Universidade do Minho
Escola de Engenharia

André Pimenta Ribeiro

Detecting and Monitoring Mental Fatigue: A Non-invasive Approach

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

COMPETE
PROGRAMA OPERACIONAL FACTORES DE COMPETITIVIDADE

QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007.2013

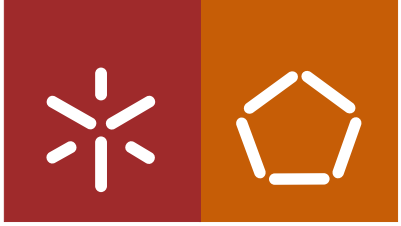
Governo da República Portuguesa

UNIÃO EUROPEIA
Fundo Social Europeu

André Pimenta Ribeiro **Detecting and Monitoring Mental Fatigue: A Non-invasive Approach**

UMinho | 2018

january 2018



Universidade do Minho
Escola de Engenharia

André Pimenta Ribeiro

Detecting and Monitoring Mental Fatigue: A Non-invasive Approach

Doctoral Thesis
Doctoral Degree in Informatics

Supervised by:
Professor Doutor Paulo Jorge Freitas de Oliveira Novais
Professor Doutor José Carlos Ferreira Maia Neves

january 2018

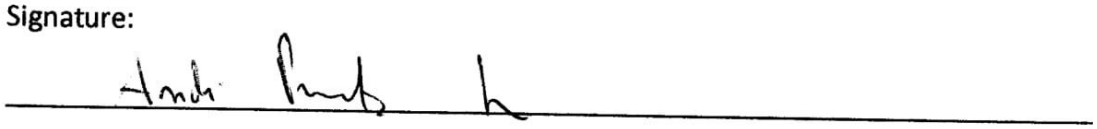
STATEMENT OF INTEGRETY

I hereby declare having concluded my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of elaboration of this thesis.
I further declare that I have fully acknowledged the code of ethical conduct of the University of Minho.

University of Minho 15/01/2018

Full Name: André Pimenta Ribeiro

Signature:



ACKNOWLEDGEMENTS

"The improvement of understanding is for two ends: first, our own increase of knowledge; secondly, to enable us to deliver that knowledge to others." - John Locke

This thesis is the result of the contribution of different people who directly or indirectly followed me during this long journey, contributed significantly to its accomplishment. I particularly highlight, viz.

- Professor Paulo Novais for all the availability and supervision of this work, in particular for the advice and suggestions, support and incentive to go further. Without it, this work would not become real;
- Professor Jose Maia Neves, who has a great critical sense, forcing me to improve, by sharing knowledge and advice that only someone with a great experience and knowledge has;
- To my friend Dr. James Miller, for accepting to help and teaching for the pleasure of sharing knowledge, but also for his great passion he has on the subject of fatigue and human factors. Thank you very much;
- To the colleagues and team of Performetric and ISLab. In particular Serafim Pinto, Daniel Carvalho, Miguel Pinto and Davide Carneiro who worked daily with me, without

forgetting Marco Gomes, Ângelo Costa, Fábio Silva and Tiago Oliveira and Professor Cesar Analide for the support through their ideas, discussions and experience, as well as the collaboration that allowed me to develop my work;

- To all the colleagues I met during my graduation at the University of Minho, especially those who started with me their academic career, my colleagues and all the colleagues I met through associativism, namely within CeSIUM.

Last but not least, I would like to dedicate this thesis and all the effort to my parents Ana Pimenta and Amilcar Ribeiro, my brother Jaime Ribeiro and my companion Francisca Guedes for the dedication, support and encouragement they have always given me, for being a safe haven that allowed me to complete this work. Thanks for being here.

This work was developed in the context of the project CAMCoF – Contextaware Multimodal Communication Framework funded by ERDF – European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-028980.



Título: Deteção e Monitorização de Fadiga Mental – Uma Abordagem não Invasiva

Resumo - No dia-a-dia, não raras vezes se tem a sensação de estarmos esgotados, quer devido ao trabalho mental ou físico, a qual se intui como sendo o processo de deterioração do desempenho na realização de simples tarefas. A capacidade de trabalho, assim como a atuação de um indivíduo, seja física ou mental, geralmente diminui à medida que o dia avança. A perda destas faculdades pode ser associada ao surgir de fadiga, o que é especialmente percetível em trabalhos longos e exigentes, ou mesmo em tarefas repetitivas. No entanto, uma boa gestão do tempo de trabalho, do esforço investido em cada tarefa, bem como o efeito de pausas no labor, pode resultar num aumento de exercício intelectual, compensando parcialmente os efeitos da fadiga.

O propósito desta tese de doutoramento é, portanto, a análise e criação de soluções capazes de avaliar o comportamento humano, e em particular a fadiga mental, de forma não-invasiva e sem interferência nas atividades quotidianas recorrendo a técnicas de Inteligência Artificial para a resolução de problemas.

Palavras Chave - Fadiga Mental, Deteção e Medida de Fadiga, Sonolência, Saúde Mental, Medidas de Desempenho, Stress, Carga de Trabalho, Inteligência Artificial, Padrões de Interação

Title: Detecting and Monitoring Mental Fatigue – A Non-Invasive Approach

Abstract - On a daily basis, people at work often feel a sense of exhaustion due to mental or physical work, with a degraded sense of accomplishment in performing simple tasks. The working capacity and the individual performance, whether physical or mental, generally decreases as the day goes by. The loss of these capabilities may be associated with the onset of fatigue, which is especially noticeable in long and demanding jobs or repetitive tasks. However, effective work and time management, as well as the effect of breaks, can result in better mental outcomes, partially offsetting the effects of fatigue.

Indeed, the object of this doctoral thesis will be centered on the analysis and creation of solutions to assess the human behavior, and in particular mental fatigue in a non-invasive way and without the interference of daily routines. It will be grounded on an Artificial Intelligence based approach to problem solving.

Keywords - Mental Fatigue, Sleepiness, Fatigue Measurement, Mental Health, Performance Measures, Stress, Workload, Artificial Intelligence, Machine Learning, Interaction Patterns.

Contents

1 Introduction 1

1.1 Fatigue 3

1.1.1 What is Fatigue? 3

1.1.2 Physical Fatigue 9

1.2 Detecting Mental Fatigue 10

1.2.1 Performance Measures 11

1.2.2 Subjective Ratings of Mental Workload, Mental Fatigue, and Sleepiness 18

1.2.3 Physiological measures 23

1.2.4 Systems for Detecting and Monitoring Mental Fatigue 24

1.2.5 Summary 33

1.3 Artificial Intelligence in Fatigue Detection and Monitoring 34

1.3.1 How Artificial Intelligence Can Improve Fatigue Detection and Monitoring 35

1.3.2 The Intersection of AI with Fatigue Detection and Monitoring 35

1.3.3 Summary 39

1.4 Discussion 40

1.4.1 The Current State 40

1.4.2 Improving Non-Intrusive and Non-Invasive Fatigue Detection 41

1.5 Investigation Hypothesis 42

1.6 Objectives 43
1.7 Research Methodology 44
1.8 Work Programme 44
1.9 Structure of the Document 46
2 Publications Composing the Doctoral Thesis 47
2.1 Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns 47
2.2 Analysis of human performance as a measure of mental fatigue 58
2.3 A neural network to classify fatigue from human–computer interaction 72
2.4 A multi-modal architecture for non-intrusive analysis of performance in the workplace 87
2.5 Non-intrusive quantification of performance and its relationship to mood 94
2.6 Improving User Privacy and the Accuracy of User Identification in Behavioral Biometrics 102
2.7 Mental workload management as a tool in e-learning scenarios 115
3 Conclusions 125
3.1 Contributions 126
3.2 Validation of the Research Hypothesis 128
3.3 Dissemination of Results 129
3.3.1 Other Publications 129
3.3.2 Collaboration and Participation in Scientific Reviews and Events 131
3.3.3 Supervision of Students 132
3.3.4 Entrepreneurship 133

3.3.5 Awards and Recognition 133
3.4 Final Remarks and Future Work Considerations 134
References.....137

LIST OF FIGURES

Figure 1 - Mental fatigue`s perceptions. 11

Figure 2 - Influences on task performance (James C Miller, 2012). 15

Figure 3 - Åhsberg's fatigue construct (Åhsberg, 1998). 21

LIST OF TABLES

Table 1 – Work program goals and phases. 44

Glossary

A

AI Artificial Intelligence

AmI Ambient Intelligence

ANAM The Automated Neuropsychological Assessment Metrics

API Application Programming Interface

APTS Automated Performance Test System

AVECLOS Average of Eye Closure

C

CAMCoF Context-aware Multimodal Communication Framework

CBR Case-Base Reasoning

CFS Chronic Fatigue Syndrome

CNS Central Nervous System

D

DAVE Divided Attention Visual Experiment

DMS Driver State Monitor

DIPT Discrete Information Processing Test

DPS Department of Public Safety

E

EDR Electrodermal Response

EDS Excessive Daytime Sleepiness

EEG Electroencephalography

EPC Evoked Potential Covariance

ESS Epworth Sleepiness Scale

F

FAID Fatigue Audit InterDyne

FAST Fatigue Avoidance Scheduling Tool

FMRI Functional Resonance Imaging

H

HCI Human–Computer Interaction

HOSVD Higher Order Singular Value Decomposition

HR Heart Rate

HRV Hear Rate Variability

K

KSS Karolinska Sleepiness Scale

KPCA Kernel Principal Component Analysis

L

LD Laplacian Derivation

M

MEQ Morningness-Eveningness Questionnaire

N

NASA-TLX NASA Task Load Index

O

OSA Occlusive Sleep Apnea

OSPAT Occupational Safety Performance Assessment Technology

P

PERCEOS Percentage of Eye Closure

PETER Performance Evaluation Tests for Environmental Research

PSQI Pittsburgh Sleep Quality Index

POMS Profile of Mood States

PVT Psychomotor Vigilance Task

S

SAFTE Sleep, Activity, Fatigue and Task Effectiveness

SOFI Swedish Occupational Fatigue Inventory

STAI State-Trait Anxiety Inventory

SSS Stanford Sleepiness Scale

SVM Support Vector Machine

SWAT Subjective Workload Assessment Technique

T

TOPS Truck Operator Proficiency System

U

USRT Unprepared Simple Reaction Time Test

V

VDT Visual Display Terminal

VTS Vienna Test System

W

WAM Wrist Activity Monitor

1 INTRODUCTION

Fatigue is considered a major problem, affecting individuals in their daily lives, especially at work, causing failures and consequent errors in the tasks being performed, leading to some kind of discomfort (Tucker, 2003; Williamson et al., 2005). It is common for people to stretch their physical and psychological boundaries to find enough time for work or play, family and fun. This extra time is usually obtained at the expense of sleep time, which is usually seen as unnecessary or less important subject, which, in turn, may lead to fatigue.

The effects of fatigue may not be immediate or even visible but may have serious consequences for individual health and safety at work and labor productivity (Folkard, Lombardi, & Tucker, 2005; Folkard & Tucker, 2003; J C Miller, 2006; James C. Miller, 2005). In more extreme cases, fatigue can even put people working on safety-sensitive jobs at risk, where any error on their part can lead to loss of life (Turner, 2013). The exercise of driving a vehicle is an example of the importance of fatigue in an activity or task in which small errors often have a significant influence on human lives (Mitler et al., 1988).

In general, fatigue was seen as a set of characteristics, such as a decrease in alertness, deficiency in decision making, a feeling of fatigue or decreased performance of certain types of tasks. There are many factors, internal and external, that modulate the onset and presence of fatigue. These include sleep deprivation, naps, noise, heat, mood, motivation, time of day or workload (Hockey, 1983). These fatigue effects can occur at any time and can persist from just a few hours to several consecutive days. Depending on its duration and intensity, fatigue can make it difficult or even impossible to perform daily tasks (Folkard & Tucker, 2003; Rosekind, 2005).

In the subsequent review, the concept of update and analysis from the point of view of an Artificial Intelligence (AI) system to detect and monitor its mental effects it is presented. It begins with an introduction to the concepts and domains of mental fatigue. Two of the three well-known classes of measures used to monitor mental fatigue, i.e., decrease in objective performance are subjective perceptions, are discussed. The physiological indicators are not reviewed here.

With regard to AI, the most widely used techniques and their purpose in detecting and monitoring mental fatigue were examined. The review highlights the advantages of various AI troubleshooting methods as well as some of its shortcomings and serves as a starting point for the development of a new system for the detection and monitoring of fatigue. The content of this review resulted from a GAP analysis (S. W. Brown & Swartz, 1989) on systems that detect and monitor fatigue. The search engines and databases used were the Web of Knowledge, Google Scholar, SCOPUS, PubMed, Science Direct, and Microsoft Academic Search, and examined using the following terms, viz.

- *Fatigue;*
- *Mental health;*
- *Artificial Intelligence;*
- *Machine Learning;*
- *Fatigue countermeasures;*
- *Fatigue measurement;*
- *Fatigue detection;*
- *Fatigue monitoring;*
- *Fatigue model development and validation;*
- *Fatigue risk management;*
- *Fatigue predictors;*
- *Physiological indicators;*
- *Performance measures;*
- *Shift work;*
- *Time on task;*
- *Workload;*
- *Sleep disorders; and*
- *Circadian Rhythm.*

The conduct of the review was motivated by the possibility of being able to detect and monitor fatigue in individuals through non-intrusive and non-invasive real-time measurements. To this end, a novel contribution is put forward in this document through the analysis of human operator interactions with the mouse and keyboard of a computer

(Pimenta, Carneiro, Novais, & Neves, 2013). The detailed approach falls within the domain of performance measurement for the detection of fatigue. For this reason, performance as a measure of fatigue is addressed in some detail in this work. Essentially, other approaches in this domain look at performance indicators and error indices to quantify the effects of mental fatigue.

1.1 Fatigue

1.1.1 What is Fatigue?

During our daily lives, we often experience a sense of tiredness during mental or physical work. This sensation is named fatigue. Generally, the term fatigue is used to describe a series of manifestations that range from drowsiness or lack of concentration to lack of physical strength or agility (Van Der Linden & Eling, 2006). Thus, the perception of fatigue is very broad and subjective. It can be seen as a combination of symptoms, including impaired performance (i.e., loss of attention, slowed reaction and response times, impaired decision making or poor performance on tasks that usually reflect good performance), and subjective feelings of sleepiness and tiredness (Perelli, 1980; Williamson et al., 2005).

To validate it, fatigue can be defined as a degree of failure of physical or mental factors associated with loss of physical or mental performance, hindering the natural or spontaneous achievement of a usual activity. Bartlett (Bartlett F, 1953) provides one of the usual definitions of fatigue in relation to day-to-day tasks, viz.

Fatigue is a term used to cover all those determinable changes in the expression of an activity which can be traced to the continuing exercise of that activity under its normal operational conditions, and which can be shown to lead, either immediately or after delay, to deterioration in the expression of that activity, or, more simply, to results within the activity that are not wanted.

On the other hand, Brown (I. D. Brown, 1964; May, 2011) stamped mental fatigue in the form, viz.

Fatigue may be conceptualized as the subjective experience of individuals who are obliged...to continue working beyond the point at which they are confident of performing their task efficiently… [Fatigue is] the subjectively experienced disinclination to continue performing the task at hand… [The] main effect of fatigue [is] a progressive withdrawal of attention from the task at hand. [This withdrawal] may be sufficiently insidious that [operators] are unaware of their impaired state and hence in no position to remedy.

Thus, routine fatigue can be viewed as a physical or mental impairment that can result in physical or mental discomfort as well as reduced performance. This insufficiency can be manifested by symptoms such as psychological disturbances, loss of energy or to increase the tension of the beam during the performance of simple tasks (Faber, Maurits, & Lorist, 2012; James C. Miller, 2008). In addition to the effects on mood or energy, fatigue is or may become the cause or partial reason for errors and accidents. Often this occurs when tired workers ignore the degree of their impaired mental state (James C. Miller, 2013). These errors and accidents take on special importance when we consider the domain of high risk jobs, such as driving large vehicles, military operations, firefighting and medical practice, just to name a few. In addition to these immediate problems, fatigue can also lead to long-term health problems. Gartner & Murphy (Gartner & Murphy, 1977) presented four main categories of immediate and undesirable effects, viz.

- *Motivation decrement* — include negative feelings and attitudes towards assigned tasks or equipment, diminished vigor or aggressiveness and disturbances, all occurring without observable decrement in performance or proficiency (Hartman, 1967);
- *Skill or proficiency decrement* — noticeable deterioration in virtuosity, precision, timing or coordination (Bartlett F, 1953) or reserve capacity (I. D. Brown, 1964), with no significant degradation of task performance;
- *Psychological stress* — detectable impairment of perceptual, cognitive or psychomotor function and/or disorganization of regulatory or adaptive mechanisms required for effective task performance (Benoit, 1980; Cohen, 1980; Welford, 1968); and

- *Performance decrement* — operationally significant blunders, erratic behavior, errors, delays or omissions which are clearly related to system performance (Hockey, 1983; McClernon & Miller, 2011; Roscoe, 1974).

