, we could say that as Hockey (2013) has noted, **subjective fatigue acts as an organismic alarm signal against resource depletion** (similarly to thirst or hunger), even when there are still enough available resources to tackle the task.

In summary, **the compensatory mechanism allows the individual to maintain good performance on task development with TOT, even under high difficulty conditions, and this will be reflected by higher physiological activation. But we can also state that this need for greater effort would result in subjective fatigue feelings, which act**

**as an alarm signal against resource depletion in such a manner that the more resources are being reallocated (higher effort), the more subjective fatigue increases.** Thus, for example, a worker may be maintaining an appropriate level of performance and high physiological activation in a task with a high level of mental demand, yet still experience high levels of subjective fatigue. This indicates the use of extra resources, but because of the risk of resource depletion, the risk of accidents is increased. Therefore, returning to the car metaphor used in the introduction, subjective fatigue acts as a low fuel warning light that is activated when there is still enough fuel to continue while warning us that our fuel is nearly running out.

Several limitations must be addressed before concluding. First, we only had one physiological measure, pupil diameter. Future research should also utilize **other physiological measures**, such as electrodermal activity, heart-rate variability and EEG. We hypothesize that every physiological measurement would reflect activation changes produced by the compensatory control mechanism. That is, higher activation would be reflected in higher electrodermal activity, lower heart-rate variability and higher EEG activity, and the opposite for low activation. However, inconsistencies have been found not only between performance, psychological and subjective measures, but **also within different indicators of each kind of measure; thus, further research is needed to untangle this issue.** It also needs to be noted that we conducted this research with students under simulated conditions; hence, we think it would be necessary to **replicate this experiment under real working conditions** to improve validity. On the other hand, the design aspect involved use of two different subject groups who were assigned either the low or high density task, but no subjects were tested in both tasks, which means that any differences could be in part attributed to group difference. In any case, we think that possible inter-group differences may have been in part solved statistically through randomization procedure (N=70). We also checked baseline values (1st interval) of both groups in dependent variables and found they are very similar. However, despite performance and pupil size presented no statistical differences, we found statistical differences in subjective experienced fatigue. This baseline similarities between groups gives us certain confidence that no differences could be in part attributed to group differences, except in the case of subjective experienced fatigue. Finally, participants were alerted by an alarm every 5 minutes to assess subjective mental fatigue scale and it may be possible that these breaks could have influenced the impact of time-on-task and, consequently, the accumulated mental fatigue. However, we think that as the fatigue scale

was designed to be very easy handling and to be answered within 1 or 2 seconds, we believe that these breaks to assess mental fatigue are not very determining and do not brake the situational awareness of participants but, must be taken into consideration in future research.

In conclusion, we could say that, based on our results, **some inconsistencies between fatigue measurements would be explained according to the compensatory control mechanism action. Moreover, our results support the current resource theories** (i.e. Hockey, 1997, 2011 2013) that replaced outdated classical resource theories (i.e. Kahneman, 1973).

Our results suggest a need for **caution about considering psychophysiological indexes such as pupil size as the only indicator of mental fatigue** (Cañas, 2017; Hopstaken et al, 2016; Sakamoto et al, 2009). Fatigue recognition systems have been introduced in some cars which claim to recognize drivers' fatigue by only taking into account drivers' eye behaviour. According to our results, ocular behaviour could indicate high levels of physiological activation, which would not be, in principle, an indicator of mental fatigue, but could instead indicate the use of extra resources to maintain optimal driving performance even though the driver has already started to feel fatigued.

## 5. Resumen

**Inconsistencias entre medidas de fatiga mental bajo las teorías del control compensatorio.** Tradicionalmente, la fatiga mental ha sido definida como una condición de reducción en los niveles de eficiencia cognitiva y rendimiento, acompañada de una sensación subjetiva de fatiga mental. A pesar de que podríamos esperar encontrar asociaciones (convergencia) entre las tres características definitorias de la fatiga mental (deterioro en el rendimiento, reducción en los niveles de activación fisiológica y surgimiento de fatiga subjetiva), la literatura ha revelado que la aparición de inconsistencias (divergencia) entre las medidas de fatiga es más frecuente de lo esperado: la gente se muestra capaz de mantener niveles adecuados de rendimiento a pesar de haber declarado encontrarse fatigados. Esto puede explicarse a partir de los modelos del mecanismo de control compensatorio, los cuales afirman que los seres humanos son capaces de proveerse con recursos adicionales bajo condiciones de elevada demanda cognitiva, únicamente a expensas de un coste psicofisiológico y del surgimiento de la

sensación subjetiva de fatiga mental. En el presente estudio, ponemos a prueba esta explicación manipulando el tiempo y la complejidad de una tarea de simulación de control de tráfico aéreo. Recabamos datos psicofisiológicos, de rendimiento y subjetivos. Nuestros resultados desvelan una disminución del diámetro pupilar en la condición de baja densidad de tráfico aéreo, en tanto que se mantiene constante en la condición de alta densidad. El nivel de rendimiento de los participantes resultó ser óptimo en ambas condiciones, a pesar de que se aprecia un incremento lineal en los niveles de fatiga subjetiva, especialmente en la condición de alta complejidad. Así, la complejidad parece activar el mecanismo compensatorio, el cual provee al organismo con recursos adicionales que mantienen fisiológicamente activados a los participantes de la condición de alta densidad de tráfico al objeto de hacer frente a una tarea de mayor dificultad, mientras que la fatiga subjetiva podría estar actuando como una señal del organismo para impedir el agotamiento de los recursos cognitivos. Nuestros hallazgos apoyan las teorías del control compensatorio y ofrecen una posible explicación sobre algunas inconsistencias entre las medidas de fatiga mental. La ciencia necesita seguir investigando el fenómeno del mecanismo compensatorio para favorecer la gestión de los efectos de la fatiga mental y prevenir los accidentes laborales.

## References

- Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623-636.
- Bailey, A., Channon, S., & Beaumont, J. G. (2007). The relationship between subjective fatigue and cognitive fatigue in advanced multiple sclerosis. *Multiple Sclerosis*, *13*(1), 73-80. Doi: <http://dx.doi.org/10.1177/1352458506071162>
- Bailey, B. P.; Iqbal, S. T. (2008) Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Trans. Comput. Interact.* *14*, 1–28.
- Barker, L. M., & Nussbaum, M. A. (2011). The effects of fatigue on performance in simulated nursing work. *Ergonomics*, *54*(9), 815-829.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.*, *91*, 276–292.

- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of Psychophysiology*, 2, 142-162.
- Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver workload and eye blink duration. *Transportation research part F: traffic psychology and behaviour*, 14(3), 199-208.
- Bruce, J. M., Bruce, A. S., & Arnett, P. A. (2010). Response variability is associated with self-reported cognitive fatigue in multiple sclerosis. *Neuropsychology*, 24(1), 77-83. doi:10.1037/a0015046
- Cañas, J. J. (2017). Fatigue and Theories of Resources: The Energetic Metaphor of the Mind Functioning Might Be Reconsidered. In: *Advances in Neuroergonomics and Cognitive Engineering* (pp. 269-277). Springer, Cham.
- Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessely, S., Wright, D., & Wallace, E.P. (1993). Development of a fatigue scale. *Journal of Psychosomatic Research*, 37(2), 147-153. doi:10.1016/0022-3999(93)90081-P
- Christodoulou, C. (2005). The assessment and measurement of fatigue. In: J. DeLuca (Ed.), *Fatigue as a window to the brain* (pp. 19-36). Cambridge, MA: MIT Press.
- Dawson, D., Ian Noy, Y., Härmä, M., Åkerstedt, T., & Belenky, G. (2011). Modelling fatigue and the use of fatigue models in work settings. *Accident Analysis & Prevention*, 43(2), 549-564.
- De Alwis Edirisinghe, V. (2017). Estimating Mental Workload of University Students using Eye Parameters (Master's thesis, NTNU).
- DeLuca, J. (2005). 19 Fatigue: Its Definition, Its Study, and Its Future. *Fatigue as a window to the brain*, 319.
- Di Stasi, L. L., Marchitto, M., Antolí, A., Baccino, T., & Cañas, J. J. (2010). Approximation of on-line mental workload index in ATC simulated multitasks. *Journal of Air Transport Management*, 16(6), 330-333.
- Fothergill, S., Loft, S., & Neal, A. (2009). ATC-labAdvanced: An air traffic control simulator with realism and control. *Behavior Research Methods*, 41(1), 118-127.
- Gilbert, D. (2009). *Stumbling on happiness*. Vintage: Canada.
- Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*, Vol. 2. Cognitive processes and performance (p. 1-49). John Wiley & Sons.

- Grandjean, E. (1989). *Fitting the task to the man: A textbook of occupational ergonomics* (4th ed.) Taylor & Francis/Hemisphere.
- Hancock, P. A. (2017). Whither Workload? Mapping a Path for Its Future Development. In: *International Symposium on Human Mental Workload: Models and Applications* (pp. 3-17). Springer, Cham.
- Hess, E. H.; Polt, J.M. (1964) Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science*, *143*, 1190–1192.
- Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1), 73-93.
- Hockey, G. R. J. (2011). A motivational control theory of cognitive fatigue. Cognitive fatigue: Multidisciplinary perspectives on current research and future applications, 167-87.
- Hockey, R. (2013). *The psychology of fatigue: Work, effort and control* Cambridge University Press.
- Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological psychology*, *110*, 100-106.
- Hopstaken, J. F., van der Linden, D., Bakker, A. B., Kompier, M. A., & Leung, Y. K. (2016). Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. *Journal of Experimental Psychology. Human Perception and Performance*, doi:2016-01220-001 [pii]
- Horiuchi, R., Ogasawara, T., & Miki, N. (2017). Fatigue evaluation by detecting blink behaviour using eyeglass-shaped optical sensor system. *IET Micro & Nano Letters*, *12*(8), 554-559.
- Ishii, A., Tanaka, M., Shigihara, Y., Kanai, E., Funakura, M., & Watanabe, Y. (2013). Neural effects of prolonged mental fatigue: A magnetoencephalography study. *Brain Research*, *1529*, 105-112. doi:10.1016/j.brainres.2013.07
- Jainta, S., & Baccino, T. (2010). Analyzing the pupil response due to increased cognitive demand: An independent component analysis study. *International Journal of Psychophysiology*, *77*(1), 1-7.
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*(1), 2.

- Johnson, S. K., DeLuca, J., Diamond, B. J., & Natelson, B. H. (1998). Memory dysfunction in fatiguing illness: Examining interference and distraction in short-term memory. *Cognitive Neuropsychiatry*, 3(4), 269-285. doi.org/10.1080/135468098396099
- Johnson, S. K., Lange, G., DeLuca, J., Korn, L. R., & Natelson, B. (1997). The effects of fatigue on neuropsychological performance in patients with chronic fatigue syndrome, multiple sclerosis, and depression. *Applied Neuropsychology*, 4(3), 145-153. doi:10.1207/s15324826an0403\_1
- Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154, 1583-1585.
- Kanfer, R. (2011). Determinants and consequences of subjective cognitive fatigue. In: L. P. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 189-208). Washington, DC: American Psychological Association.
- Krupp, L. B., & Elkins, L. E. (2000). Fatigue and declines in cognitive functioning in multiple sclerosis. *Neurology*, 55(7), 934-939. doi.org/10.1212/WNL.55.7.934-a
- Kohl, A. D. (2016). *Understanding mental fatigue: A grounded theory study* (Doctoral dissertation, Saybrook University).
- LeDuc, P. A., Greig, J. L., & Dumond, S. L. (2005). Involuntary eye responses as measures of fatigue in US army apache aviators. *Aviation, Space, and Environmental Medicine*, 76(Supplement 1), C91.
- Lemonnier, S., Brémond, R., & Baccino, T. (2014). Discriminating cognitive processes with eye movements in a decision-making driving task. *Journal of Eye Movement Research*, 7(4).
- Lorist, M. M., Boksem, M. A., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research*, 24(2), 199-205.
- Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior research methods*, 50(1), 94-106.
- Nakayama, M.; Takahashi, K.; Shimizu, Y. (2002). The act of task difficulty and eye-movement frequency for the "Oculo-motor indices.". In: *Proceedings of the*

*Symposium on Eye Tracking Research & Applications—ETRA'02*, New Orleans, Louisiana, ACM Press: New York, USA, pp. 37–42.

- Nie, Nie Xin Xin (聂欣 nie xin). (2005). Quantitative assessment of mental fatigue for human body based on non-linear analysis of physiological signals (Doctoral dissertation).
- Paul, R. H., Beatty, W. W., Schneider, R., Blanco, C. R., & Hames, K. A. (1998). Cognitive and physical fatigue in multiple sclerosis: Relations between self-report and objective performance. *Applied Neuropsychology*, 5(3), 143-148. doi:10.1207/s15324826an0503\_5
- Roy, R. N., Charbonnier, S., & Bonnet, S. (2014). Eye blink characterization from frontal EEG electrodes using source separation and pattern recognition algorithms. *Biomedical Signal Processing and Control*, 14, 256-264.
- Sakamoto, K., Aoyama, S., Asahara, S., Mizushina, H., & Kaneko, H. (2009, July). Relationship between emotional state and pupil diameter variability under various types of workload stress. In *International Conference on Ergonomics and Health Aspects of Work with Computers* (pp. 177-185). Springer, Berlin, Heidelberg.
- Sauer, J., Wastell, D. G., Robert, G., Hockey, J., & Earle, F. (2003). Performance in a complex multiple-task environment during a laboratory-based simulation of occasional night work. *Human Factors*, 45(4), 657-670.
- Shen, K. Q., Li, X. P., Ong, C. J., Shao, S. Y., & Wilder-Smith, E. P. (2008). EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clinical Neurophysiology*, 119(7), 1524-1533.
- Smets, E. A., Garssen, B., Bonke, B., & de Haes, J. M. (1995). The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research*, 39(3), 315-325. doi:10.1016/0022-3999(94)00125-O
- Stern, J. A. (1997). The pupil of the eye: What can it tell us about 'mental processes'. *Human Engineering for Quality Life (HQL) Quarterly*, 8, 1-2.
- Van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: Effects on perseveration and planning. *Acta Psychologica*, 113(1), 45-65. doi:10.1016/S0001-6918(02)00150-6
- Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors*, 43(1), 111-121.



- Venables, L., & Fairclough, S. H. (2009). The influence of performance feedback on goal-setting and mental effort regulation. *Motivation and Emotion*, 33(1), 63-74.
- Vrijkotte, S., Meeusen, R., Vandervaeren, C., Buysse, L., Cutsem, J. V., Pattyn, N., & Roelands, B. (2018). Mental Fatigue and Physical and Cognitive Performance During a 2-Bout Exercise Test. *International journal of sports physiology and performance*, 13(4), 510-516.
- Walker, L. S., Berard, J. A., Berrigan, L. I., Rees, L. M., & Freedman, M. S. (2012). Detecting cognitive fatigue in multiple sclerosis: Method matters. *Journal of The Neurological Sciences*, 316(1-2), 86-92. doi:10.1016/j.jns.2012.01.021
- Yokoyama, H., Eihata, K., Muramatsu, J., & Fujiwara, Y. (2018). Prediction of Driver's Workload from Slow Fluctuations of Pupil Diameter. In: 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 1775-1780). IEEE.
- Zhang, C., Zheng, C., Ouyang, Y., & Yu, X. (2008). Analysis of mental fatigue basing on power spectrum features of EEG. *Space Medicine and Medical Engineering*, 21(1), 35.
- Zhang, C., Zheng, C., Yu, X., & Li, X. (2006). Analysis of Mental Fatigue Based on Approximate Entropy of EEG. *Space Medicine and Medical Engineering*, 19(6), 459.
- Zhang, C., & Yu, X. (2010). Estimating mental fatigue based on electroencephalogram and heart rate variability. *Polish Journal of Medical Physics And Engineering*, 16(2), 67-84.

(Manuscript received: 24 May 2019; accepted: 16 February 2020)

## **Chapter VI**

---

*Fundamental Frequency as an Alternative  
Method for Assessing Mental Fatigue*



# **Fundamental frequency as an alternative method for assessing mental fatigue**

Enrique Muñoz-de-Escalona<sup>1</sup>

José Juan Cañas<sup>1</sup>

Jessica F. Morales Guamán <sup>1</sup>

<sup>1</sup> Mind, Brain and Behaviour Research Centre,  
Department of Experimental Psychology, Faculty of Psychology,  
University of Granada

---

This article has been published: Muñoz-de-Escalona, E., Cañas, J. J., & Morales-Guaman, J. F. (2020, December). Fundamental Frequency as an Alternative Method for Assessing Mental Fatigue. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 58-75). Springer, Cham. [https://doi.org/10.1007/978-3-030-62302-9\\_4](https://doi.org/10.1007/978-3-030-62302-9_4)

## Abstract

The development of methods for measuring mental workload and fatigue is still a central issue in the research agenda of Human Factors and Ergonomics researchers. There exists several different methodologies which offers indexes which are sensitive to mental workload and fatigue changes. One of these methodologies, the physiological measures offers many advantages, but comes with the main disadvantages of being very intrusive and require expensive and specialized equipment. To overcome this disadvantages, we propose the use of voice parameters, particularly the fundamental frequency, as an alternative valid method for assessing mental fatigue in complex and dynamic situations. In this study we tested our proposal. We manipulated the time during which participants performed the MATB-II task battery with high and constant task demands in order to induce mental fatigue over time; while subjective reports of fatigue, performance, pupil size and voice speech were collected as our dependent variables. Results showed that all of this measures, including the fundamental frequency obtained from participants' voice, were sensitive to the emergence of mental fatigue. High correlations among all measurements also confirmed the appropriate use of fundamental frequency as an alternative sensitive measure for assessing mental fatigue.

**Keywords:** mental fatigue, mental workload, fatigue measures, fundamental frequency.

## 1. Introduction

Research over the past 50 years has clearly shown that performance on complex tasks depends, among many other factors, on the mental resources that are demanded by tasks, but also on the operators' available mental resources to cope with those demands. The relation between demanded and available resources is what we call mental workload [1-4]. This relationship can be expressed as a fraction where the numerator is the amount of demanded resources and the denominator is the amount of available resources. When this fraction is greater than one, we say that we are in an overload condition, in which demanded resources are higher than available resources. If, on the contrary, the result of the fraction is less than one we have an underload condition, in which available resources are higher than task demands. The conditions of extreme overload and extreme underload negatively affect the performance of the task and the operator's welfare [1-4]. For this reason, human work is designed to avoid these situations. It should be noted that performance would not be impaired with intermediate levels of mental workload [5].

An overload condition is mainly caused by an excessive demand on resources when, for example, the task is very complex. However, overloading can also occur when there is a decrease in available resources as a result of the emergence of mental fatigue (during long periods performing the tasks) or when the tasks do not change the objectives for a long time [6,7]. For this reason, estimating the amount of available resources at any given time has been and continues to be a very important research topic in the area of Human Factors and Ergonomics. Therefore, in the agenda of researchers who study the variables that affect performance in a complex task, there is one main objective which is to develop methods that would allow us to measure the mental resources that a person has at a given time. However, all the methods that have been proposed have advantages but also have disadvantages and more research is still needed to find methods to measure available resources that would avoid them.

In this paper we will describe an experimental study carried out to explore the possibility of estimating the available resources from an acoustic parameter of the voice, its fundamental frequency (F0). Previous research has shown that the fundamental frequency of operators' voice increases as mental workload gets higher [8]. We hypothesize that this increase in fundamental frequency that occurs with time on task may be due to a decrease in available resources because of the gradual emergence of mental fatigue. In the experiment we manipulated the time spent performing the task while

keeping high constant task demands. We measured the participants' performance, their subjective estimation of fatigue, their pupil diameter and the fundamental frequency of their voice. We expected to find that as participants spent more time performing the task, their execution would worsen, their subjective perceptions of fatigue would increase, their pupil would shrink reflecting less physiological activation, and especially the fundamental frequency would increase.