Beyond these immediate effects, Hartman (Hartman, 1967) suggested three categories of fatigue, viz.

- *Acute fatigue* — which occurs normally between major sleep periods. This is the common fatigue that affects people on a day-to-day basis. It is often called tiredness (one night of good-quality sleep will may offset it);
- *Cumulative fatigue* — which accumulates over period of days or weeks owing to the insufficient recovery from successive periods of the acute one. Several nights of recovery sleep, beyond the normal sleep need, is required to recover from it; and
- *Chronic fatigue* — a psychoneurotic syndrome characterized by difficulty in committing oneself to an aggressive course of action, and by a generalized withdrawal or retreat from conflict which is intolerable for situation or personality reasons.

According to more recent definitions from Miller, who continued Hartman's line of thinking (James C. Miller, 2013), one may have, viz.

- *Acute fatigue builds up normally and unavoidably within in one waking period, such as a work day, and recovery from acute fatigue occurs as the result of one good-quality, nighttime sleep period. This is the normal pattern of fatigue that we experience during a 40-hour-per-week, nine-to-five, weekday-only job; and*
- *Cumulative fatigue builds up normally and unavoidably across major waking periods when there is inadequate recovery due to inadequate sleep between them. Recovery from cumulative fatigue cannot be accomplished in one good-quality, nighttime sleep period.*

Acute and *Cumulative* fatigue appears to have somewhat different biochemical effects within the brain. *Cumulative* fatigue is a widespread problem in shiftwork and also when people cannot get enough sleep or choose not to sleep in favor of social activities. The latter problem instigated the *National Sleep Foundation* to identify and discuss a *national sleep debt*

in the *United States* (J C Miller, 2013). Miller also redefined *chronic* mental fatigue, which is not related to the *chronic* fatigue syndrome (CFS), viz.

[M]ay be specified when an individual displays concurrently four or more of the following symptoms: the desire to sleep; apathy; substantial impairment in short-term memory or concentration; muscle pain; multi-joint pain without swelling or redness; headaches of a new type; pattern or severity; unrefreshing sleep; and post-exceptional malaise lasting more than 24 hours. The symptoms must have persisted or recurred for at least one month and should have a known source that includes lack of sleep, night work, long work hour, and/or irregular work hours (Miller JC. White Paper: Chronic Fatigue, 2014, posted at ResearchGate).

Acute and *Cumulative* fatigue interacts each day with circadian rhythm effects. For acute fatigue, the effect of the hours of the day that have passed since the end of the night of previous sleep is counteracted by the increase in body temperature. This compensation allows a reasonable level of mental performance over a normal work day. During sixteen hours of daytime wakefulness, the circadian peak in body temperature has passed and sleep homeostasis is requiring the onset of sleep. So, we slept at night and recovered from a day. However, there is a mid-afternoon increase in accidents and incidents that paralleled a mid-afternoon reduction in the daytime increase in body temperature. This mid-afternoon effect is due to the circasemidian rhythm, which also deepens the pre-auroral trough at body temperature (R. Broughton et al., 1988; Roger Broughton & Mullington, 1992; Folkard et al., 2005). The circasemidian rhythm has a cycle length of twelve hours and is suspected to be the first harmonic of the circadian rhythm (J C Miller, 2006).

For *Cumulative* fatigue, insufficient recovery between the waking periods generates at least three interactions with the circadian rhythm and its first harmonic. In the case of pre-dawn at body temperature, it is important to note that in the case of pre-auroral troughs, also for night workers, a circadian effect on daytime sleep reduces sleep periods and reduces the probability of daytime sleep provide sufficient recovery of fatigue (Admi, Tzischinsky, Epstein, Herer, & Lavie, 2008; Åkerstedt, 2003). Finally, for daytime workers who fail to properly recover at night, the circadian increase in body temperature day may decrease, leading to a much-impaired mental performance during the day.

In this classification based on three categories, normal acute mental fatigue is responsible for minimal daily performance decreases, mild subjective perceptions of fatigue or physiological responses. It generates mental fatigue that is widely recognized, accepted, and planned. With cumulative mental fatigue, deficits in cognitive performance are greater, but more subtle than acute ones.

The transition from cumulative mental fatigue to chronic disease is difficult to detect. Probably follows the persistence of cumulative fatigue for weeks or months. The symptoms of chronic fatigue may be similar to some symptoms of chronic fatigue syndrome. However, in the transition from cumulative fatigue, chronic symptoms are known.

When one speaks of fatigue in general terms, it is this perception that incorporates physical and mental aspects. Hartmann distinguished physical and mental fatigue (Benoit, 1980). Physical fatigue usually arises from "physical activity, sport, or mixed physical-intellectual activity without worry or anxiety" (Reilly & Edwards, 2007). Acute mental fatigue usually occurs after a day of "emotional stress or a difficult day, not totally enjoyable, intellectual work and more emotional intellectual work" (Åkerstedt, Folkard, & Portin, 2004; I. D. Brown, 1964). Hartmann also suggested that although physical fatigue does not produce well-defined mental effects, mental fatigue produces discomfort, irritability, anger and lack of energy, difficulty concentrating, loss of social adaptability, and loss of ability to standardize care or long-term planning.

The literature that ties together physical and mental fatigue is scarce. Michael Bonnet failed to detect a relationship between exercise and sleep (Bonnet, 1980). Generally, it is thought that the other is not affected, except, perhaps, by motivation, as Hartman suggested (Hartman, 1967). Thus, fatigue can be divided into two dimensions, i.e., mental and physical fatigue, linked by motivational factors.

Mental fatigue is a state that involves a range of effects on a set of cognitive, emotional, and motivational abilities and usually results in general discomfort as well as the emergence of performance limitations (van der Linden, Frese, & Meijman, 2003). Some of these limitations imply that a tired person is often less willing to engage in stress tasks or to perform the task in a conditioned or outdated way, well below their normal capacity (Boksem,

Meijman, & Lorist, 2006; Lorist et al., 2000). However, despite many studies on fatigue, it is extremely difficult to monitor mental fatigue (Muscio, 1921; Robert & Hockey, 1997).

Mental fatigue can be characterized by the perception of lack of mental energy. People who are affected by mental fatigue may feel that they have less energy than usual and are unusually tired and lethargic. Excessive mental workload can make a person feel mentally exhausted (Gartner & Murphy, 1977).

An operator profile provides valuable information regarding the potential level of fatigue. A profile can be seen as a predictor of a basic level of fatigue, according to the individual's lifestyle. These aspects have been carefully studied, mainly by psychologists, and include, viz.

- *Age* - Defines the mental age of the individual. It is important to understand the expected cognitive abilities of the individual, which may have a tendency to degrade with age;
- *Gender* - Mental states are different between men and women;
- *Professional occupation* - Many occupations are intrinsically more tiresome or exhausting than others; and
- *Consumption of alcohol and drugs* - The use of certain substances for short or prolonged periods may cause dependencies and other effects that lead to a state of mental fatigue.

Mental fatigue is also affected by a number of other external factors. They may or may not be directly related to the individual's behavior, and may embrace, viz.

- *Mood* - The mood of the individual may influence decisively his/her mental state, with a particular effect on his/her motivation to work. Although tired, the individual may overcome, even if only temporarily, the effects of fatigue with a positive mood and motivation;
- *Stress* - Stress may be defined as the demands placed upon the individual's mind or body by external stimuli, requiring the individual to acclimatize to the dynamic requirements of the environment. However, these processes of acclimatization require an additional effort from the brain which, when prolonged over long or intense periods, will result in mental fatigue; and

- *Sleepiness* - Sleepiness is often mistaken for mental fatigue or generalized as such. A difference exists and must be pointed out. However, the mistake is understandable since sleepiness is a symptom that is strongly connected to mental fatigue, i.e., it is one of the methods our brain uses to tell one that it is running out of resources. Sleepiness often results in a general loss of individual vitality.

1.1.2 Physical Fatigue

Our focus here is on acute muscle fatigue, especially on the hands, arms, shoulders and lower back in sedentary jobs. We also consider task-specific fatigue to answer questions such as "If a fatigue-like effect is detected by the user's interaction with the keyboard, how can it determine whether it was caused by tired typing muscles or a lapse of attention, a co-worker talking, and so on? "

Perhaps the first research project that focused on the concept of fatigue was a muscle study published by Mosso (Mosso, 1915). It reported an earlier study on muscle fatigue in 1884, in which the primary measure was "measurable change in production into a productive activity" as measured by an ergographer (Safety Net, 2009). The view of fatigue as a decrease in work or performance continues in the current times in studies of physical and mental fatigue. Muscle fatigue is defined as the inability to maintain the required or expected strength of the muscle (Edwards, Hill, Jones, & Merton, 1977). In a broader definition, muscle fatigue may be viewed as "an acute loss of performance that includes both an increase in the perception of the effort required to exert desired force and any failure to produce this force (Enoka & Stuart, 1992)."

Physical fatigue is usually associated with decreased muscle strength, speed or agility, thus limiting the performance of physical tasks (Williamson et al., 2005). Running, practicing sports or simply climbing stairs or carrying heavy grocery bags are examples of activities that can cause physical fatigue (Chaffin, 1973; Sumner & Xu, 2002). Clinically, a resistance or resistance test can be used as a diagnostic tool to try to find the causes of local physical fatigue (Chaffin, 1973).

Numerous studies have evaluated fatigue in physical activities (Cote, Feldman, Mathieu, & Levin, 2008). For example, Hammarskjold and Harms-Ringdahl investigated the influence of fatigue on muscle activities, such as repetitive movements during nails, saws and rotating screws (Hammarskjöld & Harms-Ringdahl, 1992). They showed that although some characteristics of the task, such as difficulty of perception and speed of movement, remain constant, the average EMG amplitude usually produced signs of fatigue.

1.2 Detecting Mental Fatigue

With mental fatigue caused by inadequate sleep, circadian effects or excessive mental workload, humans may exhibit reduced processing power or may not respond in any way to stimuli and information received. In contrast, when the mental workload is much lower than the appropriate level, they become bored and tend to make mistakes. Cognitive performance is usually measured as the combination of speed, success or accuracy of an individual when performing a task. One can also detect and monitor fatigue through subjective classifications of mental workload. Finally, physiological measures reveal subtle aspects of mental fatigue that were not revealed by performance measures and subjective classifications.

These measures may and should be used as complementary practices whenever possible. It should be noted that performance and physiological measures are objective, as opposed to subjective ones (Figure 1).

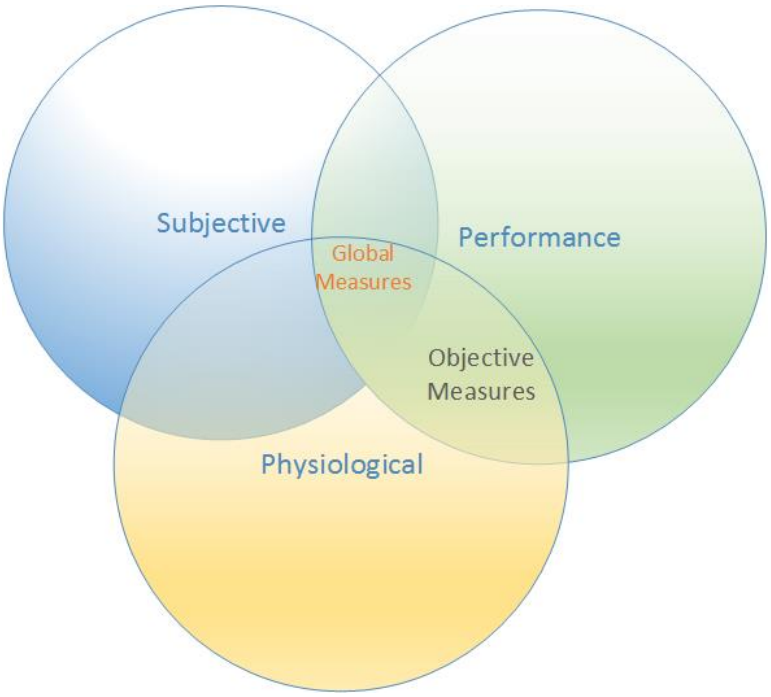


Figure 1 - Mental fatigue`s perceptions.

1.2.1 Performance Measures

Human achievement is a factor that is considered critical in any activity or work. With regard to performance measurement, we examined several human acquisition measures that are useful for detecting fatigue states, as well as detecting the possibility of failure to acquire. An individual connects performance directly to task execution. The performance of the task can be defined as the relative competence of the individual interpretation of the stimuli associated with the task and the effectiveness of its responses linked to it. Excessive or too low mental and physical workload observed in monotonous tasks are usually the situations in which performance decreases. Perhaps the most commonly used indicator of fatigue is an observation of the degradation of task performance over time or in response to changes in task demands.

Welford (Welford, 1968) distinguished four types of performance changes that may occur, viz.

- *Impairment of sensory or perceptual functions;*
- *Slowing of sensory-motor performance;*

- *Irregularity of time; and*
- *The disruption of performance.*

The decrease in performance due to fatigue can be considered as total quantity and / or quality of work performed as a function of time spent on a task. Cognitive performance is usually measured as the combination of an individual's response time accuracy when performing a task. This combination is often referred to as throughput. The response speed is inverse of the response time and is generally distributed in terms of response time in data collected from tired operators. Both speed and accuracy are generally acceptable when the individual has normal cognitive abilities, and either or both are diminished when those abilities are impaired or lessened by fatigue. With mental fatigue, human performance is not consistent. Changes in speed, accuracy, and variability allows one to measure an individual's performance on a task.

The main criticism of the use of performance indicators for fatigue reduction is based on inaccuracy in the fall accounting. The effects of fatigue may not be disclosed when individuals compensate by exerting more effort. Usually, a substantial or excessive time is required in the task to induce task-specific fatigue or acute fatigue. However, a very high workload can cause this same effect in less time. In both cases, there are demands for excessive tasks during the time required to perform a task. Excessive workload conditions can be considered as accelerating the build-up of fatigue, while normal workload results in a slower fatigue emergency.

However, the decrease in performance is often not observed, although performance requirements or workloads have increased. This phenomenon can be explained in terms of extra capacity, a concept originally developed for the evaluation and modeling of humans as drivers and pilots (Casali & Wierwille, 1983; W. W. Wierwille & Gutmann, 1978). Replacement capacity theory proposes that people rarely operate at their maximum capacity and therefore may temporarily absorb additional cognitive or physical workload and maintain their standard performance as the available capacity is not exceeded. Thus, increased fatigue is often seen as a decrease in people's resilience, and models of workload and fatigue are considered interconnected. Since mental fatigue manifests itself primarily as an impediment to cognitive abilities, it is natural that its performance can be assessed through cognitive processes, viz.

- *Visual and Auditory Perception* – One the one hand, visual perception stands for one`s ability to interpret the surrounding environment by processing information that is contained in visible light. On the other hand, auditory perception denotes one`s aptitude to decode the surrounding environment by processing information that is confined at perceptible sound pressure waves;
- *Response Time* - This is the elapsed time between the occurrence of a stimulus and the response of our body or mind. Responses may vary from a thought or a change in the mental state to a physical movement or an alteration in our physiology. Accuracy and precision may be seen as the achievement of results that are within the expected quality and timeframe for a given individual given his cognitive and physical skills. They also represent the absence of unexpected errors;
- *Memory* - The use of our memory includes all the processes related to encoding, storage and retrieval of information in our brain; and
- *Attention* - This may be seen as a cognitive process in which the intellect focuses and selects stimuli, establishing a relationship between them. Concentration is the cognitive process of selectively focusing on one aspect of the environment while ignoring others. The loss of concentration may be caused by external factors and may sometimes even be desirable. For example, a sudden and potentially hazardous event that must be analyzed before continuing with a given task. However, we frequently lose interest or focus in our tasks, at a rate that increases with the increase of the mental fatigue.

Performance variability, delayed attention and impaired vigilance are three classic features of the effects of fatigue on mental performance. One source of response variability is mental lapse. Bills pioneered the systematic study of mental fatigue in terms of brief attention failures - the famous Bills Block (A. G. Bills, 1937; Arthur G. Bills, 1931). Subsequently, Wilkinson and Houghton developed a simple, portable, 10-minute "excitation" and "continuous and concentrated attention" test, called a tape-based Unprepared Simple Reaction Test (USRT) based on a cassette player (Wilkinson & Houghton, 1982). Dinges then introduced a solid-state version of the USRT called Psychomotor Surveillance Task (PST), which is now widely used in sleep loss studies (Dinges & Powell, 1985; Lim & Dinges, 2008).

Currently, the delay in response to irregularly repeated simple stimuli is usually defined as taking more than 500 milliseconds to respond or not respond for two seconds.

The USRT / PST evaluates the delay in response to discrete stimuli. There are parallels in the field of continuous control, as exercised by a pilot guiding an aircraft or a driver who drives a vehicle on the highway. The study of this phenomenon is embedded in the theory of manual control (D.T. McRuer & Hofmann, 1968; D.T. McRuer & Jex, 1967; Duane T. McRuer, Weir, Jex, Magdaleno, & Wade Allen, 1975). In relation to response time variability and manual control, Morris and Miller conducted a study that analyzed the variability of pilots' performance during simulated flight as a means of detecting and measuring fatigue (T. L. Morris & Miller, 1996). In that work it was verified that the decrease in flight performance consistency occurred due to changes in fatigue, this performance impairment was also reflected in physiological measures. The performance variability approach was pioneered by Mackie and Miller (James C Miller & Mackie, 1978) and was recently expanded by McClernon (McClernon & Miller, 2011). Response variability is a powerful tool in the arsenal of fatigue detection.

Surveillance may be defined as the ability to remain alert for a rare but important signal embedded in a background of frequent and similar events. The surveillance research dates back to World War II studies of radar operators by Norman Mackworth, the Mackworth Clock Test, and Mackworth's description of the classic surveillance decrement. This history is available in Chapter 2 of the Operator's Fatigue Manual (James C. Miller, 2012). Until 1980, little of this research was directly applicable to real-world operations (James C Miller, Mackie, & Standards, 1980). However, since then, several relevant studies have been published and reviewed (Warm, Parasuraman, & Matthews, 2008). Historically, we attributed a decrease in vigilance to annoyance and related theoretical constructs (Mackie, 1977). However, Warm and his colleagues, based on their review, concluded that "The convergence of evidence using behavioral, neural and subjective measures shows that surveillance requires difficult and stressful mental work" (Warm et al., 2008).

Regarding attention, Smith and Miles conducted a study that examined the effects of lunch on cognitive surveillance tasks (Smith & Miles, 1986). In this study, the authors looked at the performance of individuals who did not eat lunch before performing tasks compared to those

who did. The study also showed some differences between morning and afternoon performance in individuals who abstained from lunch. An additional Tucker industrial data assessment "examined the effect of rest of temporal trends on the risk of industrial accidents, by evaluating accident records of a large engineering firm, obtained in 3 years" (Tucker, 2003). The relative risk of an accident in the last half hour of a two-hour shift was 95% greater than in the first half hour and did not differ in three two-hour work periods within a work day.

Task difficulty and perceived workload also affect task performance and details the interactions of system and task characteristics, perception, sleep, and circadian rhythm in task performance (Figure 2). The black boxes describe the characteristics of the task and the red boxes describe the characteristics of the human operator. The dotted line indicates a positive correlation, measured with the Pearson product-moment correlation coefficient, between the actual and measurable workload of the task and the worker's perception of that workload, estimated with a mental workload tool as NASA-TLX or SWAT (James C. Miller, 2012).

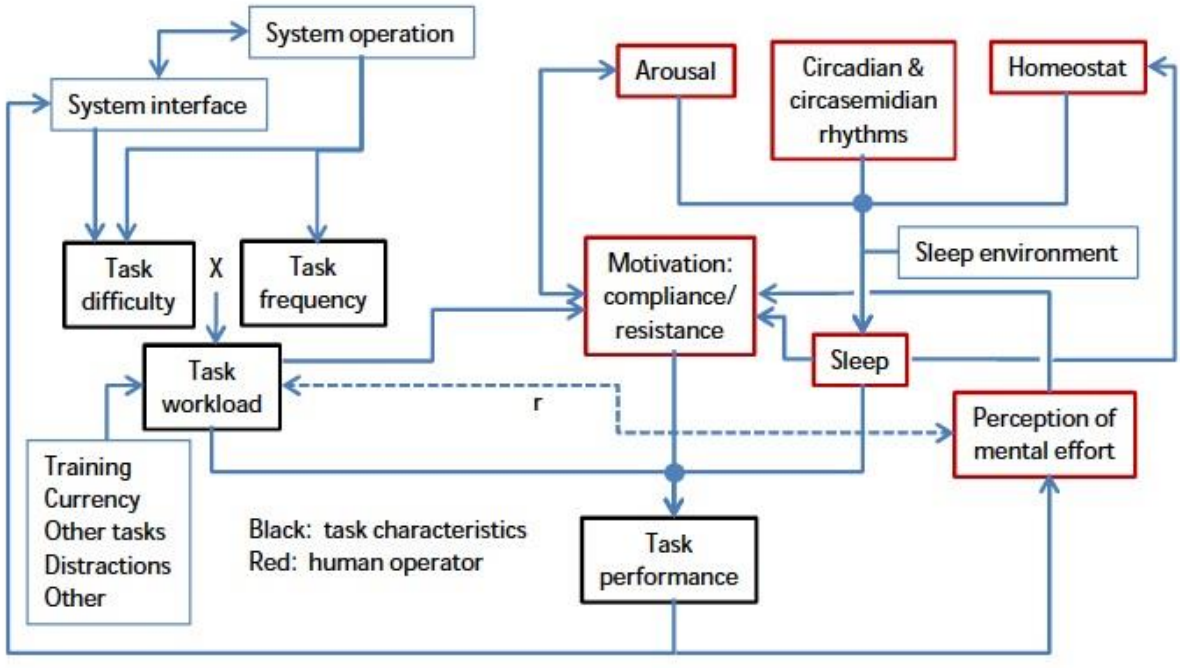


Figure 2 - Influences on task performance (James C Miller, 2012).

Performance may also be affected by various environmental conditions, such as lighting, temperature and noise. Performance challenges can occur when these conditions become uncomfortable, distracting, or simply not appropriate to the task at hand. A decrease in high noise levels is an obvious example, but levels that are too low can also become uncomfortable. Ambient lighting is an interesting case because, on the one hand, light can activate the Central Nervous System (CNS) and can improve visual quality, but on the other hand, high intensity can cause problems of reflection and visual acuity.

Circadian rhythm effects on mental performance

Humans exhibit oscillations in their physiology known as biological rhythms. These rhythms are classified according to their frequency in circadian (one cycle per day), ultradian (more than one cycle per day), and infradian (less than one cycle per day. Circadian rhythms are found in almost all human functions, including body temperature, secretion of almost all hormones, cardiac, pulmonary and metabolic activity, nervous activity and sleep-wake cycle variations. Circadian rhythms modify our perceptions of alertness and sleepiness. During the day, alertness is usually perceived to be high and sleepiness usually perceived to be low, while the opposite occurs during the nighttime. Circadian variations have been found in human performance, including motor reaction time, time estimation, verbal memory, mental arithmetic, and simulated driving tasks. Performance is usually higher during the day and lower at night, especially in the midnight to dawn period. Performance also depends upon a homeostatic sleep process. Quantitative models exist that predict the interactive effects of circadian rhythmic and sleep homeostatic processes upon alertness and task performance (Åkerstedt et al., 2004; Belyavin & Spencer, 2004; Fletcher & Dawson, 2001; Steven R. Hursh, Redmond, et al., 2004).

The different feelings of drowsiness and fatigue that are a consequence of circadian rhythms affect the cognitive abilities of the human being and result in more frequent and more serious errors that can compromise the decision making, the learning and the solution of problems. These variations in human performance may be the result of circadian rhythms in cognitive processes that are crucial to the performance of all tasks. The rhythms were found

in three basic neuropsychological processes, namely attention, working memory and executive functions (i.e., initiative, planning, action, continuous activity monitoring and behavior adjustment). These processes underlie the performance of cognitive tasks (Valdez, García, Ramírez, & García, 2012).

In research labs and some field settings, mental performance is measured by having subjects perform a set of cognitive tasks, such as reaction time, short and long-term memory tasks, reading comprehension, mental arithmetic, time, logical reasoning, coding and code resolution, and so on (Reeves, Winter, Bleiberg, & Kane, 2007). The performance of these tasks by humans depends on the basic cognitive processes that constitute the building blocks of higher brain functions and human behavior (Orchard & Stern, 1991). Cognitive processes such as attention, memory and executive functions vary throughout the day and night as functions of sleep deprivation and fatigue. This variation may be due to effects on a tonic alert state (i.e., wakefulness, general alert) which, in turn, affects other basic cognitive processes, resulting in errors.

When collecting and then analyzing performance data, one must deal with interindividual variability. Some of these variations may be due to the subjects' chronotypes, i.e., they are definitive types of morning ("larks"), definitive nocturnal types ("owls") or something intermediate. This classification is best performed with the Morningness-Eveningness Questionnaire (MEQ) (Terman, Rifkin, & White, 2001). The MEQ is a self-assessment questionnaire that was validated against variations in the circadian rhythm of oral temperature. Its main purpose is to determine whether a subject's maximum alertness is in the morning, evening or in the middle. The MEQ consists of 19 multiple-choice questions, with each question with four answer choices.

The two categories of defined Morningness and defined Eveningness represent approximately 15% to 20% of the human population and the medium range category applies to the majority (60% to 70%) of humans. Morning types can be quite sensitive to delays in nocturnal sleep and the duration of sleep during morning sleep may be short. Morning types can also report low satisfaction with night work and may choose to leave the tour and work through the night. Night types may tend more toward nighttime and shift work, and therefore

may be the people who frequently develop the kinds of health problems usually associated with night and shift work.

1.2.2 Subjective Ratings of Mental Workload, Mental Fatigue, and Sleepiness

Fatigue effects can be detected and monitored through subjective classifications of mental fatigue, mental workload, and drowsiness. Validated instruments are available to measure these subjective states. Such instruments will be useful to validate the measure of proposed fatigue performance. However, the measurements require the manual input of data. We discuss here only the measures that we hope will be most useful for our purposes.

Subjective measures of mental workload are often used to assess the mental workload associated with a task (Hart & Staveland, 1988; Reid, Eggemeier, & Shingledecker, 1982). Obtaining mental workload levels during task performance can be a difficult procedure. However, estimates of subjective workload levels can help to isolate task characteristics that affect performance. The two instruments frequently used in the research were developed in parallel in the 1980s, one at the NASA-Ames Research Center in California and another at the US Air Force's Human Factors Research Group at Wright-Patterson AFB, Ohio.

The NASA Task Load Index (NASA-TLX) is a multidimensional assessment tool (Hart & Staveland, 1988). The main seven-point scale is given below, viz.

Overall Performance: How successful were you in performing the task? How satisfied were you with your performance?

The TLX has five seven-point subscales that help identify difficult task characteristics. The subscales are denoted beneath, viz.

- *Mental Demand,* i.e., how much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?
- *Physical Demand,* i.e., how much physical activity was required? Was the task easy or demanding, slack or strenuous?
- *Temporal Demand,* i.e., how much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?

- *Frustration Level,* i.e., how irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?; and
- *Effort,* i.e., how hard did you have to work (mentally and physically) to accomplish your level of performance?

The Subjective Workload Assessment Technique (SWAT) uses a three-dimensional workload model, namely time load, mental stress load, and psychological stress load. There are three levels of classification for each dimension, i.e., low, medium and high. Typically, subjects perform a pre-type of the possible combinations of dimensions (3 x 3 x 3 = 27) and their type is transformed into a workload scale ranging from 0 to 100. For the actual task, each activity or event is evaluated by assigning an integer value in the interval 1 ... 3 in each of the three dimensions. The value of the preset scale associated with this combination becomes the value of the workload for that activity or event.

A seven-point mental workload scale was created by the Crew Performance Branch of the USAF School of Aerospace Medicine in the late 1970s, and then re-examined, linearized, and verified by the Human Factors Engineering Branch of the Air Force Flight Test Center (Ames, George, Air, & Base, 1993). This is the easiest rating instrument to use in high workload environments, but it provides less insight than the TLX or SWAT instruments. Individuals choose one of seven sets of statements describing their average mental workload during the preceding work period, viz.

- Nothing to do, i.e., no system demands;
- Light activity, i.e., minimum demands;
- Moderate activity, i.e., easily managed; considerable spare time;
- Busy, i.e., challenging but manageable; adequate time available;
- Very busy, i.e., demanding to manage; barely time enough;
- Extremely busy, i.e., very difficult; non-essential tasks postponed; and
- Overloaded, i.e., system unmanageable; essential tasks undone; unsafe.

Mental fatigue can be estimated with the Mood Status Profile (MSP) Vigor-Activity and Fatigue-Inertia factors. MSP measures five aspects of affect or mood (McNair, D M, Lorr, M, Droppelman, 1971; Shahid, Wilkinson, Marcu, & Shapiro, 2011). It consists of 65 adjectives

that describe the feeling and mood to which the subject responds according to a five-point scale ranging from "Nothing" to "Extremely". The results are reported as six mood factors, viz.

- *Tension-Anxiety,* i.e., eightened musculoskeletal tension including reports of somatic tension and observable psychomotor manifestation;
- *Depression-Defection,* i.e., depression accompanied by a sense of personal inadequacy;
- *Anger-Hostility,* i.e., anger and antipathy toward others;
- *Vigor-Activity,* i.e., vigorousness, ebullience, and high energy;
- *Fatigue-Inertia,* i.e., weariness, inertia and low energy level; and
- *Confusion-Bewilderment,* i.e., bewilderment, muddleheadedness; appears to be an organized-disorganized dimension of emotion.

Because of its length, MSP is practical only for occasional use, such as establishing a baseline and estimating the effects of excessive sleep deprivation or restriction. Åhsberg created the Swedish Occupational Fatigue Inventory (SOFI) (E. Åhsberg, Gamberale, & Gustafsson, 2000; Elisabeth Åhsberg, 1998). In this instrument, the following twenty symptoms are graded on a seven-point scale. From these twenty descriptors, a value is calculated for each of the five subscales, viz.

- *Lack of energy,* i.e., general feelings of diminishing strength;
- *Physical exertion,* i.e., whole-body sensations that may be the result of dynamic work and to a certain extent a sign of metabolic exhaustion;
- *Physical discomfort,* i.e., more local bodily sensations that may result from static or isometric work load;
- *Lack of motivation*, i.e., feelings of not being involved or enthusiastic; and
- *Sleepiness, i.e.,* feelings of sleepiness.

This is a good instrument to use when it is necessary to differentiate physical fatigue, mental fatigue and drowsiness. In fact, the fatigue construction of Åhsberg was as shown in Figure 3.

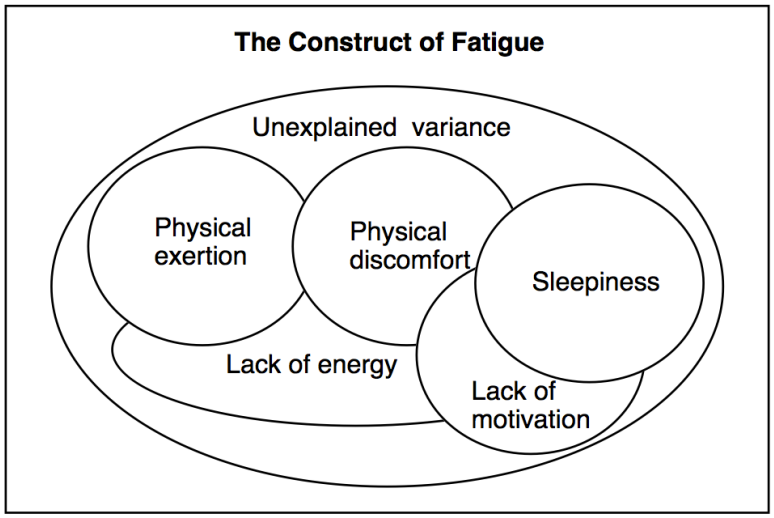


Figure 3 - Åhsberg's fatigue construct (Åhsberg, 1998).

The seven-point USAFSAM Mental Fatigue Scale was created in about 1979 by Dr. William F. Storm and Captain (Dr.) Layne P. Perelli of the Crew Performance Branch of the USAF School of Aerospace Medicine, Brooks AFB, San Antonio, Texas, and then used in many field and laboratory tests (James C. Miller, 2013), viz.

- Fully alert, i.e., wide awake; extremely peppy;
- Very lively, i.e., responsive, but not at peak;
- Okay, i.e., somewhat fresh;
- A little tired, i.e., less than fresh;
- Moderately tired, i.e., let down;
- Extremely tired, i.e., very difficult to concentrate; and
- Completely exhausted, i.e., unable to function effectively; ready to drop.

The level of an individual's fatigue becomes worrisome when their ratings reach and exceed five. One use of the scale was in the design of a proposed crew fatigue prediction instrument (Samn & Perelli, 1982). The fatigue scale was referred to in the research literature as the "Samn-Perelli" scale, but this is a misnomer. In fact, it is the "USAFSAM Fatigue Scale", created by Storm and Perelli (Woodrow, 2014).

Excessive Daytime Sleepiness (EDS) can be assessed clinically with the Epworth Sleepiness Scale (ESS). ESS was developed at Epworth Hospital in Melbourne (Johns, 1991, 1992). The

ESS correlated well with the measured ElectroencEphaloGraphically (EEG) sleep latencies measured at night or during the day, and is considered a validated and reliable self-reported measure of sleepiness. Subjects use an integer in the range 0 ... 3 corresponding to the probability (i.e., never, light, moderate, and tall, respectively) that they would sleep in eight situations, such as sitting and reading, watching TV, car for an hour, and so on. The eight numbers are added together. A total score of 15 or more of a possible 24 is a cause for concern regarding acceptable work performance, while a score of five or less indicates that there is no EDS. Plots around 10 (ten) are a gray area. This is an excellent instrument for use in subject selection before an experiment.