In Section 2 we will review the different methodologies for measuring mental fatigue to focus, afterwards, in the acoustic parameters, particularly in the fundamental frequency. Section 3 describes the design and methodology used in our study. In Section 4 we describe the obtained results and in Section 5 the discussion about our findings and future possible research in this line. Finally, in Section 6 we conclude the study and provide implications about using fundamental frequency as a mental fatigue index.

## 2. Related Work

### 2.1 Methods for measuring available resources

Over the past few decades, researchers have designed several methods to measure the mental resources a person has while performing a task. All these methods can be classified in various ways according to various criteria, but a first criterion that is often used is the temporary moment at which the amount of resources is measured. According to this criterion we can distinguish between "off-line methods" and "on-line methods".

Off-line methods basically consist of estimating the amount of mental resources once the task has been completed. These methods have the advantage that they normally do not interfere with the performance of the task because they are collected at the end of the task, but have the disadvantage that they do not allow an estimation during the performance of the task. Sometimes, what we are interested in, is how many mental resources a person has left after the task is completed. For example, we would be in that situation when we want to know how tired a person is at the end of his or her working day and not how the mental resources have been managed during the performance of his tasks at work. In these situations, off-line methods would be appropriate for our purposes. However, if our objectives were to estimate mental resources during task performance, we would need to use the so-called on-line methods.

Regardless of whether the evaluation is off-line or on-line, the methods can be classified into three big categories. The first category of methods is the so-called "concurrent task performance-based methods" which consist of asking the person to perform a task that is called the secondary task while he or she performs his or her primary task, which is the one he or she has to perform at work. Basically, these methods measure the performance of the secondary task and apply the following logic: since both the primary task and the secondary task have to share the same resources, the performance of the secondary task is a reflection of how many resources are "left over" after performing the primary task [4]. If we observe that the secondary task is performed very well, we can interpret that the primary task requires few resources. Conversely, if the secondary task is poorly performed, it will mean that the primary task is consuming most of the resources. These performance-based methods are good on-line methods for diagnosing the causes of overload or underload. However, these methods have two fundamental drawbacks. First, in real situations (not in the laboratory or in a simulator) sometimes we cannot afford to have a secondary task drawing resources that may be needed to perform well the primary task (which is the main task the person has to perform). Secondly, it is often difficult to ask the person whose mental resources we are estimating to understand that he or she has to "prioritize" the primary task over the secondary task.

The second category of methods is made up of subjective methods that consist of asking the operators themselves to give a subjective estimation of their current available mental resources, normally using a numerical interval scale where the available mental resources can be expressed from less to more during the performance of the task. The main disadvantage of these methods is that they can interfere with the completion of the task when performed on-line. Although most of these methods only require operators to verbally give their estimation of the amount of mental resources, doing so may distract their attention from the task and impair their performance. Therefore, they are usually used off-line at the end of the task or working day.

Finally, the third category of methods, which are known as psychophysiological methods, are based on our knowledge about the relationship between mental resources and certain psychophysiological parameters. There are currently several psychophysiological methods, all of these with advantages and disadvantages. Most of these psychophysiological recording methods that are currently being proposed (records of ocular parameters such as blink rate, pupil diameter, records of the electrical activity of the cerebral cortex, etc.) require high technical expertise and the use of some particular



recording equipment, which is expensive and needs the operator to be equipped with it so that the physiological parameters can be properly recorded. For example, records of ocular parameters require the operator to be equipped with special eye-trackers, and the records of cortical electrical activity have to be done by equipping the operator with a set of electrodes on the head that can be annoying and interfere with the task. Therefore, there is currently a great need to explore other psychophysiological methods for the estimation of mental resources, especially on-line, which meet two main requirements: they should not require expensive recording equipment and they do not need to ask the operator to be equipped with it during the performance of his task. The methods that we should explore must record psychophysiological parameters in a natural way from some aspect of the operators' open behavior while she or he is performing the task. One of these methods may be the estimation of mental resource states from acoustic parameters of the voice. In many situations where it is necessary to estimate the amount of mental resources that a person has, verbal communications are (or can be) recorded automatically, often for security reasons. One of these situations is the one we have at Air Traffic Management (ATM) [9].

## **2.2 Effects of mental workload on the acoustic parameters of the voice**

In the communication between two people it is very important that both are able to perceive the emotional state of the other. For this perception of the emotional state of the person we are communicating with, we can use several signals, the most important of which is the gesture of the face [10]. We humans manifest our internal psychological states very well through gestures of the face. However, another characteristic of human communication that also shows the internal psychological state is the voice.

In the field of ATM, the relationship between mental states and communications has been known for a long time. For example, Bailey, Willems and Peterson [11] found that the number of communications between route controllers and pilots increased in high mental load conditions. However, the current interest in this area is primarily to explore the relationship between the acoustic characteristics of verbal communications and the mental states. In this line, many of the first studies that were carried out on the acoustic parameters of the voice were aimed at exploring the relationship between these parameters and stress. In an internal EUROCONTROL report, Hagmüller, Kubin and Rank [12] reviewed the empirical evidence on the influence of stress caused by mental

workload on the acoustic parameters of the voice. The authors argue that since voice is a common communication tool of Air Traffic Controllers (ATCOs), the analysis of its acoustic characteristics can be very useful to identify its psychological states during the performance of the control task. The authors considered mental workload as a "psychological" stress-causing factor that is different from other physical (e.g., vibrations produced by machines), physiological (e.g., illness, lack of sleep), or perceptual (e.g., noise) factors. One of the most important conclusions of the review conducted by these researchers was that the Fundamental Frequency is sensitive to changes in stress caused by mental workload.

The fundamental frequency is the lowest frequency of a periodic waveform. In the case of the human voice, when a person produces a sound, the vocal cords vibrate at a certain speed by rapidly opening and closing with small bursts of air. In this way the sound produced is composed of a spectrum of frequencies that we can break down to obtain the lowest frequency of the spectrum. That lower frequency of the produced sound is what we call the fundamental frequency. We should not confuse the fundamental frequency with the tone of a sound even though they are related. Tone is a unit of measurement for "perception". We could say that fundamental frequency is a "physical" measure of sound that can be obtained by placing a wide-range microphone directly in the throat, above the vocal cords, but below the resonant structures of the vocal tract. While pitch is a "psychological" measure that reflects how frequencies are perceived in the human nervous system and is measured by the Mel Scale. The name Mel comes from the word melody to indicate that it is based on human perception of tone. The relationship between fundamental frequency and tone would be similar to the relationship between physical sound intensity and the decibel scale which is a perceptual psychological measure of intensity. In any case, since the relationship between frequency and tone is a complex relationship that leads us to the field of Psychophysics, researchers have preferred to measure fundamental frequency directly in their research on the relationship between mental states and the physical parameters of the voice.

Effects on fundamental frequency have been consistently found in studies where participants were asked to identify whether or not a person was stressed by hearing what this person said. In these studies, researchers have found that the characteristic that listeners use to make a meaningful classification is Fundamental Frequency [13,14]. These studies soon distinguished between the different causes of stress. Although, in a general way we can say that stress is a psychological and physiological response of the

organism to a danger, we can distinguish between the different types of dangers, called stressors, to which the organism responds. In this line, Scherer, Grandjean, Johnstone, Klasmeyer, and Bänziger pointed out that one thing is the stress produced by mental overload and another is the stress produced by other situations of danger [15]. Consider two examples: the stress of an ATCo may be caused by a high taskload, but it is also possible that the stress we observe in an ATCo is due to a situation of danger from an imminent conflict. In the latter case the observed stress has a causal component with an emotional component due to the perceived risk of collision. In the experiment carried out by these authors, the participants had to perform a logical reasoning task in one of two conditions (1) focusing on this task without any distraction; (2) performing the task while simultaneously attending to an auditory monitoring task. In the logical reasoning task, they had to make deductions based on premises that appeared on a screen. In the auditory monitoring task, they had to respond to one sound while ignoring another sound, both appearing randomly. The authors hypothesized that in the condition of the logical reasoning task there would only be a mental workload effect, while in the dual task, in addition to increasing this mental workload by having to perform two tasks simultaneously, there would be an effect of emotional stress due to time pressure and the same stressful noise defect. In this study, the participants were asked to say a few sentences appearing on the screen at intervals while they were performing the tasks. The duration of the intervals was random. These sentences were of the type "This is task number 345629". In the sentence, only the number changed from one time to the next. In this way, there was one part that was always fixed and another part that was changing, the number.

The results of the experiment showed a clear statistical independence of the effects of stress and workload on F0. The two effects did not interact, allowing the researchers to conclude that they can be considered independent effects. However, although the effect of stress was statistically significant, the effect of workload was only marginally so. The authors argued that this lack of significance was due to the large individual differences they found in the sample of participants. Based on the results of questionnaires that the participants had to fill out, the authors suggested that the participants with higher levels of anxiety caused by the desire to perform the task well were those who showed the most significant results on the effect of F0.

Huttunen, Keränen, Väyrynen, Pääkkönen, and Leino have conducted a study with military pilots in a flight simulator piloting task [16]. The three mental load conditions in

which the pilots performed their task were (1) mental workload due to the complexity of Situation Awareness; (2) mental workload due to the complexity and critical value of information; and (3) mental workload due to the difficulty of the decision making processes. Taking as a baseline the recordings collected in a period of time before the simulation, the authors calculated the differences in F0 as a function of mental workload in these three conditions. The results showed that F0 increased as a function of mental workload in all three conditions.

In a word recall task under controlled laboratory conditions, Boyer, Paubel, Ruiz, El Yagoubi and Daurat found that the F0 increased with the number of words participants had to remember [17]. The authors interpreted that the number of words can be considered a factor that increases cognitive demand and, therefore, its effect on the F0 can be interpreted as evidence of the relationship between mental workload and this acoustic parameter of the voice.

Therefore, we can say that there is experimental evidence of the effect of mental workload on an acoustic parameter of the voice, the fundamental frequency of people who are performing complex tasks. The fundamental frequency increases when the mental workload increases. However, it would be interesting to know if this effect of mental workload can be due to an increase on mental demands or a decrease in available resources. We believe that a decrease in the available resources due, for example, to the time spent performing the task may be also a factor affecting the observed increase in the Fundamental Frequency due to the emergence of mental fatigue.

### **2.3. Effects of fatigue on fundamental frequency**

Fatigue research has found results that show how acoustic parameters of the voice, such as fundamental frequency, can be sensitive to fatigue. For example, research in recent years has shown that fatigue can affect different phases of speech production. According to the review by Krajewski, Batliner, and Golz these phases would be as follows [18]:

1. Cognitive speech planning has been found to result in a slowdown in cognitive processing, impaired speech planning, impaired neuromuscular motor coordination processes, impaired fine motor control and slow movement of the articulator, and sluggish articulation and slower speech.

2. There are effects on breathing that manifest themselves in decreased muscle tension, flat and slow breathing, reduced subglottal pressure, and lower fundamental frequency, intensity, articulatory accuracy and joint rate.
3. The effects on phonation that have been found are as follows: decreased muscle tension, increased vocal cord elasticity and decreased vocal cord tension, decreased body temperature, change in vocal cord viscoelasticity, change in spectral energy distribution, breathable and lax voice, non-lifting larynx, decreased resonant (formant) frequency positions and increased formant bandwidth.
4. The effects on the articulation/resonance are decreased muscle tension, unrestricted pharynx and softening of the vocal tract walls, loss of speech signal energy, wider formant bandwidth; postural changes, lowering of the torso and head, change in the shape of the vocal tract, change in the position of the formant, increased salivation, loss of energy, decreased body temperature, reduced heat conduction, change in friction between the walls of the vocal tract and the air, changes in laminar flows, jet streams and turbulence, change in spectral energy distribution, increased formant bandwidth, increased formant frequencies especially in lower formants.
5. As for the radiation the effects are a decrease of the orofacial movement, facial expression and lip extension (visualization of open and relaxed mouth), lengthening of the vocal tract, lower positions of the first and second formant, reduction of the articulation effort, lower degree of openness, relaxed articulation, decrease of the first formant, oropharyngeal relaxation, decrease of the veil, coupling of the nasal cavity, increase of the nasality, band width of the extended Formant 1, lower amplitude of Formant 1.

Based on these known effects of mental fatigue on speech production, research has been aimed at identifying which acoustic parameters best reflect the effects of mental fatigue. In a study by Cho, Yin, Park, and Park, the authors found that when participants were divided into two groups according to a subjective fatigue scale, some acoustic parameters such as fundamental frequency shaking, brightness, HNR (the ratio of harmony to noise), SNR (the signal-to-noise ratio), and shaking amplitude were predictors of mental fatigue [19]. These results suggest that it would be possible to use some of these parameters to assess mental fatigue.

Whitmore and Fisher conducted a study in which a group of American bomber pilots, divided into groups of four, participated in a pilot task for periods of 36 hours in a flight simulator [20]. The piloting periods were interspersed with rest periods of 36 hours. Approximately every 3 hours the participants had to perform cognitive tasks, subjective fatigue assessments and repeat the sentence:

“Futility Magellan, this is xxx yyy. The time is zz:zz Zulu”.

where:

XX was the participant's rank

YY was the Participant's Name

ZZ:ZZ was the time when you say the phrases

The results showed that the fundamental frequency and duration of words were good indices of subjective fatigue and performance in the cognitive tasks.

Other researchers have conducted research using methodologies based on voice data recorded during conversations in a natural context. This is what Krajewski, Batliner, and Golz did, who conducted a validation experiment to examine whether automatically trained voice database models can be used to recognize subject drowsiness [21]. Their methodological approach can be summarized in four steps:

- They collected individual speech data as well as associated sleepiness scores for each subject.
- Then, they extracted relevant acoustic characteristics from the speech data.
- With that data, they constructed statistical models of sleepiness scores based on the acoustic characteristics.
- Finally, they tested learned models on new speech data.

With this methodological procedure the researchers did not need to conduct a study in which, in a controlled manner, a group of people were required to verbally express a certain text every certain period of time, as has been done in other empirical studies designed to study this same topic. In contrast, with the procedure used by these authors, the voice data were collected in the natural context in which the people performed their tasks. Pattern recognition algorithms were applied to the data collected with this procedure. These algorithms were trained to recognize basic acoustic characteristics

according to the acoustic-preceptual concepts of (1) Prosody (tone, intensity, rhythm, pause pattern and speed of speech); (2) Articulation (speech difficulty, reduction and elision); and (3) Speech quality (breathable, tense, high-pitched, hoarse or modal voice). The algorithms were also trained according to signal processing categories (time domain and frequency domain) and state space characteristics. With this procedure, the classification algorithms, working with an unusually large set of data, were able to determine whether a subject's sleep was beyond a critical threshold. Subsequently, the authors conducted a validation study with new data in which they achieved an accuracy rate of over 86% in unseen data, but from known speakers, with an SVM (artificial intelligence algorithm) classifier.

As this review showed, there is enough empirical evidence about the effect of mental fatigue on some acoustic parameters of the human voice. Empirical evidence on the effect of fatigue on the acoustic parameters of the voice supports the hypothesis that the effect of mental workload on these parameters may be reflecting a decrease in available resources. According to the definition of mental workload as a relationship that can be expressed as a division between demanded and available resources, a decrease in available resources due to factors such as fatigue may be the cause of the effect of the observed mental workload.

The aim of the experiment described below has been to test this hypothesis. In the experiment a group of participants performed a set of tasks in which their performance was measured at the same time as their pupil diameter, the subjective estimation of their fatigue and the fundamental frequency of their voice. The pupil diameter measurement was measured because it has proven to be a good index of the level of activation [22]. Task demands remained constant throughout the task. The variable that was manipulated was the time spent performing the task. We expected to find that the longer time performing the task, the greater the fatigue, causing a decrease in performance, a smaller pupil diameter, a greater subjective feeling of fatigue and, most importantly according to our hypothesis, an increase in the fundamental frequency of participant's voice.

### 3. Design and Methodology

#### 3.1 Materials and Instruments

**MATB-II Software.** The participants of the experiment had to perform the Multiple Attribute Task Battery (MATB-II) [23]. This task battery is designed to assess the performance and workload of operators by means of different tasks similar to those performed by flight crews. The software used to perform the tasks has a user-friendly interface that allows non-pilot participants to use it. MATB-II comes with default event files that can be easily modified to suit the needs or objectives of an experiment. The program records the events presented to the participants, as well as their responses. MATB-II contains the following four tasks: the System Monitoring Task (SYSMON), the Tracking Task (TRACK), the Communications Task (COMM), and the Resource Management Task (RESMAN) (see Figure 1).

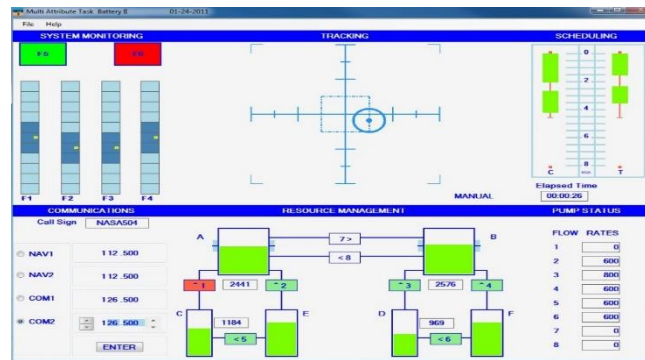
1) The task of SYSMON is divided into two sub-tasks: the lights task and the scales task. For the sub-task of lights, participants are required to respond as quickly as possible to a green light turning off and a red light turning on, and to turn the lights back on and off, respectively. For the scale sub-task, participants are asked to detect when the lights on four moving scales deviate from their normal position and respond accordingly by clicking on the deviated scale.

2) In the TRACK task, there are two modes. Participants can work only in a manual mode. During this manual mode, participants have to keep a circular target in the centre of an inner box using a joystick with their left hand (the dominant hand was necessary for the use of the mouse). During the automatic mode of the task, the circular target will remain in the inner box by itself.

3) In the COM task, an audio message with a specific callsign is displayed and the participant is asked to respond by selecting the appropriate radio channel and by setting the correct frequency, but only if the callsign matches his or her own (callsign: "NASA504"). The participant is not required to respond to messages in other callsigns.

4) In the RESMAN task, participants have to maintain the fuel level in tanks A and B, within  $\pm 500$  units of the initial condition of 2500 units each. To maintain this objective, participants must transfer the fuel from the supply tanks to A and B or transfer the fuel between the two tanks.





**Figure 1:** MATB-II task display. Taken from <https://matb.larc.nasa.gov/>

**Praat Software.** “Praat” is a scientific license free tool for analyzing spectrograms of audio records. It was developed at the University of Amsterdam by Paul Boersma and David Weenink in 1992 and it is constantly being updated with improvements implemented by authors, some of them suggested by users [24]. Once the audio file is loaded you can obtain multiple audio parameters such as fundamental frequency, intensity, volume, formants, etc. In this study we used Praat software for obtaining the fundamental frequency intervals average.

**Tobii T120 Eyetracker.** Pupil diameter measurements were obtained using an infrared eye tracking system with a sampling frequency of 120 Hz, the model Tobii T120 marketed by Tobii Video System. This equipment is completely non-intrusive, does not have a visible eye movement tracking system and provides high accuracy and an excellent compensating head movement mechanism, ensuring high quality data collection.

**Instantaneous Self-assessment Scale.** We employed an easy and intuitive instantaneous subjective fatigue scale called Instantaneous Self-assessment, which provides momentary subjective ratings of perceived mental fatigue during task performance. Participants evaluated the mental fatigue they experienced at any given time on a scale ranging from 1 (no mental fatigue) to 9 (maximum mental fatigue). Participants were taught to use the scale just before starting the experimental stage.