Similarly, the Pittsburgh Sleep Quality Index (PSQI) can be used to select the pre-experimental subject. The PSQI provides estimates of sleep disturbances and habitual sleep habits during the previous month. The PSQI consists of 19 items, scaled from 0 to 3. These items generate estimates in seven domains clinically derived from sleep difficulties, i.e., sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances. sleep, use of sleeping pills, and daytime dysfunction.

The Stanford Sleepiness Scale (SSS) is a valid, brief, seven-item scale that can be used repeatedly during the day and at night (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). To use the SSS, the subject selects one of seven sets of descriptors, viz.

- Feeling active and vital, i.e., alert; wide awake;
- Functioning at a high level, but not at peak, i.e., able to concentrate;
- Relaxed, i.e., awake; not at full alertness; responsive;
- A little foggy, i.e., not at peak; let down;
- Fogginess, i.e., beginning to lose interest in remaining awake; slowed down;
- Sleepiness, i.e., prefer to be lying down; fighting sleep; woozy; and
- Almost in reverie, i.e., sleep onset soon; lost struggle to remain awake.

Similarly, the Karolinska Sleepiness Scale (KSS) is a validated, easy to use instrument (Akerstedt & Gillberg, 1990; Gillberg, Kecklund, & Akerstedt, 1994). It consists of a nine-point scale, with three semantic anchors, viz.

1. Extremely alert;
2. Very alert;

3. Alert;
4. Rather alert;
5. Neither alert nor sleepy;
6. Some signs of sleepiness;
7. Sleepy, but no difficulty remaining awake;
8. Sleepy, some effort to keep alert; and
9. Extremely sleepy, fighting sleep.

For more information, about subjective scales associated with fatigue and sleepiness, see the book, *STOP, THAT and One Hundred Other Sleep Scales* (Shahid, Wilkinson, Marcu, & Shapiro, 2012).

1.2.3 Physiological measures

Physiological measurements may be important in validating our approach to measuring performance. ElectroEncephaloGraphy (EEG) is a technique that has been used frequently for the detection of mental fatigue patterns and also for the analysis of cognitive overload. Alan Gevins performed EEG studies to analyze workload and fatigue (A. Gevins et al., 1995; A. Gevins & Smith, 1999). Oculography and blinking measures, such as PERCLOS, were used for the detection of fatigue (Goldstein, Bauer, & Stern, 1992; Orchard & Stern, 1991; Stern, Walrath, & Goldstein, 1984). The Wrist Activity Monitor (WAM) was used to track sleep and sleep debts. The best validated WAM is the ActiGraph, which uses the Cole-Kripke algorithm (Jean-Louis et al., 1997; Tang & Harvey, 2004). Other physiological measures such as Heart Rate (HR), Heart Rate variation (HRV), and ElectroDFermal Response (EDR) were found to be unreliable for the detection of mental fatigue (Segerstrom & Nes, 2007; Stork, 2012). The main problem with physiological approaches was that, despite their accuracy, their continued use in work environments is somewhat impractical, perhaps with the exception of controlling ocular and ocular movements remotely.

A relatively new discipline, Neuroergonomics, has emerged in the realm of physiological measurement (Wickens C. D., 2008). Its objectives are parallel to the goal of the performance-based approach that is the focus of our review. Neuroergonomics harnesses modern tools

that provide information about brain function. These may include EEG, magnetoencephalography and functional Magnetic Resonance Imaging (fMRI). The course uses the knowledge gained from these types of technologies to better understand the relationships between brain functions and task performance in real-world operations. With this knowledge, the expectation is that we can design systems for safer and more efficient operations.

Parasuraman and Wilson reviewed neuroergonomics studies in workload and vigilance, adaptive automation, neuroengineering, molecular genetics, and individual differences (Parasuraman & Wilson, 2008). They concluded that the neuroergonomics approach can enrich our understanding of the use of technology by humans and can inform technological design. They suggested that applications should include the assessment of operator workload and vigilance, the implementation of real-time adaptive automation, neuroengineering for people with disabilities, and the design of selection and training methods.

There is one caveat about physiological measurement in the workplace that must be stated. In the United States, some individuals and some workers' unions object to physiological measures in the workplace. In some cases, they label them as "medical" measures, though none have any clinically diagnostic value. Nonetheless, there is common aversion to these measures in the workplace on the basis of fears of invasion of privacy.

1.2.4 Systems for Detecting and Monitoring Mental Fatigue

There are several approaches that can be taken to test for fatigue or deficiency in the workplace, namely performance, neurological measurement and biochemistry (James C Miller, 1996). These approaches are likely to be complementary, i.e., biochemical and neurological measures assess lifestyle, which is appropriate for some occupations, and performance measures assess immediate and work-relevant problems. The tests can be categorized as follows, viz.

- Biochemical
    I. Direct, e.g., brain (not used); and
    II. Indirect, e.g., blood, expired air, urine.

- Behavioral (Neurological)
    i. Subjective, e.g., DOT Standardized Field Sobriety Test (SFST); and
    ii. Objective, e.g., EEG evoked response, pupillometry.
- Performance (Motor)
    iii. Abstract, e.g., finger tapping
    iv. Industrial, e.g., simulated driving

b. Perceptual
    i. Abstract, e.g., pattern comparison
    ii. Industrial

c. Higher Cognitive
    i. Abstract, e.g., code substitution
    ii. Industrial

d. Mixed
    i. Abstract
    ii. Industrial

Performance tests may be sorted into several sub-categories, i.e., visual-motor, perceptual, and higher cognitive. These labels each refer to the highest level of "thinking" (cognition) required to perform a given test. For example, a visual-motor test might simply require that the hand be connected to the eye by normally functioning neural circuits. One reliable motor test measures finger-tapping speed and coordination.

A perceptual test might require that one accurately compare patterns, identify colors, or recognize tones as high, medium or low pitch. Examples of higher-cognitive tests include, viz.

- Simple mathematical processing, such as adding three-digit sums together with reasonable speed and accuracy;
- b) Code substitution, where numbers (digits) are substituted rapidly and accurately for letters (symbols) in a simple code-solving process; and
- Short term memory tasks where for example a set of letters (e.g., four) is briefly presented (e.g., one second), followed by a series of letters presented one at a time. The subject's task is to determine if the letters in both sets match.

The validity of facial performance tests can be abstract, i.e., not similar to tasks in the workplace. Alternatively, performance tasks can be directed directly from the workplace or industry. Performance tests may also require or require no qualified performance. For example, the finger test engine test requires only a demonstration and little practice before an individual can be tested. On the other hand, the performance of qualified compensatory tracking may require 30 minutes to two hours of practice.

Within the performance category (above), it appears that fatigue or environmental stressors may impair the major cognitive functions in first place, perception and visual-motor functions will last. Thus, measures of higher cognitive functions may provide the most sensitive indexes for assessing fitness for duty. In other words, there will be a relatively large proportion of compromise detections among those who are truly harmed. On the other hand, visual motor function measurements can provide the most specific indexes for assessing fitness for duty, i.e., there will be a relatively large proportion of no detections of impairment among those who are not actually harmed.

Performance tests may use an absolute or relative approval / failure criterion or a mixture of both. An absolute criterion uses a distribution of data collected from a representative sample of the population of interest. The performance of the individual in a given test is compared to the expectations derived from the distribution of the score of this sample. A qualified criterion is based solely on the distribution of punctuation provided by the individual. In the mixed approach, a moving window of an employee's recent performance is used to calculate that worker's baseline. The approval level is defined as a standard position within the individual's performance score distribution, as two standard deviations from the mean. Thus, all employees have an absolute passing criterion (standard deviation units) based on a relative comparison, i.e., their individual performance should fall within a specific portion of their own distribution of recent scores.

For a statistically binding sample, the sensitivity of a screening test increases with the number of samples due to the decrease in the sampling variance of the estimate. The effect size is the average of an independent variable on the measure of interest, for example, the average effect of a 0.05 blood alcohol content on tracking performance. One aspect of sensitivity is the quotient of effect size and variability. Sensitivity is enhanced as the useful

data set is increased in size and is diminished as the useful data set is reduced in volume when the effect size is held constant. Thus, a desirable test provides sensitivity within a short test period, but is not so short that it is insensitive. Another way to increase sensitivity is to measure several performance functions at once, and then look for any one of them to fall outside their respective standard limits.

One way to increase test specificity is to use test-retest strategies. For example, Wade Allen and Henry Jex at Systems Technology, Inc., described the "1-of-n" strategy. In this strategy, the test is usually passed in a first trial. Test failure requires failing every one of several trials in a test session. This strategy emphasizes the correct identification of un-impaired workers. This may be called "exclusion testing." A failure is reason to exclude the employee, at least temporarily, from safety-sensitive work. However, other necessary cognitive or motor functions have not been tested. Thus, we may not conclude from a test pass that the job will be performed properly.

Three subsections follow. They focus on cognitive and visual-motor test batteries, on performance tests, and on driver fatigue detection systems. Some of the information is historical, but provides background for present-day systems.

1.2.4.1 Cognitive and Visual-Motor Test Batteries

The former three batteries (Delta, ANAM, CogScreen) were influenced by research conducted within the U.S. Department of Defense and led by Dr. Fred Hegge (Englund, Reeves, Shingledecker, Thorne, & Wilson, 1987a, 1987b; Perez, Masline, Ramsey, & Urban, 1987).

Automated Performance Test System (APTS), which, in turn, was derived from the Navy's Performance Evaluation Tests for Environmental Research (PETER) battery (Bittner, Carter, Krause, Kennedy, & Harbeson, 1983). The battery assessed higher cognitive functions, including associative memory, linguistic information integration and manipulation, spatial information integration and manipulation, perceptual input, and output and response execution. All of the task implementations had been studied carefully and extensively (Turnage, Kennedy, Smith, Baltzley, & Lane, 1992a, 1992b). All of the tasks were learned

quickly and performance was stable across individuals within tests. Data remained stored in computer files and was processed *post hoc* with computer spreadsheets and statistics packages to make decisions about fatigue or impairment.

The *Automated Neuropsychological Assessment Metrics* (ANAM) resides now at the University of Oklahoma. This test battery has been used widely since its inception in the 1980s (Reeves et al., 2007). The battery includes 22 individual tasks that assess attention, concentration, reaction time, memory, processing speed, decision-making, and executive function.

CogScreen is a cognitive test battery that was designed by Gary Kay and Richard Horst (Banbury, Dudfield, Hoermann, & Soll, 2007) primarily for re-certifying pilots. It is also used extensively in clinical neuropsychological contexts. The functions assessed include attention, spatial perception, reasoning, and response time. It allows a varied battery to be selected, and it compares results to norms. It has been used several times in fatigue research by the U.S. Federal Aviation (Serber et al., 2010).

Additionally, the *Vienna Test System* (VTS; Schuhfried GmbH, Mödling, Austria) comes in different forms. It allows testing at of visual-motor and higher cognitive functions. The latter include attention, working memory, long-term memory, executive functions, and spatial processing. The TRAFFIC version (originally, the ART-90) of the VTS assesses fitness for driver licensing. It was developed, built and programmed between 1979 and 1982 (Kuratorium für Verkehrssicherheit, Viena). The system is "officially authorized for assessment of fitness to drive in the statutory context" in 16 countries in Europe, South America, and South Africa.

Gilliland and Schlegel provided a review of 14 computer-based performance tests and batteries, many created by government agencies (Gilliland & Schlegel, 1993).

1.2.4.2 Performance Tests

The *FACTOR 1000* (Performance Factors, Inc.) stands for the commercial version of the Critical Tracking Task developed by Jex, McRuer and colleagues (Jex, McDonnel, & Phatak, 1966; Jex, McDonnell, & Phatak, 1966; James C Miller, 1996). It generated an estimate of the minimum delay time from the hand to the eye for a visual-motor continuous tracking task (as

opposed to a discrete reaction time task). The FACTOR 1000 used mixed criteria based upon an individuals' changing distribution of scores from day to day, four samples per trial to support sensitivity, and the 1-of-n strategy across two trials to support specificity.

NovaScan assessed divided attention capabilities, requiring the user to switch attention among concurrently operating tasks (James C Miller, 1996). The tasks might involve, for example, spatial visualization and tracking, and replaced each other on the screen on a random basis. This strategy allowed a measure of the individual's ability to switch from one task to another. Measures of attention were obtained with the introduction of a third task, i.e., an indicator appeared in the corner of the screen for this third task. The net effect of this paradigm was to split the overall attention system into a series of separate sub-skills, including monitoring, attention switching, and interference with monitoring from each of the main tasks. Pass/fail criteria were determined using traditional psychometric approaches.

The *Personal Safety Analyzer* (CAE-Link) was a matrix of self-administered, computerized tests that used personal baselines. The matrix of tests assessed an employee's performance in key representative work tasks, i.e., the speed with which an employee extracts information from stimulus displays, how quickly and accurately may use that information in decision-making tasks, and their response patterns. This framework was designed to be representative of the man-machine interfaces that would be encountered in the workplace. The tests encompassed the acquisition of information (encoding), working memory, long-term memory, decision-making, and response selection and execution.

The *Divided Attention Visual Experiment* (DAVE; Atlantis Aerospace Corp.) was developed to assess levels of driving impairment in individuals affected by Occlusive Sleep Apnea (OSA) syndrome. DAVE tested focuses in measuring their ability to perform a sub-critical tracking task, while responding to a 4-choice, simple reaction time task.

The *Occupational Safety Performance Assessment Technology* (OSPAT) is a visual-motor tracking task similar to the FACTOR 1000 (Hartley, Horberry, Mabbott, & Krueger, 2000). The test uses mixed criteria based upon an individuals' changing distribution of scores from day to day. No relevant validation data is available.

CogSpeed (Gray Matter Metrics, San Antonio, Texas) began as a Discrete Information Processing Test (DIPT) (Perelli, 1980). The subjects beyond the threshold of information

processing speed is determined through an adaptive five-choice reaction time task. This task is a discrete-response analog of the Critical Tracking Task, referred to above. Instead of providing an estimate of the fastest eye-hand connection speed in a continuous tracking task, it provides an estimate of the highest speed with which an individual may respond to a discrete stimulus. The maximum speed is determined through repeated, adaptive trials, i.e., "the rate of stimulus presentation was a function of the latency of the prior response time and error rate" (Perelli, 1980).

1.2.4.3 Performance-Based Driver Fatigue Detection Systems

The *Truck Operator Proficiency System* (TOPS; Systems Technology, Inc., Hawthorne, CA) was developed with funding from the State of Arizona (Stein, Parseghian, & Allen, 1990). It was a fitness for duty testing device that tested commercial vehicle drivers' hand-eye coordination and ability to divide attention. The driver sat behind a steering wheel in a simulated truck cab and steered an imaginary vehicle down a road displayed on a computer screen. This was an early version of the STISIM driving simulator used worldwide, in research laboratories. A boring eight-minute test was administered, and results were available immediately. The Arizona Department of Public Safety (DPS) used the device to test commercial truck drivers at weigh stations. The DPS put drivers out of service for failing the test and believed that the courts will uphold this regulatory use of the technology.

The TOPS device was enhanced by Evaluation Systems, Inc. as ReadyShift, to include a non-decreasing personal baseline for a five-minute test (James C Miller, 1996). This enhancement enabled the skill developed by a driver who would take the test day after day, month after month. ReadyShift I was a desk device, while the ReadyShift II was installed in the long-haul truck taxis, to be used when the truck was parked. The test used mixed criteria based on the variable distribution of individuals' scores on a day-to-day basis, parallel measurements of various visual-motor and attention-by-test functions to support sensitivity and 1-of-n strategy in two trials to support the specificity. The test took five or ten minutes, depending on the result of the first five minutes. The results were available immediately and, for Readyshift II, transmitted via satellite to supervisors.

The *ARRB Pro-Active System* includes a real-time monitor that measures operator alertness levels during a shift for haul truck drivers on mining sites (Mabbott, 2003). ARRB also incorporates user fatigue training and user fatigue management profiles that contain their performance data from the device as well as data from take-home sleep research equipment. The profiles can be used for performance monitoring and the initiation of countermeasure discussions with individual employees where appropriate. The monitoring device is an in-cab, two-choice serial reaction time task. The test uses mixed criteria based upon an individuals' changing distribution of scores from day to day.

A two-choice serial reaction time task has been used extensively and successfully in fatigue research using the STISIM driving simulator (Systems technology, Inc., Hawthorne, California). A four-choice serial reaction time task was proposed as an improvement to the Train Sentry III alert for locomotive engineers (J C Miller, Wagner, & Bezos, 1993). Generally, it appears that a serial reaction time task is a good fatigue-monitoring device in some highway and rail transportation systems.

The *Advisory System for Tired Drivers* (ASTiD™; Fatigue Management International, UK) was designed by Professor Jim Horne and Dr. Louise Reyner of the Loughborough Sleep Research Centre (Koh, Jones, Spencer, & Thomas, 2007). It is an unobtrusive, predictive system for driver sleepiness. The ASTiD uses similar techniques to predict and detect driver fatigue. The prediction component provides hour-by-hour estimates of the likelihood of the driver falling asleep and is also influenced by prior sleep. This component uses a fatigue model such as those described below. The driver inputs a sleep judgment score of good, poor, or bad before driving, and this score modulates the fatigue model. A microcomputer uses research data to calculate the effects of task-specific (driving) fatigue and rest breaks. Finally, the system uses a direct measurement of vehicle lateral acceleration to calculate an analog of lateral lane position variability (weaving) (McClernon & Miller, 2011; James C Miller & Mackie, 1978).

For more detailed information on driver fatigue detection systems see the review by Barr and colleagues produced for the U.S. Department of Transportation (Barr, Howarth, Popkin, & Carroll, 2009).

1.2.4.4 Quantitative Prediction of Fatigue Effects

In addition to performance testing, one may use one of at least a half-dozen quantitative models to predict mental performance. These models are based upon the two-process model defined by Borbély (Åkerstedt & Folkard, 1997; Borbély, 1982; Borbely & Achermann, 1999; Kronauer, Czeisler, Pilato, Moore-Ede, & Weitzman, 1982). The two models used most widely in industry are FAID and SAFTE.

The *Fatigue Audit InterDyne* (FAID®) generates mental performance predictions based on work hours, rest periods and circadian cycles (Dawson & Fletcher, 2001; Fletcher & Dawson, 2001; Roach, Fletcher, & Dawson, 2004). This is achieved by introducing the start and end times for a shift system into the program, including shift start and end times for the previous seven days, to measure recent work history and recovery. The output consists of histograms relative to fatigue scores for each hour of the shift schedule, with supporting summary tables.

The *Sleep, Activity, Fatigue, and Task Effectiveness* (SAFTE™) simulation integrates quantitative information about circadian rhythms in metabolic rate, cognitive performance recovery rates associated with sleep, cognitive performance decay rates associated with wakefulness, and cognitive performance effects associated with sleep inertia to produce a 3-process model of human cognitive effectiveness (S R Hursh et al., 2004; Steven R. Hursh, Redmond, et al., 2004). The SAFTE model is implemented in the Fatigue Avoidance Scheduling Tool (*FAST*™) software (D R Eddy & Hursh, 2006a, 2006b; Douglas R Eddy & Hursh, 2001; Steven R. Hursh, Balkin, Miller, & Eddy, 2004a, 2004b). The output consists of continuous relative fatigue scores as a function of time, with supporting summary tables.

1.2.4.5 Quasi-Physiological Fatigue Detection Systems

There are some quasi-physiological fatigue monitors that are of interest here, since they might be useful in a static workplace.

The *DD850 Driver Fatigue Monitor* (DFM; Attention Technology Inc.) is a video-based drowsiness detection system for measuring slow eyelid closure. The DFM estimates PERCLOS (Percent Eyelid Closure), which is the proportion of time the eyes are closed 80% or more

during a specified interval (R. Grace et al., 1998; Richard Grace & Steward, 2001; W. Wierwille, 1999). In turn, PERCLOS was derived from the seminal eye blink work of the late Dr. John Stern at Washington University, St. Louis (Stern et al., 1984).

The *Driver State Monitor* (DSM) was reported to be under development by Delphi Automotive (Edenborough et al., 2005). However, it was not described on their web pages when this review was being written. Instead of PERCLOS, this video-based eyelid monitor predicted AVECLOS, the percentage of time the eyes are fully closed during a one-minute period. This less complex measure permitted the use of an automotive-grade data processor.

The *faceLAB*™ system (Seeing Machines, Canberra, Australia) uses video to track the head, face tracking, eye gaze, and eyelid. The system includes PERCLOS assessment. This system has been used extensively for research in driving simulators.

There are many fatigue detection systems that use video to track the eyelid and calculate PERCLOS. Again, for more detailed information on many of these systems see the review by Barr and colleagues (Barr et al., 2009).

1.2.5 Summary

On the one hand, in one`s assessment to detect mental fatigue, it was prominent the usefulness of using a three-pronged method, i.e., measuring task performance, physiological activities, and quantified subjective perceptions. On the other hand, given a somewhat personalized view on what Artificial Intelligence (AI) may contribute to aid performance, our focus was mainly on the measurement of task enactment. Indeed, it were de-emphasized the measurement of physiological activities, although they may be used to validate the effectiveness of AI aiding, viz.

- Having in attention these basic cognitive functions to underline the mental functioning, i.e., visual perception, visual-motor response time, memory, and attention. In terms of human behavior, it was also noted that performance variability, attention lapsing, and impaired vigilance are three classic characteristics of the effects of fatigue on mental work;

- All in all, the effects of circadian rhythms on mental performance, especially chronotype. Several validated instruments have been observed that can be used to quantify subjective perceptions of mental workload, fatigue and somnolence, namely the NASA-TLX, SWAT, POMS, USAFSAM Mental Fatigue Scale, ESS, PSQI, SSS, and KSS. Some physiological measurement domains were also object of attention, including EEG, EOG, WAM, HR, HRV, and EDR;
- Categorized systems that may be used to detect and monitor mental fatigue, and focused on performance measurement;
- Mentioned aspects of test sensitivity and specificity;
- Described cognitive and visual-motor test batteries, namely Delta, ANAM, CogScreen, and VTS;
- Described performance tests, specifically FACTOR 1000, NovaScan, Personal Safety Analyzer, DAVE, OSPAT, and CogSpeed; and
- It were discussed four driver fatigue detection systems that were performance-based, viz.
    - TOPS;
    - ReadyShift;
    - Pro-Active; and
    - ASTiD.

Also described were two predictive, quantitative fatigue-modeling tools, explicitly FAID, and SAFTE. Finally, we look at three quasi-physiological fatigue detection systems that may be instructive for one`s future work, i.e., DD850, DSM, and faceLAB.

1.3 Artificial Intelligence in Fatigue Detection and Monitoring

Artificial Intelligence (AI) has been defined as the science and engineering of making intelligent machines and as human-like intelligence, either in machines or software. AI technologies are used widely today to optimize processes, to make products easier to use, and to automate tasks. Within AI research, we look for new methods of solving problems. This search generates challenges in terms of knowledge representation and reasoning

methods, planning, learning, natural language processing, motion, manipulation, perception, social and evolutionary intelligence, feelings, and creativity. These techniques can be applied and used in areas such as medicine, psychology, weather, finance, transportation, gaming, aviation, and The Law. In the particular case of fatigue detection and monitoring, only recently sophisticated computing systems have appeared. Many of these advances have been aimed at vehicle drivers.

1.3.1 How Artificial Intelligence Can Improve Fatigue Detection and Monitoring

Computers are present in almost all areas of life, ranging from health to management and sports. Usually, computers are used as working tools, and users are required to perform tasks such as typing text in support of management, billing, programming, and communication. Computers are largely considered as working tools, and even as a mandatory requirement for intellectual work. The day-to-day life of an individual in today's society is increasingly demanding and complex, both at work and on a personal level, and in these circumstances computers are seen as excellent support. AI can enhance the use of the computer in favor of the human being. The detection and monitoring of fatigue are a good example of this potential.

As shown in Section 1.2, the detection and monitoring of fatigue have been topics discussed over many years. Some of the techniques summarized above involve the use of computational power to detect and monitor fatigue. A few of them even have recourse to AI techniques, yet most of them with some computational power are used mainly for the monitoring of vehicle drivers. In this section are examined some of the branches of AI that may be used to improve the detection and monitoring of fatigue processes.

1.3.2 The Intersection of AI with Fatigue Detection and Monitoring

Here it will be addressed some of techniques and tools that have been used for the detection and monitoring of fatigue that are inspired by and based on AI based techniques for problem solving.

1.3.2.1 Computer Vision

The computer vision field is one of the most widely used for the detection and monitoring of fatigue. This is due to the fact that a non-intrusive technique that can be used for driving scenarios that are regarded as critical tasks, as well as the areas in the field of fatigue. The growing number of traffic accidents due to the decreased level of driver`s monitoring has become a serious problem for society. In Europe, the statistics showed that 10% to 20% of all traffic accidents are due to drivers with a decreased level of vigilance caused by fatigue. In the trucking industry, about 60% of fatal truck accidents are related to the same factors (Bekiaris, Nikolaou, Panou, Van Wees, & Baten, 2002; Yannis et al., 2017).

Bergasa (Bergasa & Nuevo, 2005) presented a non-intrusive system based on computer vision to monitor the fatigue of drivers. This prototype was based upon six behavioral characteristics, namely PERCLOS, eye closure duration, blink frequency, nodding frequency, face position, and fixed gaze. Recall that a analogous measurement of several functions per trial is a good way to support test sensitivity.

The level of fatigue and driver inattention was estimated, using these parameters, under a fuzzy classifier system. Also, Cyganek (Cyganek & Gruszczyński, 2014) proposed a hybrid system for visual monitoring of the states of driver fatigue. They used a Higher Order Singular Value Decomposition (HOSVD) basis to assess sleepiness and attention, based upon tracking of the driver's eyes through planar functions.

Hwang (Kuo-An Hwang & Chia-Hao Yang, 2009) proposed a system to detect mental fatigue and restlessness in e-Learning. Such a platform used image processing to support facial recognition algorithms.

1.3.2.2 EEG

An ElectroEncephaloGram is a neurophysiological evaluation of bioelectric brain`s activity, by collecting data through electrodes placed on the scalp or even by subdural electrodes in the cerebral cortex (Montgomery, Montgomery, & Guisado, 1995). The EEG is considered the most reliable physiological assessment in the prediction and detection of fatigue (Shen, Ong,

Li, Hui, & Wilder-Smith, 2007). There has been a growing interest in tracking mental fatigue based on EEG with the widespread hope that technology will be invaluable for the prevention of accidents arising from such factor.

Alan Gevins reported topographic patterns of event-related covariance between electrodes for subjects performing a difficult memory and fine-motor control task for 10 to14 hours (A. S. Gevins et al., 1990). It was even created and published an automated drowsiness detector based on features of the ongoing EEG. This detector classified EEG segments from four bipolar channels by using decision heuristics based on increased ratios of both delta-band to alpha-band and theta-band to alpha-band spectral intensity as compared to thresholds automatically determined for each subject from a waking calibration period.

In a 1990 study Alan Gevins used Evoked Potential Covariance (EPC) analysis, based upon the idea that the neural processes underlying higher cognitive functions must involve the coordination of activity in widely distributed brain regions. It was implemented a Laplacian Derivation (LD), the second derivative in space of the potential field at each electrode. Gevins improved the accuracy of his estimates of the LD by using 3-D spline functions to compute the LD over the actual shape of the head, being examined 27 channels of scalp-recorded EEG and extracted 18 channels of Laplacian Derivation evoked potentials. It was detected and reduced EEG artifacts by applying tuned-threshold detectors and adaptive filtering and neural-network pattern recognition. EPC patterns between all pair-wise combinations of the 18 channels were computed. Striking changes occurred in the patterns after subjects performed the task for an average of 7to 9 hours, but before performance deteriorated.

Kai-Quan Shen (Shen et al., 2007) accomplished feature selection with the random forest method for multilevel mental fatigue EEG classification. The random forest is an ensemble of many randomized decision-trees. Its output is the plurality vote of all of the decision-trees (Tin Kam Ho, 1995). Also Zhang Chong (Zhang, Zheng, & Yu, 2008) investigated states of mental fatigue caused by long periods of sustained, low-workload operations at a Visual Display Terminal (VDT). They used the Kernel Principal Component Analysis (KPCA) and Support Vector Machine (SVM) to differentiate between two states. KPCA is a technique to extract features used for dimensionality reduction. The dimensionality reduction decreases the computation time and removes irrelevant structures from the collected data. These

preprocessing steps can also help in reducing the noise, making the data easier to work with as they create uncorrelated features of it (Yang, Frangi, Yang, Zhang, & Jin, 2005). SVM is a computer algorithm that learns by example to assign labels to objects. For example, an SVM can learn to recognize fatigue behavior by recording the fatigued and not fatigued people. In general, given a set of concrete examples, each marked as belonging to one of two categories, a SVM training algorithm builds a model that assigns new examples into one category to one another (Hearst, Dumais, Osman, Platt, & Scholkopf, 1998).

Research has not been limited to the *post hoc* analysis of EEG records. Real-time detection has been also object of study. Morris (A. Morris & Ulieru, 2012) proposed an architectural design for adaptive-systems agents that use brain state information to make more effective decisions on behalf of an user. This system offers fatigue detection through a multi-agent system. A multi-agent system is a computerized system that consists of multiple entities (software or hardware) that make intelligent decisions and interact within an environment. Multi-agent systems can be used to solve problems that are difficult or impossible to solve with a single entity or a monolithic system (Wooldridge, 2009). The EEG analysis and classification of the system is performed through a hybrid fuzzy neural network system.

1.3.2.3 Voice

Shiomi (Shiomi, Itano, & Suzuki, 2010) developed a method to measure the degree of activity of the brain's neocortex through the human voice. Their analytic method was based upon Chaos Theory Zhihai (Li, Li, Yao, & Li, 2011) analyzed speech signals and time domain signal of the voice through cluster analysis. Greeley (Greeley et al., 2007) used probabilistic models to detect fatigue using unsupervised learning methods. Unsupervised learning means that the system can learn to categorize new input patterns without any external feedback. It can still be seen as a process that tries to find the hidden structure in unlabeled data. Once the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution (Seret, Maldonado, Weber, & Baesens, 2013). Caraty (Caraty & Montacié, 2014) detected fatigue during extended oral readings. It were called professional speakers such as teachers and telemarketing operators, being conducted three experiments, i.e., a prosodic analysis, which may be compared with the results generated by SVM. It was possible

to classify fatigue and non-fatigue states across a large set of audio features. The classification was based upon differences between two segments of speech using a combination of multiple comparison functions executed on phonemes. Prosodic analysis stands for a language analysis based on patterns of stress and intonation in different contexts (Kohler, 2006).

1.3.2.4 Performance

Calfee (Calfee & Rowe, 2004) created a multi-agent system for the simulation of human behavior in the U.S Naval Air Force. This multi-agent system was used to create training scenarios taking into account factors such as user skills, fatigue, stress, and other aspects that may influence the performance of an individual. Sharit (Sharit, 2003) addressed the reduction of cognitive workload in computer use through Case-Based Reasoning (CBR). CBR is a problem-solving paradigm that is part of a set of AI approaches. CBR is able to utilize the specific knowledge of previously experienced situations of practical problems (cases), instead of relying only on general knowledge of a problem domain, or making associations along generalized relationships between problem descriptors and conclusions. Such paradigm seeks to solve a new problem looking a similar past cases, and reusing it to solve a new problem (Palmer, Linde, & Pons, 2004). McClernon and Miller and Morris and Miller investigated the performance of pilots during simulated flight as a means of detecting and measuring fatigue (McClernon & Miller, 2011; T. L. Morris & Miller, 1996). In both cases they used statistical variability as an indicator of fatigue.

1.3.3 Summary

The different AI techniques that have been used for detecting fatigue have focused mainly on four measurement domains, namely computer vision, EEG, voice, and performance.

Many different AI techniques have been used in each of the four domains discussed above. They have been aimed mainly at classifying the subject as fatigued or not fatigued. One must use the classification techniques best suited to each type of data. Though in most cases AI is

not used, it is clear that the use of AI in the detection and monitoring of fatigue is already a fact.

1.4 Discussion

1.4.1 The Current State

In the previous section it were presented different AI-aided approaches to the detection and monitoring of mental fatigue. However, some of the approaches are intrusive. For example, the EEG requires the placement of scalp sensors that usually lead to limitation on the individual being monitored. These approaches, especially with AI aiding, are highly effective but very limiting if one wishes to implement non-intrusive data acquisition. To a lesser degree, the same caveat applies to the use of subjective measures, as these require intrusive data collection. Detection using computer vision is quite excellent and is used successfully in driving scenarios, but may be considered somewhat invasive.

Despite the effectiveness of most of these approaches and the use of sophisticated AI methods, the user has not been seen as the central entity in most cases, except for driving. In other situations, detection and monitoring by objective indicators has not taken into account the context issue, such as the type of task to be performed during monitoring, or even the workload associated with the task being performed. Some exceptions might include the walk-up workstations of the OSPAT and the in-cab mounting of the ReadyShift II. However, this kind of context information is extremely important when one intends to design a non-intrusive, non-invasive monitoring system.

Another problem is that subjective instruments do not fully take into account inter-individual differences. There are instruments that help account for individual differences, such as the Profile of Mood States (POMS) and the State-Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, Lushene, Vagg, & Jacob, 1983; X-, 1968). However, their use in complex systems can prove to be complicated and confusing, due to the same problems that can be observed in subjective measures of fatigue detection (Curran, Andrykowski, & Studts, 1995; Millar, Jelicic, Bonke, & Asbury, 1995; Newcombe & Boyle, 1995). Subjective

measurements also have the problem of being made through questionnaires that are filled by the user himself, interfering with the primary task.