### 3.2 Participants

17 students from the University of Granada participated in the study. Participants’ ages ranged from 18 to 32, with an average of 23.6 and a standard deviation of 2.25. A total of

13 women and 4 men participated. It should be noted that there is a greater number of female participants due to the fact that psychology students at the University of Granada are mostly women. Recruitment was achieved through the dispersion of posters and flyers around the university, as well as an advertisement for the study on the university's on-line platform for experiments (<http://experimentos.psiexpugr.es/>). The requirements for participation included (1) not being familiar with the MATB-II program, (2) Spanish as a native language, and (3) visual acuity or correction of visual impairment with contact lenses, as glasses impair the utilized eye-tracking device from collecting data. Participants' participation was rewarded with extra credit.

### 3.3 Procedure

The participants went through an experimental session consisting of two phases:

3. **Training stage:** training took place for no longer than 30 minutes. The objective of this stage was for participants to familiarize themselves with the program so that they could carry out the tasks securely during the data collection stage. The procedure was conducted as follows: upon entering the lab and after filling out the informed consent form, the participant was instructed to read the MATB-II instruction manual and inform the researcher once they had finished. The researcher then sat down with the participant to allow for questions and resolve any doubts on how to use the program. Afterward, on a computer monitor, participants were presented with each MATB-II task separately and were first given a demonstration as to how to execute the task and after what they were given time (3 minutes or more if needed) to perform the task themselves. The participants were always free to consult the manual and ask the researcher questions during the training stage in case of doubts or uncertainties. Once the participants had completed all four tasks and resolved all doubts, they were ready for the data collection stage, which followed immediately afterwards. During the training stage, participants could work in one room equipped for training with the MATB-II software, and no special attention to room conditions was needed.
4. **Data collection stage:** the data collection stage lasted a period of 60 minutes that was divided into thirty intervals of 2 minutes. The first interval was left as a training

interval in order to focus the attention of participants into the task. During this first 2 minutes interval only the tracking task was activated. Then, during the second interval and until the end of the experiment, the three tasks of the MATB-II software were activated (TRACK, SYSMON and RESMAN), thus mental workload level was high and constant throughout the experimental session in order to facilitate the emergence of mental fatigue. The participants were instructed to verbalize the ISA (Instantaneous Self-Assessment) scale every 2 minutes when a scheduled alarm sounded. They were also instructed to verbalize every action they were performing in the SYSMON (e.g. “I press F3 button”) and the RESMON (e.g. “I activate fuel pump n°1”) tasks in order to collect our F0 variable through audio recordings. Prior to the start of the task-battery, the eye-tracker system was calibrated, and the participants were told to keep head and body movements to a minimum. During the data collection stage, standardizing room conditions was essential. Thus, the testing rooms were temperature controlled to 21°C, and lighting conditions (the main extraneous variable in pupil diameter measurement) were kept constant with artificial lighting; there was no natural light in the rooms. Moreover, participants always sat in the same place, a comfortable chair spaced 60 cm from the eye-tracker system.

This study was carried out in accordance with the recommendations of the local ethical guidelines of the committee of the University of Granada institution: “Comité de Ética de Investigación Humana”. The protocol was approved by the “Comité de Ética de Investigación Humana” under the code: 779/CEIH/2019. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

### **3.4 Variables**

#### **Independent Variable:**

The only independent variable was the time spent performing the task. The period of one hour performing the task was divided into 2-minute intervals resulting in 30 intervals. The first period of 2 minutes was taken as a baseline to calculate the pupil diameter so it

will not be analyzed and we are left with only 29 intervals that constitute the 29 levels of our independent variable "intervals" during the time performing the task.

### **Dependent Variables:**

*Performance.* MATB-II provides us with many indicators of participants' performance: e.g. root mean square deviation (RMSD) for the TRACK task, number of correct and incorrect responses for the SYSMON and COMM tasks, and the arithmetic mean of tanks "A-2500" and "B-2500" in absolute values for the RESMAN task. However, for the purposes of this experiment we will only consider the RMSD performance indicator. The RMSD performance indicator reflects the distance of the circle to the target point, in such a way that, performance impairment is reflected by a higher score on this variable.

*Pupil size.* Fatigue can be estimated by several physiological indexes such as EEG, HVR, and several ocular metrics. We decided to use pupil diameter as our physiological fatigue indicator, as it effectively reflects mental fatigue [19] and minimize intrusiveness. While our eye-tracking system allows continuous sampling rate recording at 120Hz, we set a total of 30 intervals lasting 2 minutes each. Since expressing pupil size in absolute values has the disadvantage of being affected by slow random fluctuations in pupil size (source of noise), we followed a procedure for standardizing the values of pupil size for each participant.. To do this, for every participant, we took his/her pupil size value during the first interval of 2 minutes, and then subtracted it from the obtained value in each of the rest of 29 intervals, thereby giving a differential standardized value allowing us to reduce noise in our data. Analyses were carried out for the average of both the left and right pupils. A negative value meant that the pupil was contracting while a positive value meant that it was dilating.

*Subjective fatigue.* We used an on-line subjective fatigue scale created for this purpose, the "Instantaneous Self-Assessment Scale". Ratings were obtained at 2 minute intervals throughout the 60 minutes of the experimental stage, obtaining a total of 30 subjective mental fatigue ratings. The rating from the first 2 minutes interval was discarded for the analysis.

*Fundamental Frequency (F0).* We recorded the verbalizations of participants through the microphone integrated in our eye-tracker system during the data collection stage. The 60 minutes obtained audio file was divided into 30 intervals which were analyzed with the “Praat” software in order to obtain the average F0 of each interval.

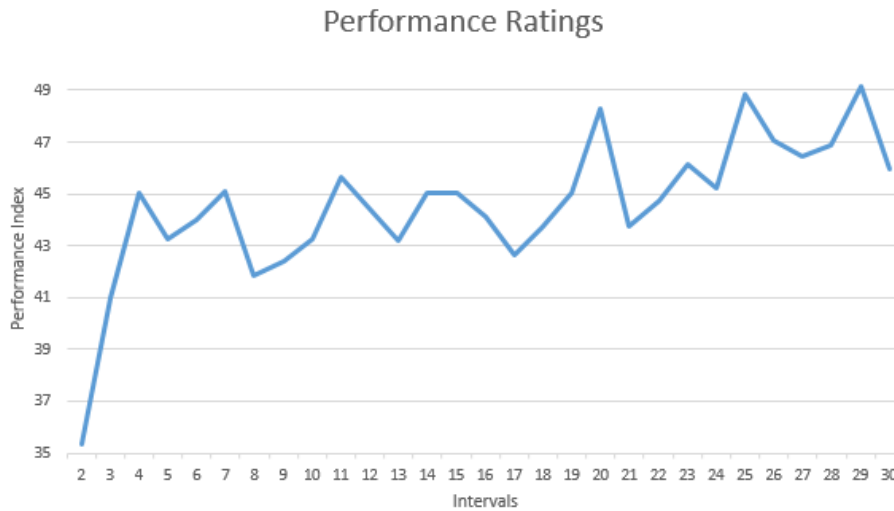
*Synchronization of measures.* Performance, pupil size, subjective, and fundamental frequency measures were obtained continuously throughout the experimental session. Synchronization between measures was simple, as the eyetracker (pupil and F0) and MATB-II performance log files began to record data simultaneously at the start of the experimental session. The scheduled alarm (every 2 minutes) was also synchronized by the experimenter, as it was simultaneously activated with the MATB-II software. This would also allow the ISA scale to be synchronized with the performance, F0 and pupil size measures.

### **3.5. Experimental design**

The experimental design was One-Way with Intervals as the only within-subject variable with 29 experimental levels.

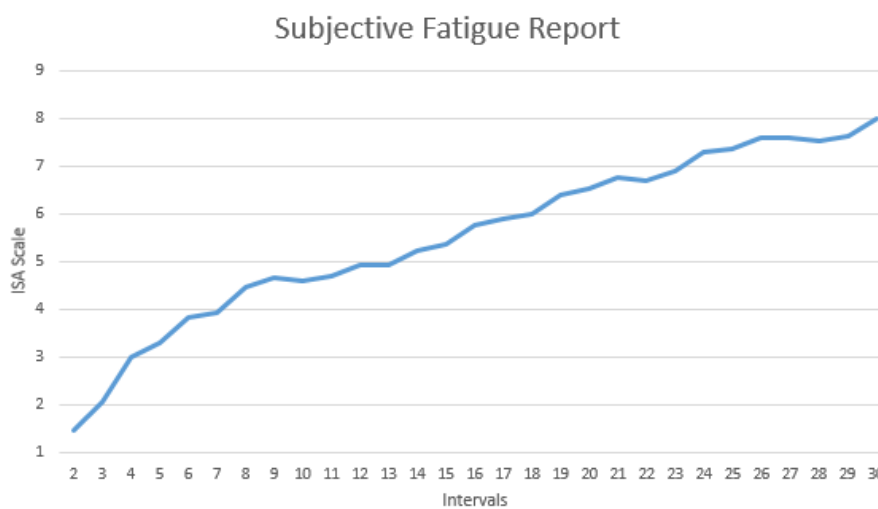
## **4. Results**

We performed four one-way within-subject ANOVA to analyze collected results, one for each different measure. Firstly, the ANOVA for our performance variable revealed a significant effect of intervals  $F(28,448) = 2.27$ ,  $MSe = 52.71$ ,  $p < .01$ . As we can see in the graph, performance worsened considerably during the first intervals (2 to 4), and then, it stabilized. However, we can appreciate that it continued worsening through intervals until the end of the experiment. Trend analysis confirmed this statements, as linear trend found to be significant,  $F(1,16) = 5.02$ ,  $MSe = 330.21$ ,  $p < .05$  (see Figure 2).



**Figure 2:** Participants’ performance during task development.

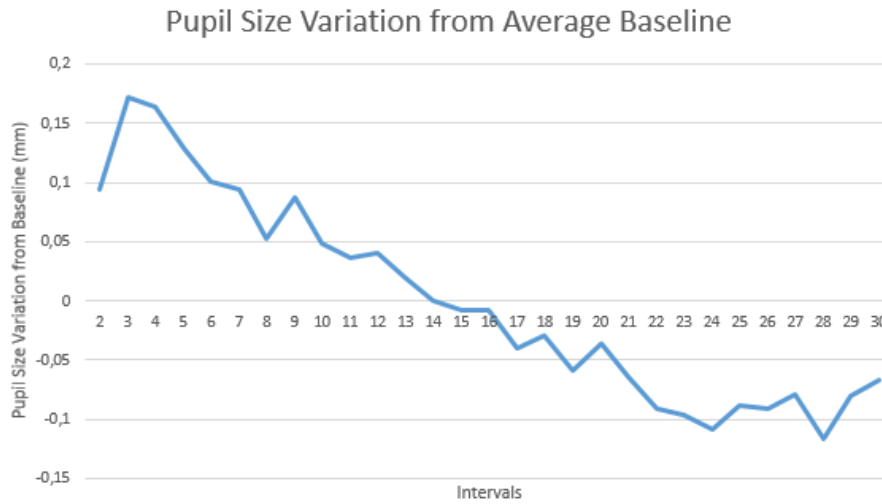
Regarding subjective mental fatigue scores, we found a linear increase through intervals. The ANOVA for this variable revealed a significant main effect of intervals  $F(28,448) = 66.14$ ,  $MSe = .79$ ,  $p < .01$ . Trend analysis supports the linear increase of mental fatigue through intervals,  $F(1,16) = 189.87$ ,  $MSe = 7.34$ ,  $p < .01$  (see Figure 3).



**Figure 3:** Participants’ subjective mental fatigue ratings during task development.

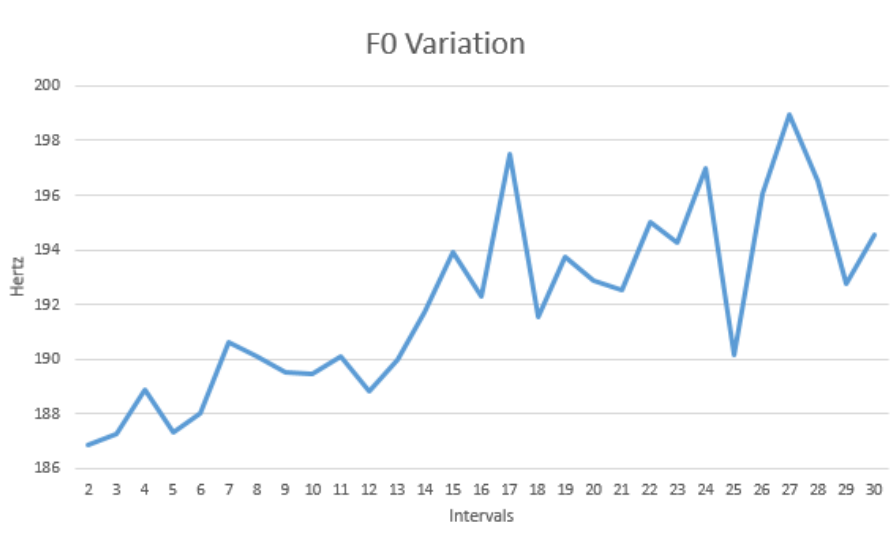
With respect to our first psychophysiological variable, pupil size, the ANOVA showed a significant effect of intervals  $F(28,448) = 20.63$ ,  $MSe = .01$ ,  $p < .01$ . The graph revealed a sudden dilation in pupil size from interval 2 to 3 and then, we can appreciate a linear decrease through intervals which tends to stabilize during the last 6 intervals. Trend

analysis also revealed a linear trend for this variable,  $F(1,16) = 33.0$ ,  $MSe = .1$ ,  $p < .01$  (see Figure 4).



**Figure 4:** Participants' pupil size variation ratings from the average baseline during task development.

The ANOVA analysis for our second psychophysiological variable, F0, revealed again a significant main effect of intervals  $F(28,448) = 2.04$ ,  $MSe = 91.33$ ,  $p < .01$ . We can see in the graph that F0 increased linearly through intervals, as trend analysis revealed,  $F(1,16) = 6.35$ ,  $MSe = 540.69$ ,  $p < .05$  (see Figure 5).



**Figure 5:** Participants' F0 during task development.

Finally, the correlation chart in Table 1 revealed significant very high correlations among measurements: we found a positive correlation between subjective fatigue and performance .73,  $p < .01$ ; a negative correlation between subjective fatigue and pupil -.94,  $p < .01$ ; a positive correlation between subjective fatigue and F0 .83,  $p < .01$ ; a negative correlation between performance and pupil size -.65,  $p < .01$ ; a positive correlation between performance and F0 .56,  $p < .01$  and a negative correlation between pupil and F0 -.86,  $p < .01$ .

**Table. 1.** Correlation chart between measures.

		Performance	Sub. Fatigue	Pupil	F0
<b>Performance</b>	Spearman Correlation	1	.725**	-.652**	.564**
	Sig. (bilateral)		,000	,000	,001
	N	29	29	29	29
<b>Sub. Fatigue</b>	Spearman Correlation	.725**	1	-.938**	.834**
	Sig. (bilateral)	,000		,000	,000
	N	29	29	29	29
<b>Pupil</b>	Spearman Correlation	-.652**	-.938**	1	-.856**
	Sig. (bilateral)	,000	,000		,000
	N	29	29	29	29
<b>F0</b>	Spearman Correlation	.564**	.834**	-.856**	1
	Sig. (bilateral)	,001	,000	,000	
	N	29	29	29	29

\*. The correlation is significant at the level .05 (bilateral).

\*\*.. The correlation is significant at the level .01 (bilateral).



## 5. Discussion

The results show a clear effect of time performing the tasks on every dependent variables: subjective fatigue, task performance, pupil diameter and the fundamental frequency of participants' voice. Particularly, the more time participants spent performing the tasks the worse their performance got, the greater their reports of subjective mental fatigue were, the more contracted their pupil got (indicating that their activation level was decreasing) and the higher the fundamental frequency in their voice got as well.

Therefore, considering the above mentioned results we can say that our hypothesis was fulfilled, and the results significantly showed that the fundamental frequency can also be considered an appropriate alternative index to other measures of mental fatigue such as EEG [25,26], HRV [27,28], blink rate [29]. The emergence of mental fatigue would be caused by a decrease in available resources (as our pupil size variable reflected) resulting from resources' depletion caused by time on task. These results were in line with literature research and proved that the fundamental frequency can be used as an appropriate index of mental fatigue. It is especially important to note that all measures of mental fatigue correlated significantly in the appropriate direction indicating that participants showed the effects of mental fatigue throughout the task on the one hand, and that every mental fatigue measure in this study was found to be sensitive to mental fatigue as well.

However, we must consider that these results should be viewed under certain methodological limitations. First, our sample was small and only composed of young university students, so we believe that these findings should also be tested under working situations in real contexts in order to verify that the fundamental frequency can be used out of the laboratory with satisfactory results, also we must take into account that the number of levels of our independent variable is higher than the number of participants and it might affect to the statistical power of the ANOVA, so our findings must be taken cautiously. Finally, it would also have been interesting to consider different time periods, some of them longer and shorter than 60 minutes and also under different task demand levels. Further research should consider this methodological modifications in order to analyse the sensitivity of the fundamental frequency to the emergence of mental fatigue to different degrees.

## 6. Conclusions

The aim of this research was to provide empirical evidence in favour of the hypothesis on the effect that mental fatigue has on an important parameter of the human voice such as the fundamental frequency. The results obtained showed that indeed, the fundamental frequency increases with fatigue. These results allow us to propose the fundamental frequency as an appropriate psychophysiological index of fatigue that may be a viable alternative to other psychophysiological indices that have been proposed. The advantage of fundamental frequency over the other psychophysiological indices is that it does not require expensive and intrusive equipment to record it. The human voice can be recorded in a natural and easy way in many tasks where operators have to communicate to perform their tasks. A direct implication would be that mental fatigue could be predicted at certain works (such as ATM) directly from human voice records, in an automated, economic and reliable way. Also, historical voice records could be used to analyse the mental fatigue that operators were experiencing at a given time or period to, for example, analyse work-related accidents or productivity at work.

## References

1. Longo, L., & Leva, M. C. (Eds.). (2017). Human Mental Workload: Models and Applications: First International Symposium, H-WORKLOAD 2017, Dublin, Ireland, June 28-30, 2017, Revised Selected Papers (Vol. 726). Springer.
2. Longo, L., & Leva, M. C. (Eds.). (2019). Human Mental Workload: Models and Applications: Second International Symposium, H-WORKLOAD 2018, Amsterdam, the Netherlands, September 20-21, 2018, Revised Selected Papers (Vol. 1012). Springer.
3. Longo, L., & Leva, M. C. (Eds.). (2019). Human Mental Workload: Models and Applications: Third International Symposium, H-WORKLOAD 2019, Rome, Italy, November 14-15, 2019, Proceedings (Vol. 1107). Springer Nature.
4. Wickens, C. D. (2017, June). Mental workload: assessment, prediction and consequences. In International Symposium on Human Mental Workload: Models and Applications (pp. 18-29). Springer, Cham.
5. Hancock, P. A. (2017, June). Whither workload? Mapping a path for its future development. In International Symposium on Human Mental Workload: Models and Applications (pp. 3-17). Springer, Cham.

6. Desmond, P. A., Neubauer, M. C., Matthews, G., & Hancock, P. A. (Eds.). (2012). *The handbook of operator fatigue*. Ashgate Publishing, Ltd.
7. Hockey, B., & Hockey, R. (2013). *The psychology of fatigue: Work, effort and control*. Cambridge University Press.
8. Haggmüller, M., Kubin, G., & Rank, E. (2005). Evaluation of the human voice for indications of workload induced stress in the aviation environment. EUROCONTROL
9. Luig, J., & Sontacchi, A. (2010, September). Workload monitoring through speech analysis: Towards a system for air traffic control. In *27th Congress of The International Council of Aeronautical Sciences* (pp. 1-10).
10. X Belin, P., Campanella, S., & Ethofer, T. (Eds.). (2012). *Integrating face and voice in person perception*. Springer Science & Business Media.
11. Bailey, L. L., Willems, B. F., & Peterson, L. M. (2001). The effects of workload and decision support automation on enroute r-side and d-side communication exchanges. FEDERAL AVIATION ADMINISTRATION OKLAHOMA CITY OK CIVIL AEROMEDICAL INST.
12. Haggmüller, M., Kubin, G., & Rank, E. (2005). Evaluation of the human voice for indications of workload induced stress in the aviation environment. EUROCONTROL
13. Protopapas, A. and Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. *Journal of the Acoustical Society of America*, 101(4):2267–2277.
14. Zhou, G., Hansen, J., and Kaiser, J. (2001). Non-linear feature based classification of speech under stress. *IEEE Transactions on Speech and Audio Processing*, 9(3):201–216.
15. Scherer, K. R., Grandjean, D., Johnstone, T., Klasmeyer, G., and Bänziger, T. (2002). Acoustic correlates of task load and stress. In *Proceedings of the International Conference on Spoken Language Processing*, pages 2017–2020, Denver, Colorado, USA.
16. Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., & Leino, T. (2011). Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. *Applied ergonomics*, 42(2), 348-357.
17. Boyer, S., Paubel, P. V., Ruiz, R., El Yagoubi, R., & Daurat, A. (2018). Human voice as a measure of mental load level. *Journal of Speech, Language, and Hearing Research*, 61(11), 2722-2734.

18. Krajewski, J., Batliner, A., & Golz, M. (2009). Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach. *Behavior Research Methods*, 41(3), 795-804. doi:10.3758/BRM.41.3.795
19. Cho, S., Yin, C. S., Park, Y., & Park, Y. (2011). Differences in self-rated, perceived, and acoustic voice qualities between high- and low-fatigue groups. *Journal of Voice*, 25(5), 544-552. doi:10.1016/j.jvoice.2010.07.006.
20. Whitmore, J., & Fisher, S. (1996). Speech during sustained operations. *Speech Communication*, 20(1-2), 55-70.
21. Krajewski, J., Batliner, A., & Golz, M. (2009). Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach. *Behavior Research Methods*, 41(3), 795-804. doi:10.3758/BRM.41.3.795.
22. Muñoz-de-Escalona, E., Cañas, J. J., & Noriega, P. (2020). Inconsistencies between mental fatigue measures under compensatory control theories. *Psicológica Journal*, 41, 103-126. <https://doi.org/10.2478/psicolj-2020-0006>
23. Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock Jr, J. R. The multi-attribute task battery ii (matb-ii) software for human performance and workload research: A user's guide (2011)
24. Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glott International*, 5(9/10), 341-347.
25. Shen, K. Q., Li, X. P., Ong, C. J., Shao, S. Y., & Wilder-Smith, E. P. (2008). EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clinical neurophysiology*, 119(7), 1524-1533.
26. Borghini, G., Vecchiato, G., Toppi, J., Astolfi, L., Maglione, A., Isabella, R., ... & Polidori, L. (2012, August). Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 6442-6445). IEEE.
27. Hong, L. Y. J. Y., & Zheng, Y. X. L. W. (2010). Study of Mental Fatigue Based on Heart Rate Variability [J]. *Chinese Journal of Biomedical Engineering*, 1.
28. Melo, H. M., Nascimento, L. M., & Takase, E. (2017). Mental fatigue and heart rate variability (HRV): The time-on-task effect. *Psychology & Neuroscience*, 10(4), 428.
29. Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: a possible measure of fatigue. *Human factors*, 36(2), 285-297.

**Acknowledgement.** This research has been supported by contract between CRIDA (<https://cida.es/webcida/>) and the University of Granada to the first two authors and by the master thesis research grant from CRIDA to the third author.

## **Chapter VII**

---

*General Discussion/Discusión General*



## 1. Introducción

La presente tesis doctoral surge de la imperante necesidad de dar respuesta a uno de los tradicionales problemas de la Ergonomía y Ciencias del trabajo en lo referente a la evaluación de la carga y la fatiga mental en contexto de laboratorio, pero, sobre todo, en contextos aplicados: **examinar las posibles causas y circunstancias en las que acontecen divergencias entre las diferentes metodologías de evaluación de carga y fatiga mental**. Así pues, el desarrollo del proyecto de tesis doctoral que aquí nos ocupa se articula principalmente en torno a un conjunto de estudios empíricos diseñados explícitamente para dar respuesta a cada una de las hipótesis barajadas sobre dichas posibles causas y circunstancias del surgimiento de divergencia entre medidas, las cuales han sido establecidas en función de una exhaustiva revisión inicial de literatura científica sobre el estado del arte del problema que nos atañe.

De otro lado, a partir del estudio de las diferentes metodologías que existen en la actualidad para medir carga y fatiga mental, junto con todas sus ventajas e inconvenientes, nos parecía interesante cerrar el presente proyecto de tesis abordando la posibilidad de **investigar alguna metodología que fuera capaz de superar dos de las limitaciones más importantes que existen respecto a las medidas psicofisiológicas, estas son, la elevada intrusividad y los elevados costes de los equipos de registro** que caracterizan a este tipo de metodologías.

A continuación, se procederá a sintetizar los resultados obtenidos en los diferentes estudios empíricos recogidos en la presente tesis, al tiempo que se discutirán las potenciales implicaciones derivadas de dichos hallazgos y las limitaciones presentes en cada uno de ellos. En última instancia, se expondrán las conclusiones generales y se propondrán futuras investigaciones en esta línea.

Podemos dividir los hallazgos obtenidos en la presente tesis en tres bloques: 1) divergencias entre medidas de carga mental, 2) divergencias entre medidas de fatiga mental y 3) exploración de nuevas medidas.



## 2. Divergencias entre medidas de carga mental

### 2.1 La importancia del tiempo

La línea de investigación de la presente tesis dio comienzo con el estudio expuesto en el capítulo 2, en el cual se exploró la posible **relación existente entre el surgimiento de disociaciones e insensibilidades y la existencia de diferencias entre medidas respecto a sus escalas temporales** (Hancock, 2017). De forma particular, con dicho estudio se buscó, en primer lugar, averiguar si efectivamente existen diferencias en cuanto a la latencia de respuesta de las diferentes metodologías de evaluación de carga mental y, en segundo lugar, se buscó determinar si ciertas divergencias entre medidas podrían ser el resultado de dichas diferencias en cuanto a la latencia de respuesta. Los resultados de dicho estudio nos desvelaron que, durante los picos de alta demanda cognitiva, la medida subjetiva de carga mental mostró una menor latencia de respuesta (reaccionó antes) que la medida psicofisiológica “diámetro pupilar”. En otras palabras, pese a existir de forma general una alta correlación entre medidas, **el hallazgo principal de dicho estudio fue que las divergencias pueden aparecer principalmente cuando acontecen picos de elevada carga mental, momento en el cual se aprecian diferencias en cuanto a la latencia de respuesta de las medidas**. Esto sugiere que ciertas divergencias entre medidas pueden ser debidas a la existencia de diferencias en cuanto a la latencia de respuesta de las medidas consideradas, que surgen en momentos determinados (picos de alta demanda cognitiva).

Una de las **implicaciones** principales de este estudio es que **la latencia de respuesta de las diferentes medidas puede variar en función del grado de demanda cognitiva de la tarea**, de este modo, las medidas de carga mental podrían estar convergiendo entre sí durante periodos de baja y media demanda cognitiva, pero podrían, sin embargo, no converger durante periodos de alta demanda cognitiva. Traducido esto al contexto aplicado, podría ocurrir que un trabajador que se encuentra experimentando niveles de demanda cognitiva bajos o moderados podría reflejar de forma equitativa dichos niveles de carga mental en las diferentes medidas, pero, sin embargo, si la situación se complica y, por ende, se produce un incremento en los niveles de demanda cognitiva, sería buena

idea confiar más en ciertos índices que poseen una menor latencia de respuesta que otros durante dichas situaciones concretas. Dicho de otro modo, podríamos utilizar la medida psicofisiológica o subjetiva indistintamente durante periodos de baja o moderada demanda cognitiva ya que reflejan los cambios de forma equivalente y, sin embargo, confiar más en la medida subjetiva que en la medida psicofisiológica “diámetro pupilar” durante picos de alta demanda cognitiva.

En lo referente a las **limitaciones** del primer estudio, la principal limitación es el **número de medidas consideradas en el mismo**: únicamente abordamos la relación entre 3 medidas, una de cada categoría, y sería conveniente estudiar la escala temporal del mayor número de medidas posible, al objeto de profundizar en el comportamiento de cada una de las medidas en cuanto a su latencia de respuesta, así como en el grado en que esto afecta a la convergencia entre las mismas. Por ejemplo, puede que la latencia de respuesta de la pupila sea superior a la de la respuesta psicogalvánica o a la respuesta de los potenciales evocados, de modo que se produzca mayores divergencias entre la medida subjetiva y la primera que entre la medida subjetiva y las últimas. Una segunda limitación tiene que ver con la **granularidad de los datos obtenidos en el estudio**. Más concretamente, los intervalos fueron definidos cada 5 minutos y quizás sería conveniente considerar intervalos más cortos en futuras investigaciones para mejorar el nivel de granularidad, sin embargo, la desventaja principal de aumentar la granularidad sería que el grado de intrusividad de algunas medidas, como la medida subjetiva, aumentaría. En último lugar, este estudio fue realizado en condiciones de laboratorio y con población universitaria y consideramos que **debería ser replicado en contexto real y con una población más heterogénea** para aumentar la validez externa de nuestros hallazgos.

## 2.2 El efecto de las transiciones en carga

Los hallazgos del primer estudio nos parecieron congruentes con la hipótesis del segundo estudio de la tesis, ya que las divergencias entre medidas surgieron precisamente durante transiciones de media a alta carga (picos de carga). En el segundo estudio del proyecto de tesis, correspondiente al capítulo 3, quisimos explorar el **efecto de las transiciones en los niveles de demanda cognitiva sobre la ocurrencia de divergencias entre medidas de carga mental**. Este estudio parte de la base teórica de que resulta

posible que exista una sensibilidad diferencial entre medidas de carga mental a la tasa de cambio en los niveles de demanda cognitiva de la tarea durante las transiciones (Hancock, 2017). Dicho de otro modo, algunas medidas de carga mental podrían mostrarse más sensibles a los cambios en carga que otras, lo cual se pondría de manifiesto principalmente durante transiciones con tasa de cambio abrupta en los niveles de demanda cognitiva de la tarea frente a aquellas caracterizadas por una tasa de cambio lineal. La hipótesis de partida fue que las diferencias entre medidas en cuanto a la sensibilidad a la tasa de cambio podrían dar lugar al surgimiento de disociaciones y/o insensibilidades, lo cual afectaría negativamente a la convergencia entre medidas; operacionalmente hablando, según nuestra hipótesis de partida debería producirse una mayor divergencia entre medidas durante transiciones con tasa de cambio abrupta respecto a aquellas con tasa de cambio lineal.

Nuestros hallazgos en dicho estudio empírico nos desvelaron la **existencia de disociaciones durante las condiciones experimentales de tasa de cambio variable (principalmente en la condición de incremento en carga) en contraste con las condiciones de tasa de cambio lineal, esto confirmaría parcialmente nuestra hipótesis de partida**. Más concretamente, la medida subjetiva de carga mental se mostró menos sensible que el resto de medidas a la tasa de cambio, dado que se produjo una variación constante y lineal de dicha medida en todas las condiciones experimentales. Sin embargo, la medida de rendimiento y la psicofisiológica (diámetro pupilar), desvelaron ser más sensibles al cambio abrupto acontecido en los niveles de demanda cognitiva que a sus niveles absolutos, ya que durante la condición de tasa de cambio variable y ascendente ambas variables mostraron los niveles más elevados en su magnitud justo al producirse el incremento abrupto en carga, siendo dichos valores incluso superiores a los obtenidos en el momento en que se produce la mayor demanda cognitiva de la tarea (hacia el final del ejercicio). En otras palabras, nuestros resultados sugieren que las disociaciones en la medida de rendimiento y en la medida psicofisiológica “diámetro pupilar” podrían aparecer cuando acontece un cambio abrupto en los niveles de demanda cognitiva de la tarea (particularmente durante incrementos), mientras que la medida subjetiva podría verse afectada no tanto por la tasa de cambio, sino por los niveles absolutos de demanda cognitiva de la tarea. De este modo, **dichos resultados confirman parcialmente la hipótesis de partida, ya que la mayor divergencia entre medidas acontece en situaciones de tasa de cambio variable ascendente**, pero no durante condiciones de tasa

de cambio variable descendente (ni durante condiciones de tasa de cambio lineal, ascendente o descendente). Respecto al comportamiento de la variable psicofisiológica, cabe señalar que debemos considerar que el diámetro pupilar lo que refleja es nivel de activación (entre otros factores que fueron controlados), de modo que cuando se produce un incremento abrupto en los niveles de demanda cognitiva, el organismo “sobre-reacciona” (reflejado con una dilatación pupilar abrupta) para preparar al individuo a enfrentar las circunstancias desafiantes del ambiente pero, sin embargo y en la línea de la teoría de los recursos, dado que los recursos mentales son limitados y se pueden agotar, si dicho incremento abrupto es seguido por un ligero incremento en los niveles de demanda cognitiva, nuestro organismo detecta que no hay necesidad de continuar activándose y por tanto se estabiliza o incluso comienza a desactivarse al objeto de ahorrar recursos disponibles que podrían hacer más falta en futuras situaciones.

Atendiendo a las **implicaciones** de dichos hallazgos, la principal implicación consiste en que, a la hora de **seleccionar las medidas** a utilizar para evaluar carga mental, deberíamos **otorgar mayor relevancia a la información proporcionada por una u otra medida en función de las condiciones de tasa de cambio de los niveles de demanda cognitiva de una tarea en particular**. Por ejemplo, si las transiciones en carga son lineales, entonces podríamos confiar en cualquiera de las diferentes medidas de carga mental (subjética, de rendimiento o psicofisiológica), pero si la naturaleza de la tarea implica transiciones abruptas en carga, entonces quizás sería conveniente diferenciar entre medidas, considerando la información particular que cada una de las medidas ofrece. Por ejemplo, la medida psicofisiológica “diámetro pupilar” puede resultar muy útil para indicarnos cuándo se ha producido una transición abrupta en carga, ya que en esta circunstancia tiende a sobre-reaccionar, sin embargo, dado que lo que refleja dicha variable es el nivel de activación del individuo y esto está mediado por numerosos factores (por ejemplo la necesidad del organismo de ahorrar recursos), podría ser que poco después de haberse producido un incremento abrupto en los niveles de demanda cognitiva de la tarea, la activación del individuo se mantenga estable o incluso se reduzca para ahorrar recursos, incluso habiéndose producido de nuevo un ligero aumento en los niveles de demanda cognitiva de la tarea. En esta circunstancia, sería entonces sensato confiar en la medida subjética de carga, dado que dicha disminución en la activación del individuo no sería un reflejo de que se está produciendo una disminución de carga mental (que permanecería elevada, y así se demostraría al preguntar al individuo por su carga

mental percibida), sino que sería el resultado de la acción del mecanismo compensatorio del organismo para ahorrar recursos y tenerlos disponibles en futuras situaciones.

En lo tocante a las **limitaciones** de dicho estudio, debemos destacar las siguientes: en primer lugar, **únicamente fueron consideradas las transiciones que acontecen en una única dirección, ascendente o descendente**, esto puede ser visto como una limitación del presente estudio ya que sería interesante considerar otro tipo de transiciones en las que cambie la dirección del nivel de demanda cognitiva (ej. baja-alta-baja; media-baja-alta, etc.) y también sería conveniente considerar el momento en que la transición abrupta en carga toma lugar (ej. al comienzo, en mitad de la sesión al final de la sesión). Una segunda limitación, que comparte con el primer estudio, tiene que ver con las medidas de carga consideradas. **Sería interesante registrar un abanico más amplio de medidas**, sobretodo psicofisiológicas, para poder analizar el comportamiento particular de cada una de dichas medidas. En tercer lugar, otra limitación compartida con el primer estudio tiene que ver con la población utilizada, una vez más estudiantes universitarios y con las condiciones del estudio, situación simulada, lo cual limita de nuevo la validez externa del mismo; por este motivo, **sería interesante validar dichos hallazgos en situaciones reales y con una población más acorde al contexto real de estudio**.

En suma, los hallazgos del segundo estudio nos desvelaron, principalmente, que **se produce una mayor divergencia entre medidas bajo circunstancias de tasa de cambio variable y ascendente en los niveles de demanda cognitiva de la tarea**. Sin embargo, dicho estudio no consideró un factor que puede tener mucho peso en el surgimiento de las divergencias entre medidas: el nivel preexistente de carga mental cuando acontecen las transiciones abruptas en carga.

### 2.3 El efecto del nivel preexistente de carga mental

En el anterior estudio, la transición abrupta (ascendente o descendente) tomaba lugar a niveles moderados de carga mental, sin embargo, dicho estudio no respondía a la incógnita de cuál sería el comportamiento de las diferentes medidas si la transición abrupta hubiera tomado lugar en situación de baja o alta carga mental. En esta tesitura, el tercer estudio del proyecto de tesis, recogido en el capítulo 4, busca arrojar luz sobre dicha

cuestión y **trata particularmente de responder la incógnita de si algunas medidas podrían ser más sensibles a reaccionar a los cambios abruptos en los niveles de demanda cognitiva en función del nivel preexistente de carga mental.** De este modo, de existir una sensibilidad diferencial entre medidas a los niveles preexistentes de carga mental cuando se produce un incremento abrupto en carga, podríamos encontrar un mayor o menor grado de divergencia entre medidas en función de dicho nivel de carga preexistente. La hipótesis de partida de dicho estudio fue que los efectos de la transición ascendente abrupta sobre las medidas de carga mental serían diferentes en función del nivel preexistente de carga mental. Particularmente, cuando el nivel de carga mental preexistente es reducido, el organismo dispone de numerosos recursos disponibles para ser movilizados, de manera que, si se produce un incremento abrupto en carga en esta circunstancia, la cantidad de recursos mentales (nivel de activación) movilizados para afrontar el incremento en dificultad de la tarea será mayor que si se parte de una situación con un mayor nivel de carga mental preexistente, lo cual resultará en un comportamiento diferente de la medida psicofisiológica respecto a las otras dos medidas de carga mental. Operacionalmente hablando, la hipótesis de partida establece que se producirá una mayor divergencia entre las medidas de carga mental cuando el cambio abrupto y ascendente en los niveles de demanda cognitiva acontece en situaciones de baja carga preexistente respecto a aquellas situaciones en las que los niveles de carga preexistentes son más elevados; esto será debido a que la medida psicofisiológica debería reaccionar en mayor medida que las medidas subjetivas y de rendimiento, ya que esta primera medida reflejará la sobreactivación que tendrá el organismo (que se encuentra repleto de recursos) para afrontar satisfactoriamente el incremento abrupto en los niveles de demanda cognitiva de la tarea.