Thus, existing non-driver solutions do not perform monitoring in a manner that adapts to the user and the working contexts. Setting these features is an essential aspect to take into consideration if one is to design a real-time, continuous monitoring system that can make a difference in the workplace.

1.4.2 Improving Non-Intrusive and Non-Invasive Fatigue Detection

Computers are increasingly present in our lives. They are used daily in one`s workplace, homes, for entertainment or to get informed. As they gradually take over the space of other objects such as television, radio or books, they start to become an irreplaceable part of our living. In this transformation, intuitive and natural interaction mechanisms between humans and computers are fundamental. This is the main goal of Human-Computer interaction (HCI), i.e., to improve the interaction between users and computers by making computer interaction mechanisms adapt to the users and the tasks accomplished. In a more long-term perspective, HCI aims to design systems that minimize the barrier between the human's cognitive model of what they want to accomplish and the computer's understanding of the user's task, as well as help in solving real world problems (Ahonen & Cowley, 2016; R. Picard & Wexelblat, 2002).

An especially interesting topic in HCI is affective computing, the branch of AI that aims to design computer systems that are capable of recognizing, interpreting and processing human emotions and behaviors. Under this approach, interaction can be driven by (among other issues) the emotional state of the user, resulting in computer systems that are sensitive to stress, tension, satisfaction or fatigue, being able to deliver better and contextualized services.

Affective computing essentially aims to make human-computer interaction closer to human-to-human interaction (R. W. Picard, 2003). In human-to-human interactions factors such as emotions, body language or signs of stress are taken into consideration. Often, these factors influence the decisions and the course of action of a person. The inclusion of these issues in computer models will allow us to develop systems that care for the user's problems,

that understand the motivation of the users and that communicate in a more appropriate manner. While computers will reveal difficulties to learn, to care or feel for real (R. Picard & Klein, 2002), the modeling and simulation of some of these traits could be advantageous.

1.5 Investigation Hypothesis

From an abstract point of view, this research work is framed in the field of Artificial Intelligence, and in particular, in the area of Ambient Intelligence (AmI), which is a relatively new paradigm, made possible to the advent of new key know-hows, namely ubiquitous computing, ubiquitous communication and intelligent user interfaces. In a few words, an AmI system can be defined as "A digital environment that proactively, but sensibly, supports people in their daily lives". This definition, albeit simple, unveils a thread of several interrelated fields, technologies and approaches. Besides the three technologies previously mentioned, that supported the birth of AmI, there are many other fields that are implicated, the most important being Intelligent Systems. Indeed, AI is such an umbrella that AmI uses many of its methodologies, tools and attitudes. The main lines of research that have been pursued in the last years, despite being naturally different, share some core ideas that bind them, viz.

- The practice of non-invasive and non-intrusive approaches borrowed from AI to problem solving; and
- The training of the appropriate models with data collected from behavioral studies.

Thus, the research hypothesis developed in order to conduct this research project states that it is possible to assess mental fatigue through non-intrusive and non-invasive means, from the observation of changes in behavioral patterns, as well as act on fatigue in a proactive way.

The analysis of users' behavior is an exciting and innovative topic in the field of Intelligent Systems, which is being researched with increasing interest in the community, i.e., knowing how a user behaves is essential to provide personalized services. The interest lies in studying the changes that occur in the individual in certain circumstances, with different levels of fatigue or in rested situations, particularly in the way they move, walk or interact with technological devices such as the mouse, keyboard or smartphone.

1.6 Objectives

This work, entitled “Detecting and Monitoring Mental Fatigue – c” has as main objective to develop and implement a system for detection and monitoring of mental fatigue using non-invasive and non-intrusive techniques. To achieve this goal, the system must be able to adapt to the user, identify his/her behavior, learn from them and being able to classify their comportments. It is mandatory vital to create a learning mechanism based on user behavior analysis.

The system should contain an application running in the background on each user device (e.g., computer, smartphone or tablet), in order to monitor the user through a non-intrusive and non-invasive way. This information should be stored not only locally but also on a central server, which will allow one to access such data through any device with internet access.

Another important component of this system involves the development of an API that will cater for the detection of mental fatigue on other platforms.

To develop this system is essential to create a learning algorithm capable to detect and monitor fatigue in different individuals with different profiles.

It is therefore possible to look at a monitoring system able to monitor mental fatigue of different users in different situations and contexts through a non-intrusive and invasive way, making it really useful and novel.

Summarizing, one`s work aims at to set the basis and the implementation of an intelligent system capable to detect and monitor, as well as to act on fatigue mental assessment, viz.

- Capture information through non-invasive and non-intrusive techniques;
- Make real-time monitoring;
- Adapt to the individual behavior and ground context;
- Predict mental fatigue behaviors; and
- Develop an API for detection of mental fatigue.

1.7 Research Methodology

To accomplish these objectives, the Action-Research methodology for problem solving will be followed (O'Brien, 1998). This methodology starts by identifying the problem so that a hypothesis can be formulated on which the problem analysis and completion will be based. Subsequently, the information is recompiled, organized and analyzed in order to build a proposal to address the computational setting, viz.

- Problem specification;
- Endless and incremental update and review of the state of art;
- Ongoing and shared development of the proposed model;
- Experimentation and implementation of the solution through the development of a prototype;
- Results analysis and conclusions making; and
- Unceasing diffusion of the results so far obtained and practices into the scientific community.

1.8 Work Programme

The work entitled "Detecting and Monitoring Mental Fatigue. A Non-invasive Approach" was carried out at the University of Minho, under the supervision of Professors Paulo Jorge Novais and José Maia Neves. The two co-supervisors have been working closely together for the last years in fields such as Intelligent Systems, Knowledge Representation and Reasoning, Behavioural Analysis and Multi-agent Systems, with application in domains such as Medicine and The Law. The work has been sub-divided into six phases that can eventually overlap.

Table 1 – Work program goals and phases.

Phase 1: State of the Art **Duration: 8 months**	Conduct a review of scientific literature in the main fields touched by this work in order to determine the main technologies and methods used in fatigue detection domain. Make an analysis of the advantages and disadvantages of current tools and

	methods to fatigue detection. At the end of this stage two Technical Reports will be produced detailing the conclusions so far obtained, one focusing on the legal domain and the other one focusing on the scientific one.
Phase 2: Architecture`s Definition **Duration: 6 months**	Study and choose the most suitable technologies to implement the monitoring system that respects the characteristics defined in the Objectives section. Several approaches will be studied and their advantages and disadvantages analyzed. At the end of this stage a Technical Report describing the conclusions so far attained will be produced.
Phase 3: Analysis of Requirements **Duration: 4 months**	Determine the requirements for the implementation of a working prototype of the monitoring system. In order to achieve this goal, human experts in the field and ergonomists will be consulted and their opinions and advices taken into account.
Phase 4: Implementation of the Architecture **Duration: 10 months**	Implementation of the architecture in the form of a working prototype, having as input the results obtained in the former two stages. The architecture should implement one or more of the approaches studied in stage 2, so that their performance and aptitude can be determined. At the end of this stage at least one Technical Report will be produced detailing the gotten conclusions.
Phase 5: Test and Assessment of the Platform **Duration: 8 months**	The platform implemented in the previous phase will be tested with volunteer's in scenarios were employees who interact directly with devices. The

	performance of the platform will be tested and eventual adjustments will be conducted as needed.
Phase 6: Diffusion and Analysis of Results **Duration: 8 months**	The results of this work will be promoted not only at the academic level but also at the commercial one through the publication of several articles and contacts with the industry, mainly with those that already have experience in the field of Human Factors and Ergonomics.

1.9 Structure of the Document

This section is intended to be a guide to help readers to find the right content and also understand some aspects related to the structure of the document.

This thesis is presented under the publication compilation scheme and therefore some requirements must be met to make it understandable and coherent. The first content to be provided is a summary of the doctoral thesis presented in both English and Portuguese. This is followed by the required List of Content, List of Figures, List of Tables, and Acronyms.

The thesis can be divided into three main chapters, namely Introduction and State of Art (Chapter 1), Publications that make up the Doctoral Thesis (Chapter 2) and Conclusions (Chapter 3)

2 PUBLICATIONS COMPOSING THE DOCTORAL THESIS

2.1 Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns

Title	**Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns**
Authors	A Pimenta, D Carneiro, P Novais, J Neves
Publication Type	Conference Proceedings
Publication Name	International Conference on Hybrid Artificial Intelligence Systems
Publisher	Springer, Berlin, Heidelberg
Volume	8073
Number	
Pages	222-231
Year	2013
Month	September
Online ISSN	978-3-642-40845-8
Print ISSN	978-3-642-40845-8
Url	http://dx.doi.org/10.1007/978-3-642-40846-5_23
State	Published
Scimago journal ranking	SJR IF 0.348 (Q2 - COMPUTER SCIENCE)
Indexed	Scopus, ISI Web of Science, DBLP

Monitoring Mental Fatigue through the Analysis of Keyboard and Mouse Interaction Patterns

André Pimenta, Davide Carneiro, Paulo Novais, and José Neves

CCTC/DI - Universidade do Minho
Braga, Portugal
pg20189@alunos.uminho.pt,
{dcarneiro,pjon,jneves}di.uminho.pt

Abstract. In our living, we often have a sense of being tired due to a mental or physical work, plus a feeling of performance degradation even in the accomplishment of simple tasks. However, these mental states are often not consciously felt or are ignored, an attitude that may result in human failures, errors and even in the occurrence of health problems or on a decrease in the quality of life. States of fatigue may be detected with a close monitoring of some indicators, such as productivity, performance or even the health states. In this work it is proposed a model and a prototype to detect and monitor fatigue based on some of these items. We focus specifically on mental fatigue, a key factor in an individual's performance. With this approach we aim to develop leisure and work context-aware environments that may improve the quality of life and the individual performance of any human being.

Keywords: Mental Fatigue, Monitoring, Fatigue Detection, Behavioral Biometric.

1 Introduction

Fatigue is regarded as one of the main causes of human error. Many times its symptoms are ignored, as well as its importance for a good mental and physical condition, elementary for human performance and health [1,2]. The activity of driving a vehicle is a good example of the importance of fatigue in our activities or tasks, in which small errors can often and easily have a significant impact on human lives. This is even more significant in high-risk jobs such as aviation, transportation, aerospace, military or medicine, in which individuals routinely operate complex systems with a high degree of responsibility [2,3]. The study of this topic is thus decisive as the continuous disregard of the effects of fatigue, seen frequently as normal consequences of our busy lifestyle, may end and up affecting one's quality of life, health or even lost of life.

Fatigue is however a very subjective concept and difficult to define from a scientific point of view. It may be seen as a combination of symptoms that include loss in performance (e.g. attention loss, slow reaction to a particular event, or

J.-S. Pan et al. (Eds.): HAIS 2013, LNAI 8073, pp. 222–231, 2013.
© Springer-Verlag Berlin Heidelberg 2013

low performance in tasks to which the individual has the necessary skills), and subjective feelings of drowsiness and tiredness. From an abstract point of view, fatigue may be seen as two-dimensional, namely mental and physical. Despite the frequent inter-dependence of these two dimensions, they may be addressed independently [2].

In this work it is detailed a monitoring system for mental fatigue. It intends to detect different mental fatigue states of an individual in a non-invasive and transparent way. The system will analyze a set of features that stem from the individual's use of the computer, namely from the mouse and the keyboard. The system goes through a prior learning phase in which the behavior of an individual while using the computer without fatigue is studied. Later, it classifies its level of mental fatigue by quantifying changes in the individual's behavior. This approach will open the door to the development of better and intelligent working environments, that may be aware of their users' mental states and take actions towards the improvement of quality of life and their performance on tasks in which they are engaged.

2 Mental Fatigue Detection

Fatigue is a non-specific symptom. It may be estimated or detected from multiple sources, including the profile of the individual (e.g. age, gender, professional occupation, consumption of alcohol or drugs), performance and precision indicators (e.g. mouse click/movement precision or tasks delivered), or attention span (e.g. time spent on a particular task versus the time spent in other non-related tasks) [4].

The user's profile provides valuable information with respect to the potential level of fatigue [5]. It can be seen as a predicted base level of fatigue in the sense that it establishes a baseline, according to the lifestyle of the individual. These aspects have been thoroughly studied, mostly by psychologists, and encompass:

- **Age** - Defines de mental age of the individual. It is crucial to understand the expected cognitive abilities of the individual, which may have a tendency to degrade over time.
- **Gender** - The mental states are different between men and women.
- **Professional occupation** - Important to understand possible causes of mental fatigue, or to foil false states, since many occupations are intrinsically more tiresome or exhausting than others.
- **Consumption of alcohol and drugs** - The use of certain substances for short or prolonged periods of time may cause dependencies and other effects that lead to a state of mental fatigue.

Mental performance is generally seen as the combination of the speed and success/accuracy of an individual when carrying out a task. It is improved when the individual has plain use of his cognitive skills and decreased when these skills are somehow impaired or diminished [6]. Cognitive skills may include:

- **Memory** - The use of our memory includes all the processes related to encoding, storage and retrieval of information in our brain. Memory loss and loss of performance in accessing memory is clearly linked to mental fatigue, besides the factor of deteriorating with age.
- **Reaction time** - The reaction time is the timespan between the occurrence of a stimulus and the response of our body or mind. Responses may vary from a thought or a change in the mental state to a physical movement or an alteration in our physiology. Independently of the nature of the stimulus and the response, slow reaction times are usually associated with mental fatigue.
- **Concentration** - Concentration is the cognitive process of selectively focusing on one aspect of the environment while ignoring the others. The loss of concentration may be caused by external factors and may sometimes even be desirable (e.g. a sudden potentially hazardous event that must be analyzed before continuing with a given task). However, we frequently loose interest or focus in our tasks, at a rate that increases with the increase of the mental fatigue.
- **Accuracy and precision** - Accuracy and precision may be seen as the achievement of results that are within the expected quality and timeframe for a given individual, given his cognitive and physical skills. It also represents the absence of unexpected errors. This is among the factors that are more easily observable and measured while, at the same time, closely related to fatigue.

Mental fatigue is also affected by a number of other external factors. They may or may not be directly related to the individual's sphere, thus adding to the complexity of their comprehension and study [5,7]:

- **Mood** - The mood of the individual may influence decisively his/her the mental state, with a particular effect on his/her motivation to work. Although tired, the individual may overcome (even if only temporarily) the effects of fatigue with a positive mood and motivation.
- **Stress** - Stress may be defined as the set of responses of the individual's mind or body to external stimuli, allowing the individual to adapt to the dynamic requirements of the environment. These processes of adaptation, however, require an additional effort from the brain which, when prolonged over long or intense periods of time, will result in mental fatigue.
- **Sleepiness** - Sleepiness is often mistaken for mental fatigue or generalized as such. The difference exists and must be pointed out. However, the mistake is understandable since sleepiness is a symptom that is strongly connected to mental fatigue: it is one of the ways our brain uses to tell us that he is running out of resources. It often results in a general loss of the individual vitality.

2.1 Indicators of Mental Fatigue

The study of mental fatigue, including its causes and symptoms, is traditionally supported by self-reporting mechanisms (generally questionnaires) or, more recently, through the use of physiological sensors. The first approach, related to

Psychology, has a certain number of disadvantages: (1) people often lie or hide truths in questionnaires; (2) people are afraid or unwilling to answer correctly; (3) people frequently have wrong and subjective perceptions of their symptoms; (4) questionnaires are often hard to define correctly, with inaccurate questions or answers. The second approach, followed by Medicine, has the advantage of being more accurate. However, this has its price on the invasive and intrusive nature of the physiological sensors used, which may even influence the variables of the study[8,9].

In order to study fatigue we are thus following an approach already validated in the study of stress [10,11]: analyzing changes in the behavior of the user while within a given technologically empowered environment, in a non-intrusive and non-invasive way. This results in an environment that may adjust, in real-time, to significant changes in its context, with context being defined by the level of fatigue of the individuals.

The detection and classification of fatigue will be based on the collection of data about behavioral biometrics, specifically keystroke and mouse dynamics. A simple logger application was developed that acquires information about each mouse and keyboard action (e.g. mouse button down, mouse button up, key down, key up, mouse movement, among others), with the detail needed to generate the features pointed out below. Following this approach it is possible to collect data that will allow to discover behavioral patterns of interaction with the keyboard and mouse, in a non-intrusive and dynamic way. Particularly, the following features are considered:

- **Keydown Time** - time spent between the key down and the key up events;
- **Errors per Key Pressed** - number of times the backspace key is pressed, versus the keys pressed;
- **Mouse Velocity** - velocity at which the cursor travels;
- **Mouse Acceleration** - acceleration of the mouse at a given time;
- **Time between Keys** - time spent between each two keys pressed;
- **Total Excess of Distance** - excess of distance travelled by the pointer when considering two consecutive clicks;
- **Average Excess of Distance** - average of the distance excess travelled by the pointer when considering two consecutive clicks;
- **Double Click Speed** - speed of the double click;
- **Number of Double Clicks** - number of double clicks in a time frame;
- **Distance While Clicking** - distance travelled by the mouse while dragging objects;
- **Signed Sum of Angles** - how much the pointer "turned" left or right during its travel;
- **Absolute Sum of Angles** - how much the pointer "turned" during its travel, in absolute terms;
- **Sum of Distances between Path and Straight Line** - considering two consecutive clicks, it measures the distance between all the points of the path travelled by the mouse, and the closest point in a straight line (that represents the shortest path) between the coordinates of the two clicks;

- **Average Distance between Path and Straight Line** - the same as above, but provides an average value of the distance to the straight line; and
- **Time between Clicks** - time spent between each two clicks.