Los resultados de dicho estudio nos desvelaron que **todas las medidas consideradas fueron sensibles al incremento en la demanda cognitiva de la tarea, sin embargo, observamos ciertas diferencias entre medidas en el modo de reflejar los cambios abruptos:** en primer lugar, la medida de rendimiento reflejó adecuadamente el incremento a lo largo de los intervalos pero también se mostró sensible a los incrementos abruptos en los niveles de demanda cognitiva de la tarea en todas las condiciones, es decir, independientemente del nivel preexistente de carga mental; en segundo lugar, la medida subjetiva de carga también reflejó adecuadamente el incremento en los niveles de demanda cognitiva a lo largo de los intervalos, sin embargo, no se mostró sensible en

absoluto a los incrementos abruptos en carga, observándose un incremento lineal de dicha medida en todas las condiciones experimentales; en última instancia, en lo referente a la medida psicofisiológica “diámetro pupilar”, de nuevo nuestros resultados desvelaron una adecuada sensibilidad al incremento en los niveles de demanda cognitiva de la tarea a lo largo de los intervalos pero, el hallazgo más destacado de dicho estudio fue que, dicha medida también se mostró sensible al incremento abrupto en los niveles de demanda cognitiva de la tarea pero únicamente durante la condición experimental de baja carga mental preexistente. **Estos resultados, por tanto, confirman nuestra hipótesis de que se producirán más divergencias entre medidas cuando se produce un incremento abrupto en los niveles de demanda cognitiva de la tarea particularmente durante condiciones de baja carga mental preexistente.** Esto es debido principalmente al comportamiento de la pupila en condiciones de baja carga mental preexistente: dada la existencia de amplias reservas de recursos disponibles, el organismo será capaz de movilizar una gran cantidad de recursos para hacer frente al incremento abrupto en la dificultad de la tarea, esta sobreactivación se verá reflejada en un incremento abrupto del diámetro pupilar, el cual no coincidirá con las medidas de rendimiento ni con la medida subjetiva. Más concretamente, la medida de rendimiento también encontrará un pico de deterioro, pero menos acentuado, al tiempo que la percepción subjetiva de carga no reflejará ningún tipo de cambio abrupto, sino un incremento lineal. Estos resultados coinciden con la teoría de los recursos, ya que el ser humano puede movilizar muchos más recursos en condiciones de baja carga mental que en condiciones de media o alta carga, en las que los recursos son más limitados y corren el riesgo de agotarse por lo que el organismo debe actuar de forma más conservadora.

La **implicación** más relevante del estudio coincide con la del estudio anterior y consiste en que **debemos ser cautos a la hora de seleccionar medidas para evaluar carga mental, siendo necesario basar dicha decisión en nuestros objetivos particulares**, habida cuenta de que cada medida de carga mental refleja de forma particular al constructo. A modo de ejemplo, si el objetivo del evaluador es detectar picos de carga mental en cualquier circunstancia, sin considerar el nivel preexistente de carga mental, entonces la medida de rendimiento será más adecuada, sin embargo, si se precisa detectar con rapidez y precisión transiciones ascendentes abruptas que acontecen en situaciones de baja carga preexistente durante periodos de tiempo prolongados, como por ejemplo tareas de monitorización, quizás sería más acertado utilizar una medida

psicofisiológica que refleje activación, dado que parecen mostrarse más sensibles a los cambios abruptos en los niveles de demanda cognitiva de la tarea bajo dicha circunstancia. Si, por el contrario, lo que nos interesa es obtener una visión general del nivel de carga mental al que se encuentra sometido un individuo entonces nos interesará utilizar la medida subjetiva de carga mental, medida que nos interesa también utilizar, sobretodo, en situaciones de alta carga preexistente, dado que no podemos confiar en otras medidas como por ejemplo la psicofisiológica por los motivos expuestos en el apartado anterior y, debemos por tanto confiar en la medida subjetiva para evitar situaciones potencialmente peligrosas de sobrecarga mental, así como para evitar el surgimiento de la fatiga mental.

En lo que respecta a las **limitaciones** de este estudio, una vez más nos encontramos con el **problema de la muestra utilizada**, constituida principalmente por jóvenes universitarios. De esta manera, sería muy **interesante realizar estudios de validación de los mismos en contextos reales y con diferentes muestras más heterogéneas** (por ejemplo, en alguna empresa concreta, con trabajadores reales). Otra limitación que vuelve a repetirse es el uso de un **conjunto restringido de medidas de carga**. Sería **interesante utilizar otro tipo de medidas**, concretamente múltiples medidas psicofisiológicas, para determinar el comportamiento diferencial de cada una de ellas y para analizar las divergencias que se producen dentro de las propias medidas psicofisiológicas.

### 3. Divergencias entre medidas de fatiga mental

La presente tesis doctoral continúa su línea de investigación con el cuarto estudio, expuesto en el capítulo 5, en el cuál se investigan las causas y circunstancias del surgimiento de divergencias entre medidas, pero esta vez, de fatiga mental. Recordemos que la carga y la fatiga mental son conceptos estrechamente vinculados entre sí, ya que el segundo depende fundamentalmente de las condiciones del primero, del paso del tiempo ejecutando la tarea y de las diferencias individuales en cuanto a los recursos disponibles de cada individuo. En el cuarto estudio de la presente tesis se busca **arrojar luz sobre las divergencias que se producen entre medidas de fatiga mental bajo el efecto del mecanismo de control compensatorio propuesto por Hockey en 1997**. Recordemos



que, según dicho modelo, el organismo es capaz de movilizar recursos extra cuando el nivel de dificultad de la tarea así lo requiere, al objeto de mantener un rendimiento adecuado bajo circunstancias más demandantes, lo cual desembocará en un mayor nivel de activación fisiológica y una mayor percepción subjetiva de carga mental, la cual actúa como señal de alarma frente al agotamiento de los recursos mentales. Nosotros pensamos que la acción de dicho mecanismo compensatorio podría ser responsable del surgimiento de ciertas divergencias entre medidas de fatiga mental. Particularmente, nuestra hipótesis de partida fue que **la medida subjetiva de fatiga no correlacionaría con la medida psicofisiológica ni con la medida de rendimiento, ya que esta primera actúa realmente como una señal de alarma frente al agotamiento de recursos mentales, pese a existir una cantidad de recursos suficiente para afrontar la tarea de forma satisfactoria**, tal y como quedaría reflejado mediante la medida psicofisiológica “diámetro pupilar” con un nivel elevado y estable de activación (en función de la dificultad de la tarea), y de rendimiento, con un rendimiento adecuado pese al incremento en dificultad.

Los **resultados arrojados en el estudio desvelaron un rendimiento óptimo y estable de los participantes en ambas condiciones experimentales (baja y alta dificultad), sin embargo, los niveles de activación se mostraron diferentes: el diámetro pupilar en la condición de alta complejidad se mostró más elevado a lo largo de los intervalos (mayor nivel de activación) que en la condición de baja complejidad. Asimismo, los niveles de fatiga subjetiva incrementaron linealmente a lo largo de los intervalos en ambos grupos, aunque con valores mayores en el caso de la condición de alta complejidad.** Por tanto, atendiendo a dichos resultados, **podemos confirmar que encontramos divergencias entre las medidas de fatiga mental**, ya que cabría haber esperado un deterioro en el rendimiento y una disminución de la activación asociados a un incremento en los niveles de fatiga mental percibida, cosa que no aconteció. Estos resultados **confirman, por tanto, nuestra hipótesis de partida** y pueden ser explicados bajo el efecto del mecanismo de control compensatorio: en la condición de elevada dificultad, la acción del mecanismo compensatorio suscitó una mayor activación (mayor movilización de recursos) para hacer frente a la mayor dificultad de la tarea en contraste con la condición de baja dificultad, donde no se requerían niveles tan altos de activación para afrontar dicha tarea de naturaleza más sencilla; el rendimiento se mostró estable y adecuado en ambas condiciones y la fatiga subjetiva incrementó con

el paso del tiempo, pero de forma más notable en la condición de alta complejidad. Cabe resaltar que el incremento en los niveles de fatiga a lo largo de los intervalos fue mayor en la condición de alta complejidad dado el mayor esfuerzo del organismo (mayor asignación de recursos) para afrontar de forma satisfactoria la tarea de mayor complejidad. De este modo, podríamos decir que tal y como Hockey postula, la fatiga subjetiva actúa como una señal de alarma del organismo para impedir el posible agotamiento de los recursos mentales, incluso cuando aún existen recursos suficientes para desempeñar la tarea de forma satisfactoria.

La **implicación** más relevante de dicho estudio es que **debemos ser cautos a la hora de considerar el uso de las medidas psicofisiológicas como único indicador de fatiga mental**, tal y como acontece por ejemplo en el caso de los sistemas de reconocimiento de fatiga mental en los vehículos de última generación, los cuales únicamente tienen en cuenta parámetros oculares para evaluar la fatiga del conductor. En este sentido y, de acuerdo a los hallazgos del presente estudio, nuestro comportamiento ocular podría indicar la existencia de niveles elevados de activación, lo cual no sería indicativo a priori de la existencia de fatiga mental y, sin embargo, podría ser el resultado de la acción del mecanismo compensatorio que provee con recursos extra para afrontar la situación demandante cuando, en realidad, si preguntamos al conductor, este podría llevar tiempo sintiéndose fatigado, incrementándose así la probabilidad de ocurrencia de un accidente de tráfico.

En última instancia, en lo tocante a las **limitaciones** del estudio, debemos destacar las siguientes: al igual que todos los estudios anteriores, podemos considerar como una limitación del estudio **el hecho de disponer de una única medida psicofisiológica**, ya que limita el estudio del comportamiento de las medidas psicofisiológicas como categoría. En este sentido, **sería interesante estudiar el comportamiento de otras variables psicofisiológicas** tales como la actividad eléctrica cerebral, la respuesta psicogalvánica o los parámetros cardiorrespiratorios ya que sabemos que las divergencias se han encontrado no solo entre diferentes categorías de medidas sino también dentro de cada categoría con diferentes índices, por lo que se necesita una mayor investigación para desentrañar los misterios de las divergencias entre medidas de fatiga mental. En cualquier caso, pensamos que cualquier variable psicofisiológica que refleje activación del organismo debería reflejar adecuadamente los cambios producidos por el mecanismo de

control compensatorio. Una segunda limitación del estudio se refiere, una vez más, a la **muestra utilizada (estudiantes universitarios) y a las condiciones contextuales (contexto de laboratorio)**, pensamos que este estudio debería ser replicado en contextos reales de trabajo para incrementar la validez externa del mismo. Una tercera limitación tiene que ver con el diseño de la investigación, cuyos **participantes fueron divididos en ambas condiciones experimentales, pero ningún participante pasó por ambas condiciones**. En este sentido, algunas diferencias entre grupos podrían ser atribuidas en parte a las diferencias grupales, sin embargo, nosotros pensamos que esta limitación fue corregida estadísticamente gracias al procedimiento de asignación aleatoria de los participantes a ambos grupos experimentales.

#### 4. Exploración de nuevas medidas

La línea de investigación de esta tesis concluye con el quinto estudio, publicado en el capítulo 6, el cual persigue por objetivo la exploración de nuevas medidas psicofisiológicas de fatiga mental que superen las limitaciones de las medidas actuales. Tal y como queda indicado en el capítulo 2, las medidas psicofisiológicas tienen numerosas ventajas que no comparten con las medidas de rendimiento ni con las medidas subjetivas, pero, sin embargo, también tienen una serie de inconvenientes particulares de entre los cuales destacan principalmente dos: 1) el coste de los equipos de registro, junto con la dificultad técnica que supone su uso y 2) la elevada intrusividad que el empleo de medidas psicofisiológicas supone para el evaluado (Young & Stanton, 2001). El objetivo del último estudio del presente proyecto de tesis es, precisamente, **la exploración de una medida de fatiga mental que según los últimos estudios parece ser capaz de reflejar adecuadamente los cambios en la magnitud del constructo y que supera ambas limitaciones** expuestas en líneas anteriores. La medida explorada consiste en un parámetro acústico de la voz denominado “frecuencia fundamental” (F0) y que guarda una estrecha relación con el “tono de voz”. Las ventajas indiscutibles de dicha medida son la sencillez en cuanto a su registro, el bajo coste que supone su registro (ej. una simple grabadora de voz) y la ausencia de intrusividad de la técnica en el evaluado, dado que el registro de la voz no impide ni afecta al funcionamiento normal del individuo. Nuestra **hipótesis de partida fue que la frecuencia fundamental de la voz**, tal y como parecen

indicar algunos estudios recientes (Hagmüller Kubin & Rank, 2005; Diepeveen, 2020), **aumentaría a lo largo de los intervalos, a medida que aumenta la fatiga subjetiva** de los participantes al realizar la tarea.

Nuestros **resultados desvelaron que se produce un aumento de la frecuencia fundamental de la voz a lo largo de los intervalos** (a medida que transcurre el tiempo ejecutando la tarea), **así como un incremento de la fatiga subjetiva percibida y una disminución tanto del rendimiento en la tarea como del diámetro pupilar** (disminución del nivel de activación). En otras palabras, **nuestros hallazgos confirmaron nuestra hipótesis de partida** de que se produciría un aumento de F0 con el paso del tiempo, esto es, a medida que incrementaba el nivel de fatiga mental. De este modo, estos datos **respaldan la idea de considerar la frecuencia fundamental de la voz como una medida de fatiga mental apropiada y válida**, que puede utilizarse como alternativa a otros índices psicofisiológicos bien establecidos, superando las limitaciones más destacadas de este grupo de medidas.

Las **implicaciones** más relevantes de este estudio tienen que ver precisamente con la **superación de dichas limitaciones**. En primera instancia, el surgimiento de fatiga mental en ciertos ámbitos de trabajo en los que los trabajadores requieren de la voz para ejercer sus funciones (ej. el ámbito “ATM” (Air Traffic Management, por sus siglas en inglés)), podría ser predicho directamente a partir de las grabaciones de la voz de los trabajadores, de forma automática, económica y fiable. No se requieren equipos costosos de registro e, incluso, en muchas ocasiones ya se encuentra automatizado el registro de la voz de los trabajadores, tal y como ocurre en el ámbito ATM, en el cual se registran continuamente las comunicaciones de los ATCo (Air Traffic Controller, por sus siglas en inglés) con los pilotos y con otros ATCo. De otro lado, en muchas ocasiones existe un histórico de registro de los datos de voz, por lo que podrían realizarse análisis a posteriori del nivel de fatiga mental experimentado por los trabajadores en un momento en concreto para, por ejemplo, realizar el análisis de accidentes laborales o simplemente para realizar un estudio de los niveles de fatiga mental experimentado por los trabajadores en momentos determinados, con fines preventivos u organizacionales.

En cualquier caso, los resultados del presente estudio deben verse a la luz de una serie de **limitaciones experimentales**. En primer lugar, tal y como ocurre con todos los

estudios expuestos en la presente tesis, la **muestra experimental quedó constituida por un grupo muy homogéneo de estudiantes jóvenes universitarios**, por lo que dichos hallazgos **deberían ser replicados en contextos reales y bajo condiciones reales** para ser validados. Asimismo, en lo referente al análisis estadístico realizado, no debemos dejar de lado el hecho de que **el número de niveles de nuestra variable independiente fue mayor que el número de participantes**, lo cual podría haber afectado al poder estadístico del ANOVA realizado, aunque también es cierto que el tamaño del efecto fue muy elevado. En último lugar, otra limitación fue el **periodo de tiempo considerado** en nuestra condición experimental: se estableció una única condición de 60 minutos de duración, por lo que habría sido interesante considerar otras condiciones experimentales con diferentes periodos temporales, mayores y menores a 60 minutos, así como con diferentes niveles de demanda cognitiva, al objeto de poder estudiar el surgimiento y evolución de la fatiga mental, así como la idoneidad del uso de la frecuencia fundamental de la voz para la detección y seguimiento de la misma.

## 5. Futuras direcciones de investigación

Considerando los trabajos recogidos en la presente tesis doctoral, podrían establecerse algunas líneas de investigación destinadas a ampliar, clarificar y complementar algunos de los hallazgos obtenidos.

- 1) En primer lugar y atendiendo a todos los estudios de la presente tesis doctoral, pensamos que los hallazgos son muy enriquecedores y relevantes para el campo de la Ergonomía y las Ciencias del trabajo, pero también pensamos que **resulta fundamental validar todos y cada uno de los hallazgos de la presente tesis en contextos aplicados**. A este respecto, los investigadores sabemos que en el laboratorio, las condiciones controladas e ideales permiten una manipulación experimental óptima, con un control impecable de las variables extrañas que pudieran contaminar el efecto de la manipulación de nuestras variables independientes sobre los resultados; sin embargo, también sabemos que la investigación básica cobra sentido especialmente cuando dichos hallazgos pueden ser validados en contextos reales y, además, esto tiene especialmente relevancia en campos tan aplicados como son la Ergonomía y las Ciencias del Trabajo. Por tanto, pensamos que resulta

fundamental replicar todos y cada uno de los estudios aquí presentados en contextos aplicados, como, por ejemplo, con la plantilla de trabajo de alguna empresa determinada. Únicamente de esta manera, podremos tener la certeza de que los hallazgos aquí expuestos pueden ser aplicados en contextos reales y, por tanto, pueden tener una utilidad y un beneficio real en nuestra sociedad.

2) En segundo lugar y en lo que respecta al estudio de las causas y circunstancias de las divergencias entre medidas de carga y de fatiga mental, debe quedar muy claro el hecho de que el surgimiento de divergencias entre medidas es un problema real que debe ser ampliamente abordado con investigación futura, dado que los factores que afectan a dicho fenómeno son múltiples y difíciles de interpretar. A este respecto, pensamos que esta segunda línea de investigación debe tener **3 vertientes**:

- i) En primer lugar, la primera vertiente debe establecerse bajo el **objetivo de replicar, corroborar y amplificar los hallazgos obtenidos con los trabajos de la presente tesis**, esto es, volver a probar las hipótesis de partida tanto en condiciones variables como en múltiples contextos, ya que sólo de esta manera se podrá validar y comenzar a afianzar el conocimiento que se está empezando a desvelar y que ha quedado reflejado en el presente trabajo.
- ii) En segunda instancia, cabe señalar que los motivos del surgimiento de divergencias entre medidas de carga y de fatiga mental son múltiples y no todos ellos han sido abordados en la presente tesis doctoral, sino solamente aquellos que considerábamos más determinantes. A este respecto, **existen numerosas posibles causas relacionadas con el surgimiento de divergencias entre medidas de carga y fatiga mental que no han sido consideradas en la presente tesis doctoral y que requieren ser investigadas** por la ciencia al objeto de continuar arrojando luz sobre dicho fenómeno. A modo de ejemplo, Hancock (2017) explica en su revisión que otro de los factores que afecta a la divergencia entre medidas tiene que ver con las influencias del efecto de la retrospección y de la prospección. El primer efecto hace referencia a la posible influencia que ejercen los contenidos de la memoria sobre la percepción de carga mental. Esto es, nuestra percepción de carga mental subjetiva depende de las vivencias pasadas: un individuo que

experimentó una elevada carga mental en el pasado mientras realizaba una actividad determinada, tendrá una predisposición a percibir elevada carga mental mientras realiza la misma actividad en el presente, a pesar de que en realidad los niveles de demanda cognitiva de dicha actividad podrían ser más reducidos. El segundo efecto tiene que ver con la influencia que ejerce la prospección del futuro sobre la percepción subjetiva de carga mental, esto es, la anticipación mental de una situación que aún no ha acontecido puede afectar al modo de percibir la carga mental subjetiva en el presente: un individuo que anticipa una situación de alta demanda cognitiva en un futuro, tendrá unos niveles presentes de carga mental percibida mayores que en el caso en que no anticipara dicha situación futura de elevada demanda cognitiva. De este modo, ambos efectos, el de retrospección y el de prospección, podrán influir sobre el surgimiento de divergencias entre medidas de carga mental, por lo que un progreso muy relevante en la comprensión de las divergencias entre medidas de carga mental podría ser llegar a adquirir un mayor conocimiento sobre la influencia de ambos efectos temporales sobre la carga mental percibida durante el momento presente.

- iii) En último lugar, la tercera vertiente del estudio de las disociaciones entre medidas de carga y fatiga mental tiene que ver con la existencia de un tipo de disociaciones concreto y son las **disociaciones que se producen entre diferentes medidas, pero, en esta ocasión, entre medidas de la misma categoría**. Esto es, dentro de cada uno de los tres grandes grupos de medidas (las de rendimiento, las psicofisiológicas y las subjetivas), existen diversas medidas que también pueden mostrar divergencias entre sí al evaluar carga o fatiga mental. En otras palabras y, tomando como ejemplo la evaluación de carga mental mediante medidas psicofisiológicas, pudiera ser que un aumento en los niveles de demanda cognitiva se viera reflejado en un incremento de la conductancia eléctrica de la piel, lo cual sería indicativo de un aumento en los niveles de activación dada la mayor carga mental, y, sin embargo, si registramos la dilatación pupilar, podríamos observar que no se ha producido tal incremento esperado o, incluso, que ha disminuido. Este ejemplo constituye un caso plausible de divergencia entre medidas del mismo tipo (psicofisiológicas) y pensamos que nuestra ciencia debe hacer un gran

esfuerzo para desentrañar las causas y circunstancias de surgimiento de este caso particular de divergencias entre medidas. Por ejemplo, una de las causas podría estar relacionada con la latencia de respuesta de las diferentes medidas psicofisiológicas y que, tal y como quedó expuesto en el capítulo 2 de la presente tesis, el hecho de existir diferencias en cuanto a la latencia de respuesta de las distintas medidas podría ser el origen de ciertas divergencias que acontecen entre medidas. Otra posible causa podría tener su origen en el efecto que ejercen ciertas variables extrañas de forma particular a alguna medida en concreto, a modo de ejemplo, sabemos que las condiciones de luz y sequedad pueden afectar a los parámetros oculares diámetro pupilar y parpadeo, de modo que estas medidas podrían verse contaminadas por dichos condicionantes, en tanto que otras medidas no, lo cual podría desembocar en el surgimiento de divergencias entre medidas.

- 3) En último lugar, la tercera línea de investigación que consideramos necesaria para el avance de nuestra ciencia tiene que ver con la **exploración de nuevas medidas psicofisiológicas de carga y de fatiga mental que sean capaces de superar o mejorar de algún modo las limitaciones existentes con las actuales medidas**. Este tema en concreto se ha abordado en el capítulo 6 de la tesis doctoral, donde se explora la posibilidad de utilizar el parámetro acústico de la voz “F0” como medida de fatiga mental que supera las principales limitaciones de las medidas subjetivas. A este respecto, consideramos que la investigación debería continuar abordando 2 vertientes particulares:

- i) En primer lugar, consideramos que debe **seguir explorándose la voz humana y sus múltiples parámetros como medida alternativa válida y fiable tanto de carga como de fatiga mental**. La ventaja fundamental de utilizar la voz ya ha sido explicitada en líneas anteriores y, recordemos tiene que ver con la inexistencia de intrusividad, así como la facilidad, economía de uso y automatización del análisis, entre otras. Existen numerosos parámetros de la voz que pueden ser analizados y que tienen que ver tanto con el contenido, como con la forma, como con la variación de los parámetros acústicos de la voz. A modo de ejemplo, sabemos que los ATCo, al encontrarse en situaciones de elevada carga mental, tienden a recortar en el contenido, ajustándose



únicamente a transmitir el mensaje relevante y dejando de lado la cortesía, sin embargo, cuando el nivel de carga mental es reducido o asumible, los ATCo suelen ser muy amables y corteses a la hora de interactuar con los pilotos. La investigación de todos estos parámetros permitirá ahondar en el conocimiento de la voz y validarla como medida de carga y fatiga mental, capaz de ser utilizada en cualquier ambiente en el que los trabajadores deban de hablar para poder ejercer sus funciones.

- ii) La segunda vertiente que consideramos que debe ser explorada con inmediatez es el **uso de las expresiones emocionales como posible medida alternativa de carga y fatiga mental**. Las emociones son el resultado de un proceso de estimación en el cual los individuos evalúan estímulos externos o representaciones mentales en términos de su relevancia con respecto a sus necesidades actuales y objetivos, incluyendo aspectos de su habilidad para afrontar las consecuencias (Scherer, 2000). Todas las emociones pueden ser clasificadas en cuanto a su nivel de arousal (activación psicofisiológica que generan en el organismo) y su valencia (positiva o negativa, según su valor afectivo) (Barrett, 1998; Anders et al., 2004; Kensinger, 2004). En este sentido, las emociones están fuertemente vinculadas con los conceptos de carga y de fatiga mental: el resultado de dicho proceso de estimación de estímulos externos (elevada demanda cognitiva) en términos de su relevancia respecto a sus necesidades y objetivos (ejecutar óptimamente la tarea) desembocará en el surgimiento de emociones de elevado arousal y valencia negativa (Bradley et al., 2008). Esta hipótesis ya fue explorada por un estudio liderado por nuestro equipo de investigación y los **resultados desvelaron que el diámetro pupilar correlacionaba significativamente con la emoción “disgusted” caracterizada por su arousal negativo y elevado nivel de activación**. Por tanto, nuestros resultados en dicho estudio apoyaron la idea de utilizar las expresiones emocionales, particularmente la emoción “disgusted” como un indicador alternativo de activación y, por ende, de carga mental (Muñoz-de-Escalona & Cañas, 2017). En última instancia, cabe señalar que el uso de expresiones emocionales como alternativa válida y fiable a las medidas de carga y fatiga mental tendría ventajas parecidas a las del uso de la voz humana, habida cuenta de que en la actualidad existen numerosos

softwares de reconocimiento automático y en tiempo real de expresiones emocionales mediante grabaciones de video, por lo que el registro de dicha medida se podría hacer de forma automatizada y en tiempo real mediante una webcam, que resulta económica y nada intrusiva.

## 6. Conclusiones finales

Con la presente tesis doctoral se ha intentado contribuir a partir de nuestros hallazgos a **incrementar el conocimiento existente sobre las divergencias que se producen entre medidas de carga mental y entre medidas de fatiga mental**. Las divergencias entre medidas constituyen un fenómeno más frecuente de lo que, a priori, cabría esperar, tanto en investigación básica como aplicada, así como en contextos reales de evaluación de carga y fatiga mental. Los problemas derivados de la aparición de divergencias entre medidas son múltiples y fundamentalmente afectan a la correcta evaluación de ambos constructos, lo cual incide, en última instancia, tanto en la seguridad en el trabajo, como en la productividad empresarial y en el bienestar de los trabajadores.

En primer lugar, los hallazgos presentados señalan que pueden producirse divergencias entre medidas de carga mental debido a numerosos factores, de entre los cuales se han explorado el papel de la latencia de respuesta de las medidas, el papel de la tasa de cambio en los niveles de demanda cognitiva de la tarea y el papel del nivel de carga preexistente durante las transiciones en carga mental. Nuestros resultados desvelan que **cada uno de estos factores afecta de forma diferencial a las diferentes medidas de carga, lo cual da lugar al surgimiento de divergencias entre medidas de carga mental en circunstancias particulares**. En segundo lugar, hemos explorado el papel del mecanismo de control compensatorio sobre las divergencias que se producen entre medidas de fatiga mental y nuestros resultados ponen de manifiesto que **la acción de dicho mecanismo compensatorio efectivamente podría estar involucrada en el surgimiento de dichas divergencias**. Finalmente, hemos explorado la posibilidad de utilizar el parámetro frecuencia fundamental de la voz como medida alternativa fiable y válida de fatiga mental, con el valor añadido de ser capaz de superar las principales limitaciones comunes a las medidas psicofisiológicas y nuestros resultados parecen indicar que **dicho índice se muestra sensible a las variaciones en la magnitud del constructo evaluado**.

En cualquier caso, a lo largo de los estudios hemos podido comprobar que no podemos confiar únicamente en el uso de una única medida para evaluar carga o fatiga mental, habida cuenta de la existencia de múltiples factores que afectan de forma particular a cada una de las medidas. A este respecto, una de las conclusiones fundamentales a las que podemos llegar hace referencia a la necesidad de **establecer una triangulación en el proceso de evaluación** de dichos constructos. Esto es, la forma más adecuada para realizar una evaluación de carga o de fatiga mental comprende **el uso de al menos una medida por cada uno de los tres grandes tipos de medida que existen**, dado que cada uno de los índices tiene sus propias ventajas e inconvenientes, además de que todos ellos contribuyen de forma efectiva a aportar información única y relevante sobre los constructos evaluados.

En último lugar, y llevando al ámbito aplicado los hallazgos de esta tesis doctoral, **quizás ahora estamos más preparados para comprender qué sucedió exactamente con los acontecimientos de nuestros protagonistas presentados en la introducción, Ginés Navarro y Jorge Torres**, quienes, recordemos sufrieron las consecuencias de la aplicación incorrecta de ciertas metodologías de evaluación de fatiga y carga mental respectivamente.

Comenzando por el caso del accidente sufrido por Ginés Navarro, recordemos que el sistema de reconocimiento automático de fatiga del tráiler de marca Premium no fue capaz de detectar el surgimiento de dicho fenómeno a partir de la monitorización del diámetro pupilar de su conductor: Ginés permanecía con las pupilas dilatadas, a pesar de que se encontraba fatigado subjetivamente, esto le llevó a tener un descuido, salirse de la calzada y volcar el camión. Volviendo a las preguntas que planteábamos al inicio ¿Qué había ocurrido?, ¿Por qué el tráiler no le había avisado del surgimiento de fatiga mental?, ¿Por qué Ginés confiaba más en dicha tecnología que en sus propios instintos?, ¿Se podría haber evitado?, quizás ahora tengamos un marco de conocimiento más amplio para esclarecer dichas preguntas: lo que ocurrió fue una suma de factores, de entre los cuales destacan el hecho de que **Ginés confiara sobremanera en dicha tecnología de detección de fatiga mental instalada en el tráiler**, la cual precisamente no fue capaz de detectar el aumento en los niveles de fatiga mental de Ginés. El tráiler no fue capaz de avisar de dicho surgimiento de fatiga al conductor debido al **uso inadecuado de una**

**metodología de detección de fatiga mental basada en una única medida**, esta es, la monitorización del diámetro pupilar del conductor, medida, además, que hemos podido desvelar que se ve influenciada por la acción del mecanismo compensatorio del organismo, el cual se encarga de reasignar recursos mentales extra cuando la situación lo requiere, lo cual provocó un aumento en el nivel de activación del sistema nervioso central que se vio reflejado en una dilatación pupilar del conductor. **Ginés confiaba en dicho sistema por encima de sus instintos porque la presión temporal de su trabajo le motivaba a confiar más en dicha tecnología que en sus propios instintos, que en esa señal de alarma que tiene el organismo para avisar de que los recursos mentales se están agotando y que se percibe en forma de fatiga mental subjetiva.** Finalmente, se podría haber evitado dicho accidente, principalmente **si Ginés no se viera sometido a dicha presión temporal** en el trabajo que lo llevó a abusar de su propio organismo para cumplir con objetivos pero, también, se podría haber evitado **si el sistema de detección de fatiga del trailer hubiera estado basado en una triangulación de medidas:** una medida de rendimiento, como por ejemplo, la detección de oscilaciones poco comunes en cuanto a la trayectoria del camión; una medida psicofisiológica, como por ejemplo la tasa de parpadeo, medida que no se basa en el nivel de activación del sistema nervioso central y por último una medida de fatiga subjetiva, como por ejemplo un recordatorio cada 2 horas que permita al conductor evaluar su estado de fatiga mental y con sugerencias de tomar un descanso que le permita recuperarse y así evitar el surgimiento de fatiga y, por ende, el posible acontecimiento de accidentes.

Finalmente, en lo tocante al suceso del controlador Jorge Torres, quizás también estemos ahora en una mejor posición para poder arrojar luz sobre las causas de su error. Recordemos que Jorge no fue capaz de detectar el conflicto que se estaba dando entre 2 vuelos bajo circunstancias en las que realmente se encontraba sobrecargado, sin embargo, él lo negaba en todo momento, pese a que su rostro se mostraba sudoroso y sus pupilas llamativamente dilatadas (indicios de sobrecarga). ¿Qué había ocurrido?, ¿por qué Jorge afirmó no estar sobrecargado, sin embargo, sus índices psicofisiológicos indicaban lo contrario?, ¿cómo se podría haber evitado dicha situación de peligro? En este caso, estamos ante un claro ejemplo de una de las principales limitaciones de las medidas subjetivas en general y de carga subjetiva en particular, y es que **las medidas subjetivas son susceptibles de ser falseadas por el evaluado por motivos varios.** Por ejemplo, puede que Jorge estuviera recién contratado en la empresa y no quería mostrar síntomas

de incapacidad o de debilidad ante su compañero o, incluso, puede que la excesiva confianza en sí mismo de Jorge le impidiera percatarse de que realmente se encontraba sobrecargado. Esto es, las medidas subjetivas son altamente influenciadas por numerosos factores intrínsecos (sesgos cognitivos, conflictos motivacionales, baja capacidad de introspección, etc.), por lo que **una evaluación de los niveles de carga o de fatiga mental no debería estar basada únicamente en medidas subjetivas**. Una vez más, este tipo de situaciones podrían evitarse de forma eficaz mediante el **uso de una metodología de evaluación de carga mental basada en la triangulación de al menos una medida de cada tipo**: una medida de rendimiento, por ejemplo el número de quasi-conflictos, o número de pares de aviones que sin haber entrado en conflicto estuvieron cerca de estarlo; una medida psicofisiológica como por ejemplo la tasa de parpadeo, que resulta fácilmente detectable y está vinculada directamente con las estrategias inconscientes que adopta el organismo para favorecer el procesamiento de la información en situaciones de sobrecarga o la frecuencia fundamental de la voz, que ofrece numerosas ventajas ya analizadas en líneas anteriores; y una medida subjetiva, que permita cerrar el círculo, ofreciendo información única sobre la percepción de carga del propio individuo.

## 7. Final conclusions (in English)

This doctoral thesis comprises an attempt to contribute, from our findings, to **increase the existing knowledge about the divergences that occur between mental workload measures and between mental fatigue measures as well**. Divergences between measures are a more frequent phenomenon than, a priori, would be expected, both in basic and applied research, as well as in real contexts of mental workload and mental fatigue assessment. The problems derived from the emergence of divergences between measures are multiple and fundamentally affect the appropriate evaluation of both constructs. This would ultimately affect safety at work, business productivity and the welfare of employees.

Firstly, our findings indicate that mental workload divergences between measures may emerge due to multiple factors, among which we have explored 1) the impact of latency differences between measures, 2) the effect of differential sensitivity to rates of change during task demand transitions and 3) the influence of task-load baseline during task

demand transitions. Our results revealed that **each of these factors affect each mental workload measures in different ways and this would lead to the emergence of divergences between mental workload measures, under particular circumstances.** Secondly, we have explored the impact of the “compensatory control mechanism” on the divergences that arises between mental fatigue measures and our results showed that **the action of such compensatory mechanism could indeed be involved in the emergence of said divergences.** Finally, we have explored the possibility of considering the fundamental frequency of the voice as a reliable and valid alternative measure of mental fatigue, with the benefit of being able to overcome the main limitations common to psychophysiological measures. Our results seem to indicate that **the fundamental frequency of the voice is sensitive to reflect changes in mental fatigue.**

In any case, throughout the different studies that comprise this doctoral thesis, we have learned that we cannot rely on assessing mental workload and/or mental fatigue by using a single measure, since we must consider the existence of multiple factors that affect each different measure in a particular ways. In this regard, one conclusion of crucial significance refers to the need to **establish a triangulation in the evaluation process of these constructs.** In other words, the most appropriate way to carry out an assessment of mental workload and/or mental fatigue involves **the use of at least one measure for each of the three major groups of measures,** since each index has its own advantages and drawbacks and all contribute to provide relevant and unique information on such constructs.

Lastly, and taking the findings of this doctoral thesis to the applied field, **perhaps now we are more prepared to understand what exactly happened to the events of our protagonists presented in the introduction, Ginés Navarro and Jorge Torres,** who, let us remember, suffered the consequences of the inadequate application of certain evaluation methodologies of mental fatigue and mental workload, respectively.

Beginning with the case of the accident suffered by Ginés Navarro, let us remember that the automatic fatigue recognition system of the Premium brand trailer was not able to detect the emergence of this phenomenon from the monitoring of its driver’s pupil diameter: Ginés remained with his pupils dilated, despite the fact that he was subjectively feeling fatigued. This led him to be careless about his feelings and finally resulted in a

traffic accident. Returning to the questions we posed at the beginning: what had happened? Why hadn't the trailer warned him against the emergence of mental fatigue? Why Ginés trusted this technology more than his own instincts? Could the accident have been avoided? Perhaps we now have a broader framework of knowledge to clarify these questions: What happened was a sum of factors, among which stands out the fact that **Ginés relied heavily on the trailer's fatigue detection technology**, which actually failed detecting the increasing levels of its driver's mental fatigue. The trailer could not warn the driver from the emergence of mental fatigue due to the **inappropriate use of a mental fatigue detection methodology based on a single measure**, that is, the monitoring of the driver's pupillary diameter. A type of psychophysiological measure, furthermore, that could be influenced by the action of the organism's compensatory mechanism, which is responsible for reallocating extra mental resources when the situation requires it, causing an increase in the level of activation of the central nervous system that was reflected as a pupil dilation in the driver. **Ginés trusted this system above his instincts because the temporary pressure of his work motivated him to trust this technology more than the alarm signal that the body has to warn us that mental resources are becoming depleted and which is perceived as a subjective feeling of mental fatigue.** Finally, this accident could have been avoided, mainly **if Ginés were not subjected to such temporary pressure at work** that led him to abuse his own organism to meet objectives, but it could also have been avoided **if the fatigue detection system of the trailer would have been based not on a single measure but on a triangulation of measurements.** For example: 1) a performance measure, such as the detection of unusual oscillations in the truck's trajectory; 2) a psychophysiological measure, such as blink rate, which is not based on the central nervous system activation levels but in a shift of cognitive processing strategies; and finally 3) a subjective fatigue measure, such as a fatigue reminder every 2 hours that would allow the driver to assess his own perceived mental fatigue, with suggestions to take a break that would let the driver to recover and thus avoid the emergence of fatigue and, therefore, the possible occurrence of accidents.

Finally, regarding the event of the ATCo Jorge Torres, perhaps we are also now in a better position to shed light on the possible causes of the mistake he made at work. Let us remember that Jorge failed to detect a conflict that was taking place between 2 flights under overload circumstances, however, he denied being overload at all times, despite the fact that his face was sweaty and his pupils conspicuously dilated (signs of mental

overload). What had happened? Why did Jorge claim not to be overloaded but, his psychophysiological indexes indicated the opposite? How could this dangerous situation have been avoided? In this case, we are faced with a clear example of one of the main limitations of subjective measures in general and subjective mental workload measures in particular, and that is that **subjective measures are liable to be falsified for several reasons**. For example, Jorge may have just been hired at the company and did not want to show symptoms of disability or weakness to his partner, or Jorge's excessive self-confidence may even prevent him from realizing that he really was overloaded. In other words, subjective measures are highly influenced by numerous intrinsic factors (cognitive biases, motivational conflicts, low introspection capacity, etc.), so **mental workload or mental fatigue assessment should not only be based on subjective measures**. Once more, this sort of situation could have effectively been avoided by **using a mental workload assessment methodology based on the triangulation of measures, that is, using at least one measure of each main groups of measures**: 1) a performance measure, for example the number of quasi-conflicts (number of pairs of airplanes that without having entered into conflict were close to being so); 2) a psychophysiological measure such as blink rate, which is easily measurable and is directly linked to the unconscious cognitive strategies to improve information processing during overload circumstances or the fundamental frequency of the voice, which offers numerous advantages already analysed in previous lines; and 3) a subjective measure, which allows closing the circle, offering unique information on the perception of the individual's own mental workload.