3 The Mental Fatigue Monitoring Framework

As seen in section 2, indicators of mental fatigue are defined by a set of metrics. In order to have a positive and optimized effect on the lives of its users, a framework for fatigue monitoring based on these metrics must perform classification and decision-making in real-time.

The architecture of the proposed framework includes not only the sheer acquisition and classification of the data, but also a presentation tier that will support the human-based or autonomous decision-making mechanisms that are now being implemented. It is a layered architecture. The first layer is the *Data acquisition* one. It is responsible for capturing information describing the behavioral patterns of the user, receiving data from events fired from the use of the mouse and the keyword. Therefore, this layer encodes each event with the corresponding necessary information (e.g. timestamp, coordinates, type of click, key pressed).

The second layer is the *Data processing*. In this layer the data received from the *Data acquisition* layer is processed and transformed in order to be evaluated according to the metrics presented. One of the most important tasks of this layer is to filter outlier values that would have a negative effect on the analysis (e.g. a key pressed for more than a certain amount of time).

The next layer is the *Classification layer*. This layer is responsible for interpreting data from the mental fatigue indicators and to build the meta data that will support decision-making. To do it, this layer uses the machine learning mechanisms detailed below.

After the classification, the *Data access layer*, is responsible for providing access to - and persisting all the information generated by - the lower layer. We are not only interested in allowing access to the data in real-time but also to persist the historic of the user as well as his/her profile, to allow analyses within longer time frames.

Finally, the *Presentation layer* includes the mechanisms to build intuitive and visual representations of the mental states of the users, abstracting from the complexity of the data level where they are positioned.

3.1 Classifying Fatigue

The classification of the mental state of a user is achieved through the use of the k-Nearest Neighbor algorithm (k-NN). It is a method of classification based on closest training examples in the feature space. The data used to train the model used by the k-NN algorithm was collected in an experiment performed with students of the University of Minho, as described in section 4. From the available features, only the ones showing the most statistical significance were

selected: Mouse Acceleration, Mouse Velocity, Keydown Time, Time between keys, and Error per key.

A different classifier is trained for each feature in order to get the level of fatigue due to each one. This is of significant importance since there are periods of time during which the framework only accesses a limited set of features (e.g. the user may not use the keyboard for a certain period of time). Thus, each feature is classified independently, assigning a binary value to each of its inputs, marking it as fatigued or not fatigued.

Afterwards, a weighted sum is computed involving all the features with available data. The weights of the features for the overall computation of the level of fatigue are given by the significance of the feature, computed during the statistical analysis of the data at hand, as described in section 4. This means that features that have been significantly affected by fatigue will have a larger contribution to the computed value than the ones that were hardly affected during the experiment. This weighted sum constitutes a fusion of the different features, resulting in the actual level of fatigue for a user. Using training data for a specific individual, when available, makes the fatigue detection model personalized, thus more accurate.

The prototype of the presentation layer (Figure 1) uses a set of emoticons to depict the state of each user in an intuitive way that is easy to understand. For more detail it is also possible to access a more detailed interface, which shows not only the overall state of fatigue of the user (in a binary form as fatigued or not fatigued) but also the level of fatigue, which actually represents the weighted sum of the features.

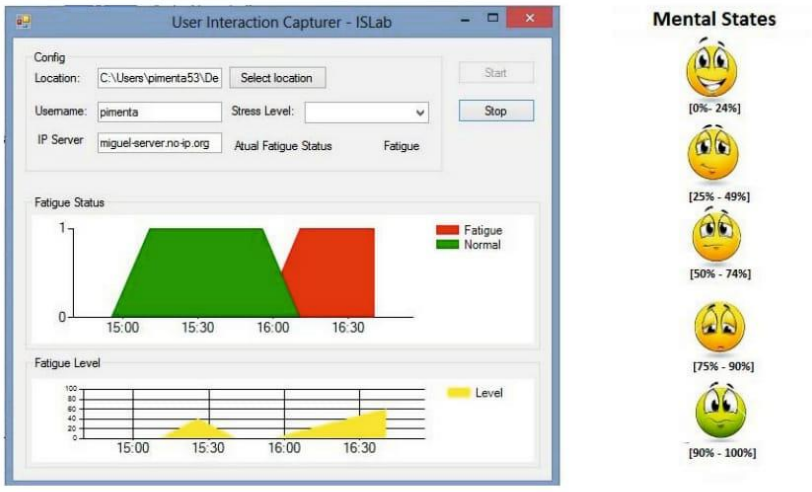


Fig. 1. The presentation layer of the monitoring system detailing the level of fatigue of an individual in part of the afternoon

4 Analysis of Results

Twenty participants took part on the experiment in order to collect data to train the fatigue detection module. In this experiment participants needed only have an application running that recorded all their interactions with the keyboard and the mouse. It did not interfere with their usage patterns nor needed the participants to perform additional or different activities. The participants (seventeen males and three females) were mostly volunteer students from our institution, aged eighteen to fifty. All these individuals were familiar with the technological devices used and the interaction with them was not an obstacle.

The collection of the data took place in two moments, for each user: the first at the beginning of the day, where participants are fully rested and in a normal mental state and the second at the end of the day. The two collection moments took place on the same day. It is also important to mention that the collection of data took place during the exam season and that the participants spent the day studying, with the support of a computer.

4.1 Statistical Data Analysis

To determine to which extent each feature considered is or is not influenced by mental fatigue, the data collected in the first phase (without fatigue) is compared with the data collected in the second phase (with expected mental fatigue). Given that most of the distributions of the data collected are not normal, the Mann-Whitney test is used to perform the analysis. This test is a non-parametric statistical hypothesis test for assessing whether one of two samples of independent observations tends to have larger values than the other. The null hypothesis is thus: $H_0 =$ *the medians of the two distributions are equal.* For each two distributions compared, the test returns a p-value, with a small p-value suggesting that it is unlikely that H_0 is true. Thus, for every Mann-Whitney test whose $p - value < \alpha$, the difference is considered to be statistically significant, i.e., H_0 is rejected, with $\alpha = 0.05$. The data analysis was performed using Wolfram Mathematica, Version 8.0.

A significant difference between the collected data in the two phases means that the feature is effectively influenced by mental fatigue for this specific individual. Statistically significant differences were observed in the features Keydown Time, Errors per Key Pressed, Time Between Keys, Mouse Velocity and Mouse Acceleration, hence only these are used for building the model.

For each feature and for each of the two moments of data collection, the average and median values were analyzed in order to determine the trend of the value, i.e., we wanted to answer the question "does it tend to decrease or to increase under fatigue?". Table 1 summarizes these observations.

The results obtained lead us to conclude that all the features have a significantly marked trend, either increasing or decreasing. The only exception to this is the feature **Time between Clicks**, which increased to half of the participants and decreased to the other half.

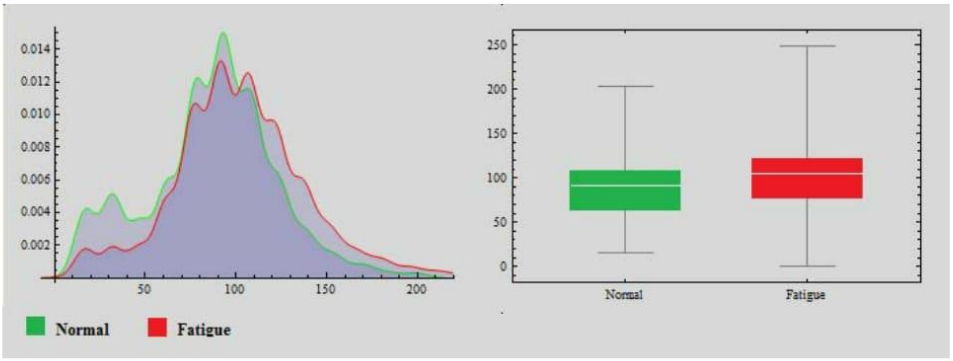


Fig. 2. Histograms and Box Plots comparing the data of the two distributions for the feature **Keydown Time** of a specific volunteer: fatigued individuals tend to write slower

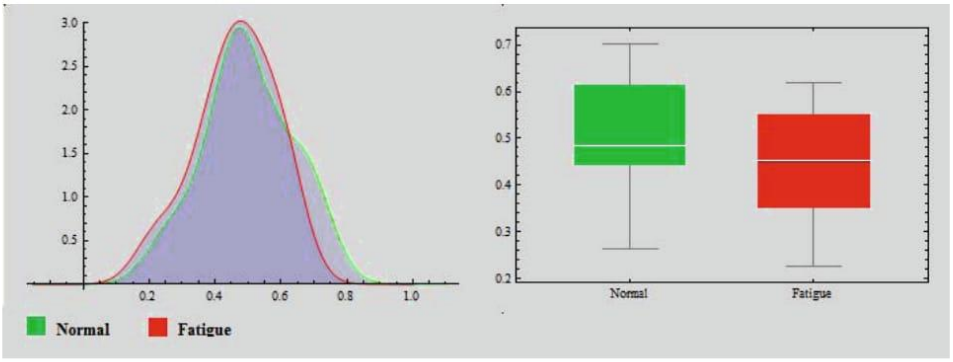
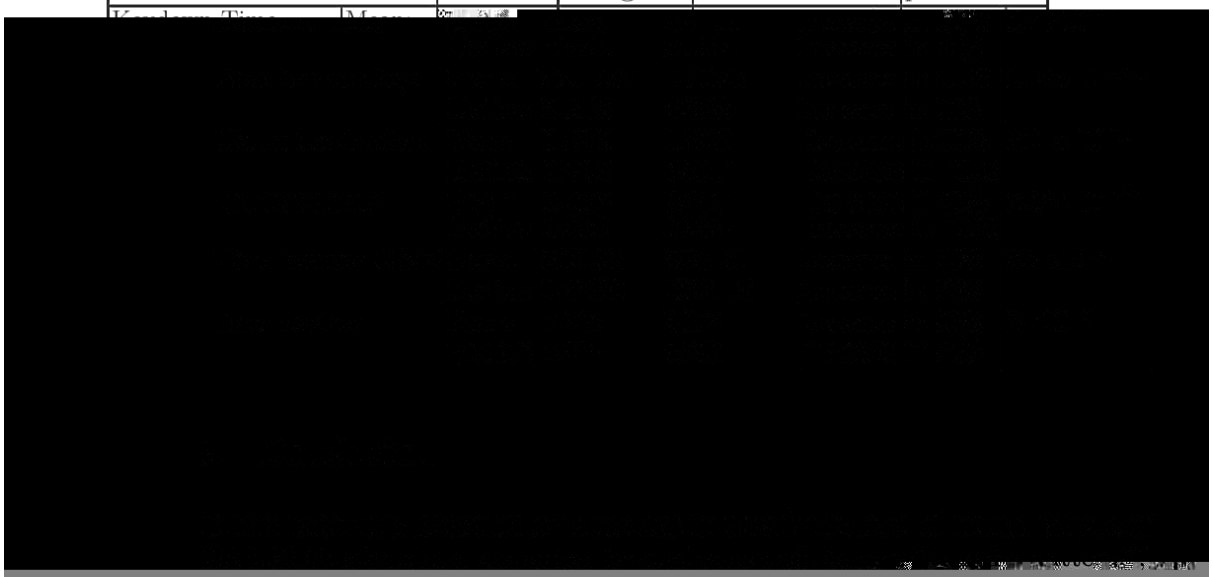


Fig. 3. Histograms and Box Plots comparing the distributions of the data collected in the two moments for the feature **Mouse Velocity** of a volunteer: fatigued individuals move the mouse slower.

After the statistical analysis performed, one may claim that it is indeed possible to detect mental fatigue through the handling of keyboard and mouse usage data. It also proves, once more, that the presence of mental fatigue is accompanied by a loss of performance and an increase in errors, and that this loss may be measured using non-invasive tools. This loss of performance and increase in the number of errors are consequences of decreased cognitive skills, which are caused by the increasing level of fatigue being measured during the day's work. The decrease in the cognitive skills is evidenced by general patterns of slower mouse (mouse acceleration and velocity) and keyboard (Keydown Time, Time between keys) interactions, as well as by an increasing number of errors (Error per key).

Table 1. Results of the statistical analysis of the data for the 20 participants. Only the features that have shown significant differences are included. The "trend" column depicts the percentage of participants that have a given trend. For example, the mean value of the velocity of the mouse decreases for 90% of the students, when fatigued

Metric		Normal	Fatigued	Trend	p-value
Keyboard: Time	Mean				





and Technology) within project FCOMP-01-0124-FEDER-028980 (PTDC/EEI-SII/1386/2012) and PEst-OE/ EEI/UI0752/2011. The work of Davide Carneiro is also supported by a doctoral grant by FCT (SFRH/BD/64890/2009).

References

1. van der Linden, D., Eling, P.: Mental fatigue disturbs local processing more than global processing. Psychological Research 70(5), 395–402 (2006)
2. Williamson, R.J., Purcell, S., Sterne, A., Wessely, S., Hotopf, M., Farmer, A., Sham, P.C.: The relationship of fatigue to mental and physical health in a community sample. Social Psychiatry and Psychiatric Epidemiology 40(2), 126–132 (2005)
3. Nussbaum, M.A., Babski-Reeves, K.L., Kleiner, B.M., Smith-Jackson, T.L., Barker, L.M.: Measuring and modeling the effects of fatigue on performance: Specific application to the nursing profession Linsey Marinn Barker Virginia Polytechnic Institute and State University to partially fulfill the requirements of a Doctor of Philosophy In Indus. PhD thesis (2009)
4. Winwood, P.C., Winefield, A.H., Dawson, D., Lushington, K.: Development and Validation of a Scale to Measure Work-Related Fatigue and Recovery: The Occupational Fatigue Exhaustion/Recovery Scale (OFER). Journal of Occupational and Environmental Medicine 47(6), 594–606 (2005)
5. Kobayashi, H., Demura, S.: Relationships between Chronic Fatigue, Subjective Symptoms of Fatigue, Life Stressors and Lifestyle in Japanese High School Students. School Health 2, 5 (2006)
6. Joyce, E., Blumenthal, S., Wessely, S.: Memory, attention, and executive function in chronic fatigue syndrome. Journal of Neurology, Neurosurgery, and Psychiatry 60(5), 495–503 (1996)
7. Hossain, J.L., Ahmad, P., Reinish, L.W., Kayumov, L., Hossain, N.K., Shapiro, C.M.: Subjective fatigue and subjective sleepiness: two independent consequences of sleep disorders? Journal of Sleep Research 14(3), 245–253 (2005)
8. Joyce, J., Rabe-Hesketh, S., Wessely, S.: Reviewing the reviews: the example of chronic fatigue syndrome. JAMA: The Journal of the American Medical Association 280(3), 264–266 (1998)
9. Neuberger, G.B.: Measures of fatigue: The Fatigue Questionnaire, Fatigue Severity Scale, Multidimensional Assessment of Fatigue Scale, and Short Form-36 Vitality (Energy/Fatigue) Subscale of the Short Form Health Survey. Arthritis & Rheumatism 49(S5), S175–S183 (2003)
10. Novais, P., Carneiro, D., Gomes, M., Neves, J.: Non-invasive Estimation of Stress in Conflict Resolution Environments. In: Demazeau, Y., Müller, J.P., Rodríguez, J.M.C., Pérez, J.B. (eds.) Advances on PAAMS. AISC, vol. 155, pp. 153–160. Springer, Heidelberg (2012)
11. Carneiro, D., Castillo, J.C., Novais, P., Fernández-Caballero, A., Neves, J.: Multimodal Behavioural Analysis for Non-invasive Stress Detection (July 2012)

2.2 Analysis of human performance as a measure of mental fatigue

Title	**Analysis of human performance as a measure of mental fatigue**
Authors	A Pimenta, D Carneiro, P Novais, J Neves
Publication Type	Conference Proceedings
Publication Name	International Conference on Hybrid Artificial Intelligence Systems
Publisher	Springer, Cham
Volume	8480
Number	
Pages	389-401
Year	2014
Month	June
Online ISSN	978-3-319-07617-1
Print ISSN	978-3-319-07616-4
Url	http://dx.doi.org/10.1007/978-3-319-07617-1_35
State	Published
Scimago journal ranking	SJR IF 0.339 (Q2 - COMPUTER SCIENCE c
Indexed	ISI Web of Science, DBLP

Analysis of Human Performance as a Measure of Mental Fatigue

André Pimenta, Davide Carneiro, Paulo Novais, and José Neves

CCTC/DI - Universidade do Minho
Braga, Portugal
{apimenta,dcarneiro,pjon,jneves}@di.uminho.pt

Abstract. In our daily life, we often have the feeling of being exhausted due to mental or physical work, and a sense of performance degradation in the execution of simple tasks. The maximum capacity of operation and performance of an individual, whether physical or mental, usually also decreases gradually as the day progresses. The loss of these resources is linked to the onset of fatigue, which is particularly noticeable in long and demanding tasks or repetitive jobs. However, good management of the working time and effort invested in each task, as well as the effect of breaks at work, can result in better performance and better mental health, delaying the effects of fatigue. This paper details a non-invasive approach on the monitoring of fatigue of a human being, based on the analysis of the performance of his interaction with the computer.