# References

---

*Referencias*



## References

- Acerra, E., Pazzini, M., Ghasemi, N., Vignali, V., Lantieri, C., Simone, A., & Babiloni, F. (2019, November). EEG-based mental workload and perception-reaction time of the drivers while using adaptive cruise control. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 226-239). Springer, Cham. [https://doi.org/10.1007/978-3-030-32423-0\\_15](https://doi.org/10.1007/978-3-030-32423-0_15)
- Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623-636. <https://doi.org/10.1016/j.ergon.2006.04.002>
- Ahsberg, E., Gamberale, F., Kjellberg, A., (1997). Perceived quality of fatigue during different occupational tasks. Development of a questionnaire. *International Journal of Industrial Ergonomics* *20*, 121–135. [https://doi.org/10.1016/S0169-8141\(96\)00044-3](https://doi.org/10.1016/S0169-8141(96)00044-3)
- Ahsberg, E., (1998). Perceived fatigue related to work. Tryckt ho CM Gruppen, Solna, Sverige.
- Anders, S., Lotze, M., Erb, M., Grodd, W., & Birbaumer, N. (2004). Brain activity underlying emotional valence and arousal: A response-related fMRI study. *Human brain mapping*, *23*(4), 200-209. <https://doi.org/10.1002/hbm.20048>
- App, E., & Debus, G. (1998). Saccadic velocity and activation: Development of a diagnostic tool for assessing energy regulation. *Ergonomics*, *41*(5), 689-697. <https://doi.org/10.1080/001401398186856>
- Backs, R. W. & Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, *23*, 243-254. [https://doi.org/10.1016/0003-6870\(92\)90152-L](https://doi.org/10.1016/0003-6870(92)90152-L)
- Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3-4), 191-204. [https://doi.org/10.1016/0025-5564\(75\)90075-9](https://doi.org/10.1016/0025-5564(75)90075-9)
- Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion*, *12*(4), 579-599. <https://doi.org/10.1080/026999398379574>
- Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science*, *5*(10), 371-372. <https://doi.org/10.3758/BF03328444>

- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J.T. Cacioppo, L.G. Tassinary, and G.G. Berntson (Ed.), *Handbook of psychophysiology* (pp. 142-162). Cambridge, UK: Cambridge University Press.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91, 276-292. <https://doi.org/10.1037/0033-2909.91.2.276>
- Becker, W. (1989). The neurobiology of saccadic eye movements. *Metrics. Reviews of oculomotor research*, 3, 13.
- Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011a). Driver workload and eye blink duration. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14(3), 199-208. <https://doi.org/10.1016/j.trf.2010.12.001>
- Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011b). Driver workload and eye blink duration. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14(3), 199-208. <https://doi.org/10.1016/j.trf.2010.12.001>
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T. & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(5), B231-B244.
- Boas, D. A., Culver, J. P., Stott, J. J., & Dunn, A. K. (2002). Three dimensional monte carlo code for photon migration through complex heterogeneous media including the adult human head. *Optics Express*, 10(3), 159-170. <https://doi.org/10.1364/OE.10.000159>
- Bodala, I. P., Li, J., Thakor, N. V., & Al-Nashash, H. (2016). EEG and eye tracking demonstrate vigilance enhancement with challenge integration. *Frontiers in Human Neuroscience*, 10, 273. <https://doi.org/10.3389/fnhum.2016.00273>
- Borghini, G., Aricò, P., Di Flumeri, G., Cartocci, G., Colosimo, A., Bonelli, S., et al. (2017). EEG-based cognitive control behaviour assessment: An ecological study with professional air traffic controllers. *Scientific Reports*, 7(1), 547. <https://doi.org/10.1038/s41598-017-00633-7>
- Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4), 602-607. <https://doi.org/10.1111/j.1469-8986.2008.00654.x>

## REFERENCES

- Bradshaw J. (1967). Pupil size as a measure of arousal during information sing. *Nature*, 216, 515-516. <https://doi.org/10.1038/216515a0>
- Bradshaw, J. (1968). Pupil size and problem solving. *Quarterly Journal of Experimental Psychology*, 20,116-122. <https://doi.org/10.1080/14640746808400139>
- Brennan, S.D., (1992). An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. DRA Technical Memorandum (CAD5) 92017, DRA Maritime Command and Control Division, Portsmouth. <https://doi.org/10.1016/j.apergo.2005.05.004>
- Cabon, P., Mollard, R., Cointot, B., Martel, A., & Besiot, P. (1997). Elaboration of a method for the assessment of psychophysiological states of ATCs in Simulation. *EEC Report*, (323).
- Cain, B. (2007). A review of the mental workload literature. *Defence Research and Development Toronto (Canada)*.
- Callan, D. J. (1999). Psychophysiological measures for human attention lapses during simulated aircraft operations. *AIR FORCE INST OF TECH WRIGHT-PATTERSONAFB OH*.
- Cameron, J., (1996). Assessing work-related body-part discomfort: current strategies and a behaviorally oriented assessment tool. *International Journal of Industrial Ergonomics* 18, 389–398. [https://doi.org/10.1016/0169-8141\(95\)00101-8](https://doi.org/10.1016/0169-8141(95)00101-8)
- Casper, P. A. (1990). Dissociations among measures of mental workload: Effects of experimenter-induced inadequacy.
- Chalder, T., Berelowitz, G., Pawlikowska, T., et al., (1993). Development of a fatigue scale. *Journal of Psychosomatic Research* 37, 147–153. [https://doi.org/10.1016/0022-3999\(93\)90081-P](https://doi.org/10.1016/0022-3999(93)90081-P)
- Christodoulou, C. (2005). The assessment and measurement of fatigue. In: J. DeLuca (Ed.), *Fatigue as a window to the brain* (pp. 19-36). Cambridge, MA: MIT Press.
- Clarke, D. D. (1999). Circulation and energy metabolism of the brain. *Basic neurochemistry: Molecular, cellular, and medical aspects*.
- Cooper, G. E., & Harper Jr, R. P. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities* (No. AGARD-567). Advisory Group for aerospace research and development Neuilly-Sur-Seine (France).
- Cope, M., Delpy, D. T., Reynolds, E., Wray, S., Wyatt, J., & Van der Zee, P. (1988). Methods of quantitating cerebral near infrared spectroscopy data. *Oxygen transport to tissue X* (pp. 183-189) Springer. [https://doi.org/10.1007/978-1-4615-9510-6\\_2](https://doi.org/10.1007/978-1-4615-9510-6_2)

- Curry, R., Jex, H., Levison, W., & Stassen, H. (1979). Final report of control engineering group. In *Mental workload* (pp. 235-252). Springer, Boston, MA. [https://doi.org/10.1007/978-1-4757-0884-4\\_13](https://doi.org/10.1007/978-1-4757-0884-4_13)
- da Silva, F. P. (2014). Mental workload, task demand and driving performance: what relation? *Procedia-Social and Behavioral Sciences*, *162*, 310-319. <https://doi.org/10.1016/j.sbspro.2014.12.212>
- Dawson, D., Ian Noy, Y., Härmä, M., Åkerstedt, T., & Belenky, G. (2011). Modelling fatigue and the use of fatigue models in work settings. *Accident Analysis & Prevention*, *43*(2), 549-564. <https://doi.org/10.1016/j.aap.2009.12.030>
- Dawson, M. E., Schell, A. M., & Fillion, D. L. (2007). The electrodermal system. *Handbook of Psychophysiology*, *2*, 200-223.
- Delliaux, S., Delaforge, A., Deharo, J. C., & Chaumet, G. (2019). Mental workload alters heart rate variability, lowering non-linear dynamics. *Frontiers in physiology*, *10*, 565. <https://doi.org/10.3389/fphys.2019.00565>
- Díaz Robredo, L. A., & Robles Sánchez, J. I. (2018). La actividad electrodérmica de la piel como indicador de activación psicofisiológica en pilotos de caza españoles: un estudio preliminar. *Sanidad Militar*, *74*(1), 7-12. <https://dx.doi.org/10.4321/s1887-85712018000100002>
- Diepeveen, H. (2020). *Vocal intensity and fundamental frequency as a measure of fatigue in an aviation setting* (Master's thesis).
- Di Flumeri, G., Aricò, P., Borghini, G., Sciaraffa, N., Ronca, V., Vozzi, A., ... & Babiloni, F. (2019, November). EEG-based workload index as a taxonomic tool to evaluate the similarity of different robot-assisted surgery systems. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 105-117). Springer, Cham. [https://doi.org/10.1007/978-3-030-32423-0\\_7](https://doi.org/10.1007/978-3-030-32423-0_7)
- Di Stasi, L. L., Catena, A., Canas, J. J., Macknik, S. L., & Martinez-Conde, S. (2013a). Saccadic velocity as an arousal index in naturalistic tasks. *Neuroscience & Biobehavioral Reviews*, *37*(5), 968-975. <https://doi.org/10.1016/j.neubiorev.2013.03.011>
- Di Stasi, L. L., Catena, A., Canas, J. J., Macknik, S. L., & Martinez-Conde, S. (2013b). Saccadic velocity as an arousal index in naturalistic tasks. *Neuroscience & Biobehavioral Reviews*, *37*(5), 968-975. <https://doi.org/10.1016/j.neubiorev.2013.03.011>

## REFERENCES

- Di Stasi, L. L., Marchitto, M., Antolí, A., Baccino, T., & Cañas, J. J. (2010). Approximation of on-line mental workload index in ATC simulated multitasks. *Journal of Air Transport Management*, 16(6), 330-333. <https://doi.org/10.1016/j.jairtraman.2010.02.004>
- Di Stasi, L. L., Marchitto, M., Antolí, A., & Cañas, J. J. (2013). Saccadic peak velocity as an alternative index of operator attention: A short review. *Revue Européenne De Psychologie Appliquée/European Review of Applied Psychology*, 63(6), 335-343. <https://doi.org/10.1016/j.erap.2013.09.001>
- Di Stasi, L. L., Renner, R., Catena, A., Cañas, J. J., Velichkovsky, B. M., & Pannasch, S. (2012). Towards a driver fatigue test based on the saccadic main sequence: A partial validation by subjective report data. *Transportation Research Part C: Emerging Technologies*, 21(1), 122-133. <https://doi.org/10.1016/j.trc.2011.07.002>
- Di Stasi, L. L., Renner, R., Staehr, P., Helmert, J. R., Velichkovsky, B. M., Cañas, J. J., et al. (2010). Saccadic peak velocity sensitivity to variations in mental workload. *Aviation, Space, and Environmental Medicine*, 81(4), 413-417. <https://doi.org/10.3357/ASEM.2579.2010>
- Dittner, A. J., Wessely, S. C., & Brown, R. G. (2004). The assessment of fatigue: a practical guide for clinicians and researchers. *Journal of psychosomatic research*, 56(2), 157-170. [https://doi.org/10.1016/S0022-3999\(03\)00371-4](https://doi.org/10.1016/S0022-3999(03)00371-4)
- Durantín, G., Gagnon, J. F., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural brain research*, 259, 16-23. <https://doi.org/10.1016/j.bbr.2013.10.042>
- Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). Workload assessment in multi-task environments. *Multiple-task performance*, 207-216.
- Elshtain, L. y Schaefer, T. (1968). Effects of storage load and word frequency on pupillary responses during short-term memory. *Psychonomic Science*, 12,143-144. <https://doi.org/10.3758/BF03331240>
- Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human factors*, 37(2), 381-394. <https://doi.org/10.1518/001872095779064555>
- Fan, J., & Smith, A. P. (2017, June). The impact of workload and fatigue on performance. In *International symposium on human mental workload: Models and applications* (pp. 90-105). Springer, Cham. [https://doi.org/10.1007/978-3-319-61061-0\\_6](https://doi.org/10.1007/978-3-319-61061-0_6)

- Galley, N., & Andres, G. (1996). Saccadic eye movements and blinks during long-term driving on the autobahn with minimal alcohol ingestion. *Vision in Vehicles*, 5, 381-388.
- Galley, N. (1989). Saccadic eye movement velocity as an indicator of (de) activation. A review and some speculations.
- Galley, N. (1993). The evaluation of the electrooculogram as a psychophysiological measuring instrument in the driver study of driver behaviour. *Ergonomics*, 36(9), 1063-1070. <https://doi.org/10.1080/00140139308967978>
- Galley, N. (1998). An enquiry into the relationship between activation and performance using saccadic eye movement parameters. *Ergonomics*, 41(5), 698-720. <https://doi.org/10.1080/001401398186865>
- Gilbert, D. (2009). *Stumbling on happiness*. Vintage: Canada.
- Gillberg, M., Kecklund, G., Akerstedt, T., 1994. Relations between performance and subjective ratings of sleepiness during a night awake. *Sleep* 17, 236–241. <https://doi.org/10.1093/sleep/17.3.236>
- Goldmark, J. C. (1912). *Fatigue Efficiency*. Charities publication committee.
- Goldwater, B. (1972). Psychological significance of pupillary movements. *Psychological Bulletin*, 77 (5), 340-355. <https://doi.org/10.1037/h0032456>
- González Gutiérrez, J. L., Moreno Jiménez, B., Garrosa Hernández, E., & López López, A. (2005). Carga mental y fatiga en servicios especiales de enfermería. *Revista latinoamericana de psicología*, 37(3), 477-492.
- Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*, Vol. 2. *Cognitive processes and performance* (p. 1–49). John Wiley & Sons
- Gould, K. S., Røed, B. K., Saus, E. R., Koefoed, V. F., Bridger, R. S., & Moen, B. E. (2009). Effects of navigation method on workload and performance in simulated high-speed ship navigation. *Applied ergonomics*, 40(1), 103-114. <https://doi.org/10.1016/j.apergo.2008.01.001>
- Grandjean, E. (1989). *Fitting the task to the man: A textbook of occupational ergonomics* (4th ed.) Taylor & Francis/Hemisphere.
- Granholm, E., Asarnow, R. F., Sarkin, A. J. & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, 33, 457-461. <https://doi.org/10.1111/j.1469-8986.1996.tb01071.x>



## REFERENCES

- Grech, M. R., Neal, A., Yeo, G., Humphreys, M., & Smith, S. (2009). An examination of the relationship between workload and fatigue within and across consecutive days of work: Is the relationship static or dynamic? *Journal of occupational health psychology, 14*(3), 231. <https://doi.org/10.1037/a0014952>
- Hagmüller, M., Rank, E., & Kubin, G. (2006). Evaluation of the human voice for indications of workload-induced stress in the aviation environment. *EEC Note, 18*(06).
- Hancock, P. A., & Matthews, G. (2019). Workload and performance: Associations, insensitivities, and dissociations. *Human factors, 61*(3), 374-392 <https://doi.org/10.1177/0018720818809590>
- Hancock, P. A. (2017, June). Whither workload? Mapping a path for its future development. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 3-17). Springer, Cham (2010). [https://doi.org/10.1007/978-3-319-61061-0\\_1](https://doi.org/10.1007/978-3-319-61061-0_1)
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- He, X., Wang, L., Gao, X., & Chen, Y. 2012. The eye activity measurement of mental workload based on basic flight task. In *Industrial Informatics (INDIN), 2012 10th IEEE International Conference on*, 502–507. IEEE. <https://doi.org/10.1109/INDIN.2012.6301203>
- Henelius, A., Hirvonen, K., Holm, A., Korpela, J., & Muller, K. (2009, September). Mental workload classification using heart rate metrics. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1836-1839). IEEE. <https://doi.org/10.1109/IEMBS.2009.5332602>
- Hess, E. H. & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*,349-350. <https://doi.org/10.1126/science.132.3423.349>
- Hess, E. H. (1965). Attitude and pupil size. *Scientific American, 212*,46-54.
- Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45*(1), 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. J. (2011). A motivational control theory of cognitive fatigue. *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications*, 167-87. <https://doi.org/10.1037/12343-008>

- Hockey, R. (2013). *The psychology of fatigue: Work, effort and control* Cambridge University Press.
- Hoeks, B., & Levelt, W. J. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*, 25(1), 16-26. <https://doi.org/10.3758/BF03204445>
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures* OUP Oxford.
- Hopstaken, J. F., van der Linden, D., Bakker, A. B., Kompier, M. A., & Leung, Y. K. (2016). Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. *Journal of Experimental Psychology. Human Perception and Performance*, <https://doi.org/10.1037/xhp0000189>
- Horrey, W. J., Lesch, M. F., & Garabet, A. (2009). Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. *Journal of safety research*, 40(1), 7-12. <https://doi.org/10.1016/j.jsr.2008.10.011>
- Hyönä, J., Tommola, J., & Alaja, A. M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, 48(3), 598-612. <https://doi.org/10.1080/14640749508401407>
- Iqbal, S. T., Zheng, X. S. y Bailey, B. P. (2004). Task evoked pupillary response to mental workload in human-computer interaction. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1477-1480. <https://doi.org/10.1145/985921.986094>
- Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., & Chance, B. (2004). Functional optical brain imaging using near-infrared during cognitive tasks. *International Journal of Human-Computer Interaction*, 17(2), 211-227. [https://doi.org/10.1207/s15327590ijhc1702\\_6](https://doi.org/10.1207/s15327590ijhc1702_6)
- Jones, N. P., Siegle, G. J., Muelly, E. R., Haggerty, A. y Ghinassi, F. (2010). Poor performance on cognitive tasks in depression: Doing too much or not enough? *Cognitive, Affective, and Behavioral Neuroscience*, 10, 129-140. <https://doi.org/10.3758/CABN.10.1.129>

## REFERENCES

- Jorna, P. G. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological psychology*, 34(2-3), 237-257. [https://doi.org/10.1016/0301-0511\(92\)90017-O](https://doi.org/10.1016/0301-0511(92)90017-O)
- Josten, E. J., Ng-A-Tham, J. E., & Thierry, H. (2003). The effects of extended workdays on fatigue, health, performance and satisfaction in nursing. *Journal of advanced nursing*, 44(6), 643-652. <https://doi.org/10.1046/j.0309-2402.2003.02854.x>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4), 329. <https://doi.org/10.1037/0033-295X.87.4.329>
- Hahnemann, D., & Beatty, J. (1967). Pupillary responses in a pitch-discrimination task. *Perception & Psychophysics*, 2(3), 101-105. <https://doi.org/10.3758/BF03210302>
- Kahneman, D., Onuska, L. & Wolman, R. (1968). Effects of grouping on the pupillary response in a short-term memory task. *Quarterly Journal of Experimental Psychology*, 20, 309-311. <https://doi.org/10.1080/14640746808400168>
- Kahneman, D., Peavler, W. S. & Onuska, L. (1968). Effects of verbalization and incentive on the pupil response to mental activity. *Canadian Journal of Psychology*, 22, 186-196. <https://doi.org/10.1037/h0082759>
- Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.
- Kamzanova, A. T., Kustubayeva, A. M., & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human factors*, 56(6), 1136-1149. <https://doi.org/10.1177/0018720814526617>
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S., Hudspeth, A. J., & Mack, S. (Eds.). (2000). *Principles of neural science* (Vol. 4, pp. 1227-1246). New York: McGraw-hill.
- Kartali, A., Janković, M. M., Gligorijević, I., Mijović, P., Mijović, B., & Leva, M. C. (2019, November). Real-time mental workload estimation using eeg. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 20-34). Springer, Cham. [https://doi.org/10.1007/978-3-030-32423-0\\_2](https://doi.org/10.1007/978-3-030-32423-0_2)
- Kashiwagi, S., (1969). Psychological rating of human fatigue. In: *Hashimoto, K., Kogi, K., Grandjean, E. (Eds.), Methodology in Human Fatigue Assessment. Taylor & Francis, Kyoto*, pp. 17–21. <https://doi.org/10.1080/00140137108931218>

- Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., & Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. *Biological psychology, 102*, 118-129.
- Kensinger, E. A. (2004). Remembering emotional experiences: The contribution of valence and arousal. *Reviews in the Neurosciences, 15*(4), 241-252. <https://doi.org/10.1515/REVNEURO.2004.15.4.241>
- Kinsman, R., Weiser, P., (1976). Subjective symptomatology during work and fatigue. In: *Simonson, E., Weiser, P. (Eds.), Psychological Aspects and Physiological Correlates of Work and Fatigue. CC Thomas, Springfield*, pp. 336-405.
- Kirschenbaum, A., Oigenblick, L., & Goldberg, A. I. (2000). Well being, work environment and work accidents. *Social Science & Medicine, 50*(5), 631-639. [https://doi.org/10.1016/S0277-9536\(99\)00309-3](https://doi.org/10.1016/S0277-9536(99)00309-3)
- Kirwan, B., Evans, A., Donohoe, L., Kilner, A., Lamoureux, T., Atkinson, T., et al. (1997). Human factors in the ATM system design life cycle. *Proceedings of EUROCONTROL/FAA ATM R & D Seminar*, pp. 16-20.
- Kuo, J., Lenné, M. G., Myers, R., Collard-Scruby, A., Jaeger, C., & Birmingham, C. (2017, September). Real-Time Assessment of Operator State in Air Traffic Controllers Using Ocular Metrics. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, No. 1, pp. 257-261). Sage CA: Los Angeles, CA: SAGE Publications. <https://doi.org/10.1177/1541931213601547>
- Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry A window to the preconscious? *Perspectives on Psychological Science, 7*(1), 18-27. <https://doi.org/10.1177/1745691611427305>
- Ledger, H. (2013). The effect cognitive load has on eye blinking. *The Plymouth Student Scientist, 6*(1), 206-223.
- Lee, C. W., & Cuijpers, P. (2013). A meta-analysis of the contribution of eye movements in processing emotional memories. *Journal of Behavior Therapy and Experimental Psychiatry, 44*(2), 231-239. <https://doi.org/10.1016/j.jbtep.2012.11.001>
- Lee, K. A., Hicks, G., & Nino-Murcia, G. (1991). Validity and reliability of a scale to assess fatigue. *Psychiatry research, 36*(3), 291-298. [https://doi.org/10.1016/0165-1781\(91\)90027-M](https://doi.org/10.1016/0165-1781(91)90027-M)
- Lenné, M. G., & Jacobs, E. E. (2016). Predicting drowsiness-related driving events: a review of recent research methods and future opportunities. *Theoretical Issues in*

REFERENCES

- Ergonomics Science*, 17(5-6), 533-553.  
<https://doi.org/10.1080/1463922X.2016.1155239>
- Leplat, J. (2002). *Éléments pour une histoire de la notion de charge mentale. Charge mentale: notion floue et vrai problème. Toulouse: Octarès*
- Lilley, R., Feyer, A. M., Kirk, P., & Gander, P. (2002). A survey of forest workers in New Zealand: Do hours of work, rest, and recovery play a role in accidents and injury? *Journal of safety research*, 33(1), 53-71. [https://doi.org/10.1016/S0022-4375\(02\)00003-8](https://doi.org/10.1016/S0022-4375(02)00003-8)
- Longo, L., & Leva, M. C. (Eds.). (2017). *Human Mental Workload: Models and Applications: First International Symposium, H-WORKLOAD 2017, Dublin, Ireland, June 28-30, 2017, Revised Selected Papers (Vol. 726)*. Springer.
- Longo, L., & Leva, M. C. (Eds.). (2018). *Human Mental Workload: Models and Applications: Second International Symposium, H-WORKLOAD 2018, Amsterdam, the Netherlands, September 20-21, 2018, Revised Selected Papers (Vol. 1012)*. Springer.
- Longo, L., & Leva, M. C. (Eds.). (2019). *Human Mental Workload: Models and Applications: Third International Symposium, H-WORKLOAD 2019, Rome, Italy, November 14-15, 2019, Proceedings (Vol. 1107)*. Springer Nature.
- Lorist, M. M., Boksem, M. A., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research*, 24(2), 199-205. <https://doi.org/10.1016/j.cogbrainres.2005.01.018>
- Lynn, R. (1966). *Attention, arousal, and the orientation reaction*. Oxford: Pergamon Press.
- Mansour, A., Riad, W., & Moussa, A. (2010). The occupational fatigue in anesthesiologists: illusion or real? *Middle East journal of anaesthesiology*, 20(4), 529-534.
- Matthews, G., & Desmond, P. A. (1998). Personality and multiple dimensions of task-induced fatigue: A study of simulated driving. *Personality and Individual Differences*, 25(3), 443-458. [https://doi.org/10.1016/S0191-8869\(98\)00045-2](https://doi.org/10.1016/S0191-8869(98)00045-2)
- Matthews, G., Middleton, W., Gilmartin, B. & Bullimore, M. A. (1991). Pupillary diameter and cognitive and cognitive load. *Journal of Psychophysiology*, 5, 265-271.
- Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich IV, J. (2015). The psychometrics of mental workload: Multiple measures are sensitive but

- divergent. *Human Factors*, 57(1), 125-143.  
<https://doi.org/10.1177/0018720814539505>
- McKendrick, R. D. (2016). *Nonlinear hemoDynamics of Workload and Workload Transit* (Doctoral dissertation, George Mason University).
- Meshkati, N. (1988). Heart rate variability and mental workload assessment. In *Advances in psychology* (Vol. 52, pp. 101-115). North-Holland.  
[https://doi.org/10.1016/S0166-4115\(08\)62384-5](https://doi.org/10.1016/S0166-4115(08)62384-5)
- Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the brain takes a break: a model-based analysis of mind wandering. *Journal of Neuroscience*, 34(49), 16286-16295.  
<https://doi.org/10.1523/JNEUROSCI.2062-14.2014>
- Monk, T., Folkard, S., (1985). Individual differences in shiftwork adjustment. In: *Folkard, S., Monk, T. (Eds.), Hours of Work. Temporal Factors in Work Scheduling. Wiley, Chichester*, pp. 227–237.
- Moray, N. (1979). Models and measures of mental workload. In *Mental workload* (pp. 13-21). Springer, Boston, MA. [https://doi.org/10.1007/978-1-4757-0884-4\\_2](https://doi.org/10.1007/978-1-4757-0884-4_2)
- Mulder, L. J. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological psychology*, 34(2-3), 205-236.  
[https://doi.org/10.1016/0301-0511\(92\)90016-N](https://doi.org/10.1016/0301-0511(92)90016-N)
- Munoz-de-Escalona, E., & Canas, J. J. (2017). Online measuring of available resources. In *H-workload 2017: the first international symposium on human mental work*.
- Muñoz-de-Escalona, E., & Cañas, J. J. (2018, September). Latency differences between mental workload measures in detecting workload changes. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 131-146). Springer, Cham. [https://doi.org/10.1007/978-3-030-14273-5\\_8](https://doi.org/10.1007/978-3-030-14273-5_8)
- Muñoz-de-Escalona, E., Cañas, J. J., Leva, C., & Longo, L. (2020, December). Task Demand Transition Peak Point Effects on Mental Workload Measures Divergence. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 207-226). Springer, Cham. [https://doi.org/10.1007/978-3-030-62302-9\\_13](https://doi.org/10.1007/978-3-030-62302-9_13)
- Muñoz-de-Escalona, E., Cañas, J. J., & Noriega, P. (2020). Inconsistencies between mental fatigue measures under compensatory control theories. *Psicológica Journal*, 41, 103-126. <https://doi.org/10.2478/psicolj-2020-0006>
- Muñoz-de-Escalona, E., Cañas, J. J., & Morales-Guaman, J. F. (2020, December). Fundamental Frequency as an Alternative Method for Assessing Mental Fatigue. In

## REFERENCES

- International Symposium on Human Mental Workload: Models and Applications* (pp. 58-75). Springer, Cham. [https://doi.org/10.1007/978-3-030-62302-9\\_4](https://doi.org/10.1007/978-3-030-62302-9_4)
- Muñoz-de-Escalona, E., Cañas, J. J., & van Nes, J. (2019, November). Task Demand Transition Rates of Change Effects on Mental Workload Measures Divergence. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 48-65). Springer, Cham. [https://doi.org/10.1007/978-3-030-32423-0\\_4](https://doi.org/10.1007/978-3-030-32423-0_4)
- Murata, A. (2005). An attempt to evaluate mental workload using wavelet transform of EEG. *Human Factors*, 47(3), 498-508. <https://doi.org/10.1518/001872005774860096>
- Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, 48(11), 1532-1543. <https://doi.org/10.1111/j.1469-8986.2011.01226.x>
- Novstrup, A., Goan, T., & Heaton, J. (2018, September). Workload Assessment Using Speech-Related Neck Surface Electromyography. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 72-91). Springer, Cham. [https://doi.org/10.1007/978-3-030-14273-5\\_5](https://doi.org/10.1007/978-3-030-14273-5_5)
- Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 33(1), 17-33. <https://doi.org/10.1177/001872089103300102>
- O'Donnell, R., & Eggemeier, F. (1986). Workload assessment methodology. Handbook of Perception and Human Performance. Volume 2. *Cognitive Processes and Performance*. KR Boff, L. Kaufman and JP Thomas
- Okogbaa, O., Shell, R., Filipusic, D., (1994). On the investigation of the neurophysiological correlates of knowledge worker mental fatigue using the EEG signal. *Applied Ergonomics* 25, 355–365. [https://doi.org/10.1016/0003-6870\(94\)90054-X](https://doi.org/10.1016/0003-6870(94)90054-X)
- Paxion, J., Galy, E., & Berthelon, C. (2014). Mental workload and driving. *Frontiers in psychology*, 5, 1344. <https://doi.org/10.3389/fpsyg.2014.01344>
- Payne, D. T., Parry, M. E. & Harasymiw, S. J. (1968). Percentage of pupillary dilation as a measure of item difficulty. *Perception and Psychophysics* 4, 139-143. <https://doi.org/10.3758/BF03210453>

- Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology*, 11, 559-566. <https://doi.org/10.1111/j.1469-8986.1974.tb01114.x>
- Piquado, T., Isaacowitz, D. & Wingfield A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, 47, 1-10. <https://doi.org/10.1111/j.1469-8986.2009.00947.x>
- Porges, S. W., & Byrne, E. A. (1992). Research methods for measurement of heart rate and respiration. *Biological Psychology*, 34(2-3), 93-130.
- Porter, G., Troscianko, T. & Gilchrist, I. D. (2007). Effort during visual search and counting: insights from pupillometry. *Quarterly Journal of Experimental Psychology*, 60,211-229. <https://doi.org/10.1080/17470210600673818>
- Privitera, C. M., Renninger, L. W., Carney, T., Klein, S. & Aguilar, M. (2010). Pupil dilation during visual target detection. *Journal of Vision*, 10,1-14. <https://doi.org/10.1167/10.10.3>
- Priya, D. S., Johnson, P., Padmavathi, R., Subhashini, A. S., Ayyappan, R., & Ayyappan, M. (2010). Evaluation of the relationship between workload and work capacity in petrochemical and tannery workers-A pilot study. *Life Sciences and Medicine Research*, 19(1), 2-12
- Rabinach, A., (1992). *The Human Motor-energy, Fatigue and the Origins of Modernity*. University of California Press, Berkeley.
- Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, (3), 257-266. <https://doi.org/10.1109/TSMC.1983.6313160>
- Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in Psychology*, 52, 185-218. [https://doi.org/10.1016/S0166-4115\(08\)62387-0](https://doi.org/10.1016/S0166-4115(08)62387-0)
- Reiner, M., & Gelfeld, T. M. (2014). Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *International Journal of Psychophysiology*, 93(1), 38-44. <https://doi.org/10.1016/j.ijpsycho.2013.11.002>
- Rimehaug, T., Svebak, S., (1987). Psychogenic muscle tension: the significance of motivation and negative affect in perceptual-cognitive task performance. *International Journal of Psychophysiology* 5, 97–106. [https://doi.org/10.1016/0167-8760\(87\)90014-6](https://doi.org/10.1016/0167-8760(87)90014-6)



## REFERENCES

- Rodgers, S. H. (2000). An Ergonomic Approach to Analyzing Workplace Accidents. *Applied occupational and environmental hygiene*, 15(7), 529-534. <https://doi.org/10.1080/10473220050028321>
- Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight: A decade of practical use* (No. RAE-TR-90019). ROYAL AEROSPACE ESTABLISHMENT FARNBOROUGH (UNITED KINGDOM).
- Roscoe, A. H. (1987). *In-flight assessment of workload using pilot ratings and heart rate*. BRITANNIA AIRWAYS LTD LUTON (UNITED KINGDOM).
- Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration?. *Biological psychology*, 34(2-3), 259-287. [https://doi.org/10.1016/0301-0511\(92\)90018-P](https://doi.org/10.1016/0301-0511(92)90018-P)
- Rubio, S., Díaz, E., & MARTÍN, J. (1999). La medida de la carga mental de trabajo II: Procedimientos subjetivos. *Boletín Digital de Factores Humanos* (21).
- Saito, Y., Kashiwagi, S., (1970). Factors underlying subjective feelings of fatigue. *Journal of Science Labour* 46, 205–224.
- Sarsangi, V., Salehiniya, H., Hannani, M., Marzaleh, M. A., Abadi, Y. S., Honarjoo, F. & Derakhshanjazari, M. (2017). Assessment of workload effect on nursing occupational accidents in hospitals of Kashan, Iran. *Biomedical Research and Therapy*, 4(8), 1527-1540. <https://doi.org/10.15419/bmrat.v4i08.226>
- Schaefer, T., Ferguson, J. B., Klein, J. A. y Rawson, E. B. (1968). Pupillary responses during mental activities. *Psychonomic Science*, 12,137-138. <https://doi.org/10.3758/BF03331236>
- Schaeffer, S., Darby, L., Browder, K., (1995). Perceived exertion and metabolic responses of women during aerobic dance exercise. *Perceptual and Motor Skills* 81, 671–700. <https://doi.org/10.1177/003151259508100265>
- Scherer, K. R. (2000). Psychological models of emotion. *The neuropsychology of emotion*, 137(3), 137-162.
- Schleicher, R., Galley, N., Briest, S., & Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired? *Ergonomics*, 51(7), 982-1010. <https://doi.org/10.1080/00140130701817062>
- Schneider, M., & Deml, B. (2016). An integrated approach of mental workload assessment. *Advances in ergonomic design of systems, products and processes* (pp. 191-208) Springer. [https://doi.org/10.1007/978-3-662-48661-0\\_13](https://doi.org/10.1007/978-3-662-48661-0_13)

- Smets, E., Garssen, B., Bonke, B., De Haes, J., (1995). The multidimensional fatigue inventory (MFI). Psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research* 39, 315–325. [https://doi.org/10.1016/0022-3999\(94\)00125-O](https://doi.org/10.1016/0022-3999(94)00125-O)
- Smets, E., Garssen, B., Cull, A., De Haes, J., (1996). Application of the multidimensional fatigue inventory (MFI-20) in cancer patients receiving radiotherapy. *British Journal of Cancer* 73, 241–245. <https://doi.org/10.1038/bjc.1996.42>
- Smith, A. P., & Smith, H. N. (2017, June). Workload, fatigue and performance in the rail industry. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 251-263). Springer, Cham. [https://doi.org/10.1007/978-3-319-61061-0\\_17](https://doi.org/10.1007/978-3-319-61061-0_17)
- Sokolov, E. N. (1963). Perception and the conditioned reflex. *Oxford: Pergamon Press*.
- So, W. K., Wong, S. W., Mak, J. N., & Chan, R. H. (2017). An evaluation of mental workload with frontal EEG. *PloS one*, 12(4), e0174949. <https://doi.org/10.1371/journal.pone.0174949>
- Steinhauer, S. R., Siegle, G. J., Condray, R. & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, 52,77-86. <https://doi.org/10.1016/j.ijpsycho.2003.12.005>
- Strangman, G., Boas, D. A., & Sutton, J. P. (2002). Non-invasive neuroimaging using near-infrared light. *Biological Psychiatry*, 52(7), 679-693. [https://doi.org/10.1016/S0006-3223\(02\)01550-0](https://doi.org/10.1016/S0006-3223(02)01550-0)
- Taylor, A. H., & Dorn, L. (2006). Effects of physical inactivity on stress, fatigue, health and risk of at-work road traffic accidents. *Physical inactivity and road accidents*, 27, 371-391. <https://doi.org/10.1146/annurev.publhealth.27.021405.102117>
- Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine*, 78(5), B176-B185.
- Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39(3), 358-381. <https://doi.org/10.1080/00140139608964470>
- Tsang, P. S., & Wilson, G. (1997). Mental workload. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 417-449) John Wiley & sons Canada.

## REFERENCES

- Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human factors*, *43*(1), 111-121. <https://doi.org/10.1518/001872001775992570>
- Verney S. P., Granholm, E. & Marshall S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *International Journal of Psychophysiology*, *52*,23-36. <https://doi.org/10.1016/j.ijpsycho.2003.12.003>
- Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2), 167-190. <https://doi.org/10.1016/j.trf.2005.04.014>
- Vidulich, M. A., & Tsang, P. S. (1987, September). Absolute magnitude estimation and relative judgement approaches to subjective workload assessment. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 31, No. 9, pp. 1057-1061). Sage CA: Los Angeles, CA: SAGE Publications. <https://doi.org/10.1177/154193128703100930>
- Vogt, J., Hagemann, T., & Kastner, M. (2006). The impact of workload on heart rate and blood pressure in en-route and tower air traffic control. *Journal of Psychophysiology*, *20*(4), 297.
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, *50*(3), 449-455. <https://doi.org/10.1027/0269-8803.20.4.297>
- Wang, Y., Reimer, B., Dobres, J., & Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*, 227-237. <https://doi.org/10.1016/j.trf.2014.08.003>
- Wang, L. (1998). Glissadic saccades: A possible measure of vigilance. *Ergonomics*, *41*(5), 721-732. <https://doi.org/10.1080/001401398186874>
- Wickens, C. D. (1984). Processing resources in attention. *Varieties of attention*, *40*, 63-98.
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, *50*(3), 449-455. <https://doi.org/10.1518%2F001872008X288394>
- Wickens, C. D. (2017, June). Mental workload: assessment, prediction and consequences. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 18-29). Springer, Cham. [https://doi.org/10.1007/978-3-319-61061-0\\_2](https://doi.org/10.1007/978-3-319-61061-0_2)

- Wientjes, C. J. (1992). Respiration in psychophysiology: methods and applications. *Biological psychology*, *34*(2-3), 179-203.
- Yeh, Y. Y., & Wickens, C. D. (1984). *The dissociation of subjective measures of mental workload and performance* (No. EPL-84-2). National Aeronautics and Space Administration, Ames Research Center.
- Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, *30*(1), 111-120..  
<https://doi.org/10.1177/001872088803000110>
- 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>
- Young, M. S., & Stanton, N. A. (2001). Mental workload: theory, measurement, and application. *International encyclopedia of ergonomics and human factors*, *1*, 507-509.
- Young, M. S., & Stanton, N. A. (2002). Attention and automation: new perspectives on mental underload and performance. *Theoretical issues in ergonomics science*, *3*(2), 178-194. <https://doi.org/10.1080/14639220210123789>
- Yoshino, K., Edamatsu, M., Yoshida, M., & Matsuoka, K. (2007). An algorithm for detecting startle state based on physiological signals. *Accident Analysis and Prevention*, *39*(2), 308-312. <https://doi.org/10.1016/j.aap.2006.08.003>
- Yoshitake, H., (1978). Three characteristic patterns of subjective fatigue symptoms. *Ergonomics* *21*, 231–233. <https://doi.org/10.1080/00140137808931718>
- Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O. N. M., & Atkins, M. S. (2012). Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surgical endoscopy*, *26*(10), 2746-2750. <https://doi.org/10.1007/s00464-012-2268-6>
- Zuber, B. L., Stark, L., & Cook, G. (1965). Microsaccades and the velocity-amplitude relationship for saccadic eye movements. *Science*, *150*(3702), 1459-1460.  
<https://doi.org/10.1126/science.150.3702.1459>