Keywords: Fatigue, Mental Fatigue, Performance, Behavioural Analysis, Classification.

1 Introduction

In today's busy society, people push their limits in order to find a balance between ever ambitious goals and time to dedicate to enjoyable moments, family or personal projects [1]. This balance, when it comes to exist, is generally obtained at the expenses of less time for rest and sleep, with an impact on the well-being and health of the individual. Fatigue, above all others, must be considered here. The effects of fatigue may not be felt immediately. Nonetheless, when prolonged it may have serious consequences on health and safety at the workplace as well as on labour productivity [2]. Ultimately, fatigue can put people who have safety-sensitive occupations at risk, as any mistake can lead to the loss of their own lives or those of others [3].

This paper presents a new approach on the problem of fatigue detection and monitoring, that can be included in the umbrella of Hybrid Artificial Intelligent Systems [4,5]. It looks at the effects of fatigue on the performance of the individual's interaction with the computer. This approach can be deemed both non-invasive and non-intrusive as it relies on the observation of the individual's use of the mouse and keyboard. It builds on a previous study, now to address the evolution of fatigue throughout the day and the role that task complexity and Circadian Rhythm play in the degradation of performance.

M. Polycarpou et al. (Eds.): HAIS 2014, LNAI 8480, pp. 389–401, 2014.
© Springer International Publishing Switzerland 2014

1.1 Circadian Rhythm, Environment and Their Role on Fatigue

In studying fatigue one must acknowledge the major role of the biological clock, i.e., one cannot expect to beat one's rhythm and biological need to rest. Indeed, the role of our biological clock goes beyond compelling the body to fall asleep and to wake up again. It also modulates our hour-to-hour waking behaviour, which in turn affects our sense of fatigue, alertness and performance, generating circadian rhythmicity in almost all neuro-behavioural variables [6]. This means that there is a natural fluctuation of the level of fatigue we experience during the day, that is independent of the intensity of the tasks we perform. It can be seen as a base-level of fatigue.

The context and environment in which the individual carries out the activity are also very important in determining the level of fatigue during the day. Indeed, factors such as workload, distractions, periods of boredom, motivational factors, stress, food intake, posture, ambient temperature, background noise, lighting conditions or drug intake (e.g., caffeine) are examples of contextual variables that can have a positive or negative influence on the level of fatigue of a human being [7]. Smith and Miles conducted a study which analyses the effects of lunch on cognitive vigilance tasks [8]. The authors analysed the performance of individuals who did not take lunch before performing tasks compared to those who took, to show some differences between morning and afternoon performance in subjects who abstained from eating lunch.

Figure 1 details the influence of several environment variables and external factors in task performance, as well as their relationship. Black boxes depict task characteristics and while red ones depict characteristics of the Human operator. The dotted line indicates a positive correlation (measured with the Pearson product-moment correlation coefficient) between actual, measurable task workload and the worker's perception of that workload (estimated with a mental workload tool such as NASA-TLX [9] or SWAT [10]).

1.2 Human Performance as a Measure of Fatigue

When under the effect of prolonged excessive mental workload, inadequate sleep patterns, or circadian effects, Humans may exhibit slowed cognitive processes or may not respond at all to the stimuli and information received. This is generally due to the amount of information exceeding the brain's processing capacity at that moment. In contrast, when the mental workload is significantly below the adequate level, the individual tends to get tired and bored, which will likely result in an increase in the number of errors in the tasks being performed [12]. Performance as a measure of fatigue has been used previously as an objective measure of fatigue in humans. Morris & Miller, for example, conducted a study which analysed the performance of pilots during simulated flight as a means of detecting and measuring fatigue [13]. The authors have shown that the decrease in flight performance occurs due to fatigue.

Consequently, the measurement of the performance of the individual while performing a task results in an objective way of measuring fatigue in the human being as a relationship between the two exists. Usually, tasks that imply

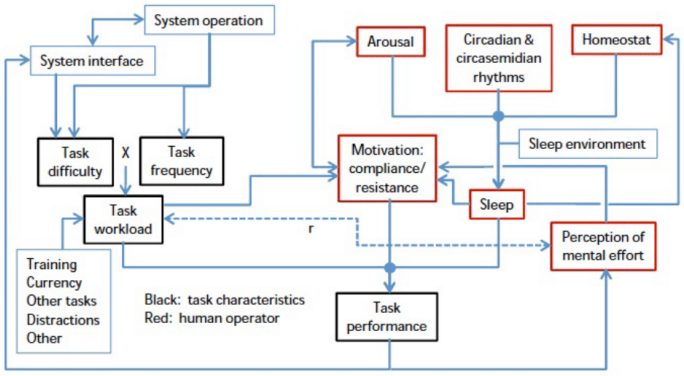


Fig. 1. Environment variables and external factors on Mental Performance Task (Source: [11])

excessive mental and physical workload as well as very low workload (observed in monotonous and repetitive activities) result in significant performance decrements. Thus being, the loss of performance due to fatigue can be studied as a function that takes as input the duration of the task and the quantity and/or quality of the work carried out [2,14].

The challenge lies in the fact that a decrease in the performance is often not "seen" immediately, which may be explained by the concept of *reverse capacity* [15]. The theory of reverse capacity states that an individual rarely works at his maximum capacity, and can thus temporarily absorb the load of cognitive or physical additional work while still maintaining performance indicators, as long as their inverse capacity is not exceeded. For this reason, the effects of fatigue may not be immediately visible [15]. The effects of fatigue on performance, in particular, may only be felt in the course of time, when it may be to late to avoid them.

2 A Non-Invasive Approach to Fatigue Monitoring

This paper introduces a system for detecting and monitoring mental fatigue through the analysis and measurement of human performance indicators. As previously postulated in [16] it uses the keyboard and mouse as sensors of user performance. Specifically, the user is monitored through particular Operating System events fired from the use of the computer's mouse and keyboard:

– MOV, timestamp, posX, posY - an event describing the movement of the mouse, in a given time, to coordinates (posX, posY) in the screen;

- MOUSE_DOWN, timestamp, [Left|Right], posX, posY - this event describes the first half of a click (when the mouse button is pressed down), in a given time. It also describes which of the buttons was pressed (left or right) and the position of the mouse in that instant;
- MOUSE_UP, timestamp, [Left|Right], posX, posY - an event similar to the previous one but describing the second part of the click, when the mouse button is released;
- MOUSE_WHEEL, timestamp, dif - this event describes a mouse wheel scroll of amount dif, in a given time;
- KEY_DOWN, timestamp, key - identifies a given key from the keyboard being pressed down, at a given time;
- KEY_UP, timestamp, key - describes the release of a given key from the keyboard, in a given time;

From these events, that fully describe the interaction of the user with the mouse and keyboard, it is possible to extract the following features:
Key Down Time - the timespan between two consecutive KEY_DOWN and KEY_UP events, i.e., for how long was a given key pressed.
Units - milliseconds

Time Between Keys - the timespan between two consecutive KEY_UP and KEY_DOWN events, i.e., how long did the individual took to press another key.
Units - milliseconds

Mouse Velocity - The distance travelled by the mouse (in pixels) over the time (in milliseconds). The velocity is computed for each interval defined by two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, mup and mdo, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$, that took place respectively in the instants $time_1$ and $time_2$. Let us also assume two vectors $posx$ and $posy$, of size n, holding the coordinates of the consecutive MOUSE_MOV events between mup and mdo. The velocity between the two clicks is given by $r_dist/(time_2 - time_1)$, in which r_dist represents the distance travelled by the mouse and is given by equation 1.
Units - pixels/milliseconds

$$r_dist = \sum_{i=0}^{n-1} \sqrt{(posx_{i+1} - posx_i)^2 + (posy_{i+1} - posy_i)^2} \tag{1}$$

Mouse Acceleration - The velocity of the mouse (in pixels/milliseconds) over the time (in milliseconds). A value of acceleration is computed for each interval defined by two consecutive MOUSE_UP and MOUSE_DOWN events, using the intervals and data computed for the Velocity.
Units - pixels/milliseconds^2

Time Between Clicks - the timespan between two consecutive MOUSE_UP and MOUSE_DOWN events, i.e., how long did it took the individual to perform

another click.
Units - milliseconds

Double Click Duration - the timespan between two consecutive MOUSE_UP events, whenever this timespan is inferior to 200 milliseconds. Wider timespans are not considered double clicks.
Units - milliseconds

Average Excess of Distance - this feature measures the average excess of distance that the mouse travelled between each two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. To compute this feature, first it is measured the distance in straight line between the coordinates of *mup* and *mdo* as depicted in equation 2.

$$s_dist = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \tag{2}$$

Then, it is measured the distance actually travelled by the mouse by summing the distance between each two consecutive MOUSE_MV events. Let us assume two vectors *posx* and *posy*, of size n, holding the coordinates of the consecutive MOUSE_MV events between *mup* and *mdo*. The distance actually travelled by the mouse, $real_dist$ is given by equation 1. The average excess of distance between the two consecutive clicks is thus given by r_dist/s_dist.
Units - pixels

Average Distance of the Mouse to the Straight Line - in a few words, this feature measures the average distance of the mouse to the straight line defined between two consecutive clicks. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. Let us also assume two vectors *posx* and *posy*, of size n, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The sum of the distances between each position and the straight line defined by the points $(x1, y1)$ and $(x2, y2)$ is given by equation 3, in which $ptLineDist$ returns the distance between the specified point and the closest point on the infinitely-extended line defined by $(x1, y1)$ and $(x2, y2)$. The average distance of the mouse to the straight line defined by two consecutive clicks is this given by s_dists/n.
Units - pixels

$$s_dists = \sum_{i=0}^{n-1} ptLineDist(posx_i, posy_i) \tag{3}$$

Distance of the Mouse to the Straight Line - this feature is similar to the previous one in the sense that it will compute the s_dists between two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, according to equation 3. However, it returns this sum rather than the average value during

the path.
UNITS - pixels

SIGNED SUM OF ANGLES - with this feature the aim is to determine if the movement of the mouse tends to "turn" more to the right or to the left. Let us assume three consecutive MOUSE_MOVE events, $mov1, mov2$ and $mov3$, respectively in the coordinates $(x1, y1)$, $(x2, y2)$ and $(x3, y3)$. The angle α between the first line (defined by $(x1, y1)$ and $(x2, y2)$) and the second line (defined by $(x2, y2)$ and $(x3, y3)$) is given by $degree(x1, y1, x2, y2, x3, y3) = \tan(y3 - y2, x3 - x2) - \tan(y2 - y1, x2 - x1)$. Let us now assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*. Let us also assume two vectors *posx* and *posy*, of size n, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The signed sum of angles between these two clicks is given by equation 4.
UNITS - degrees

$$s_angle = \sum_{i=0}^{n-2} degree(posx_i, posy_i, posx_{i+1}, posy_{i+1}, posx_{i+2}, posy_{i+2}) \quad (4)$$

ABSOLUTE SUM OF ANGLES - this feature is very similar to the previous one. However, it seeks to find only how much the mouse "turned", independently of the direction to which it turned. In that sense, the only difference is the use of the absolute of the value returned by function $degree(x1, y1, x2, y2, x3, y3)$, as depicted in equation 5.
UNITS - degrees

$$s_angle = \sum_{i=0}^{n-2} \mid degree(posx_i, posy_i, posx_{i+1}, posy_{i+1}, posx_{i+2}, posy_{i+2}) \mid \quad (5)$$

DISTANCE BETWEEN CLICKS - represents the total distance travelled by the mouse between two consecutive clicks, i.e., between each two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. Let us also assume two vectors *posx* and *posy*, of size n, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The total distance travelled by the mouse is given by equation 1.

From these features it is possible to obtain a measure of the user's performance (e.g. an increased distance between clicks or sum of angles represents decreased performance). Once information about the individual's performance exists in these terms, it is possible to start monitoring his fatigue, in real-time, and without the need for any explicit or conscious interaction. This makes this approach especially suited to be used in work environments in which people use computers as it requires no change in their working routines. This is the main advantage

of this work, especially when compared to more traditional approaches that still rely on questionnaires (with issues concerning wording or question construction), specific hardware (that has additional costs and is frequently intrusive) or the availability of human experts.

3 Case Study

In order to assess the validity of the approach described in the previous section a case study was implemented aimed at collecting data over a relatively long period of time. This allowed us to determine if and how people are affected by fatigue in each of the mentioned features.

3.1 Participants

The participants, twenty seven in total (20 men, 7 women) were students from the University of Minho who volunteered to participate. Their age ranged between 18 and 30. The following requirements were established to select, among all the volunteers, the ones that participated: (1) familiarity and proficiency with the use of the computer; (2) use of the computer on a daily basis and throughout the day; (3) owning at least one personal computer. These restrictions constituted no problem as a large slice of the population of out institution, and particularly the ones affiliated to the Department of Informatics, fulfil the requirements.

3.2 Methodology

The methodology followed to implement the study was devised to be as minimally intrusive as the approach it aims to support. Participants were provided with an application for logging the previously mentioned events of the mouse and keyboard. This application, which maintained the confidentiality of the keys used, needed only to be installed in the participant's primary computer and would run on the background, starting automatically with the Operating System. The only explicit interaction needed from the part of the user was the input of very basic information on the first run, including the identification and age.

This application was kept running for approximately one month, collecting interaction data whenever there was interaction with the computer. During the whole process, participants did not need to perform any additional task and were asked to perform their activities regularly, whether they were work or leisure-related, as they would if they were not participating on the study.

Finally, when the period of one month ended, participants were asked to send in the file containing the log of their interaction during the duration of the study. The data collected was used to build the features described previously. The resulting dataset as well as the process by means of which data were analysed is described in the following sub-sections.

3.3 Data Selection

As previously stated, the data collected spans a period of approximately 30 complete days of computer use, in a work context as well as in a leisure one. From this data, we extracted particular moments for analysis.

Specifically, four distinct periods of one our each were selected in order to analyze the effect of Circadian rhythm and the effect of acute fatigue in a work day: (1) the start of the day, when the user is mentally fresh; (2) immediately before lunch break; (3) after lunch break; and (4) the end of the day, when the individual is most fatigued. This selection stems from the willingness to study the differences between the beginning and end of the day as well the potential effect of a lunch break. Lunch break is automatically inferred from the lack of interaction during a relative long time in a given period.

3.4 Statistical Data Analysis

To determine the effects of acute fatigue on the interaction patterns of an individual it is necessary to determine that the distributions of the data collected in each moment are statistically and significantly different. To this aim, the following approach was implemented.

First, it was determined, using the Pearson's chi-squared test, that most of the distributions of the data collected are not normal. Given this, the Kruskal-Wallis test was used in the subsequent analysis. This test is a non-parametric statistical method for testing whether samples originate from the same distribution. It is used for comparing more than two samples that are independent, and thus prove the existence of distinct behaviours. The null hypothesis considered is: H_0 : all samples have identical distribution functions against the alternative hypothesis that at least two samples have different distribution functions. For each set of samples compared, the test returns a $p-value$, with a small $p-value$ suggesting that it is unlikely that H_0 is true. Thus, for every Kruskal-wallis test whose $p - value < \alpha$, the difference is considered to be statistically significant, i.e., H_0 is rejected. In this work a value of $\alpha = 0.05$ is considered, a standard value generally accepted by research.

This statistical test is performed for each of the features considered, with the intention of determining if there are statistically significant differences between the several distributions of data, which will in turn confirm the existence of effects of fatigue on the interaction patterns.

Table 1 details, for each feature, the percentage of participants whose interaction patterns were significantly affected by fatigue and the average value of the $p - value$.

After having determined that a large majority of the participants were indeed significantly affected by acute fatigue, it was analysed the evolution of user performance throughout the day. Figure 2 shows these effects for a particular participant, in two different features and in each of the four previously mentioned moments. It is possible to see how, during the day, the excess of distance between clicks and the time between keys increases gradually. Both examples essentially depict a significant loss of performance.

Table 1. Percentage of participants whose interaction patterns were significantly affected by fatigue, in each feature, and the average value of the $p - value$

Metric	significant	p-value
Time between keys	100%	9.88e-74
Key down time	88.0%	7.42e-16
Mouse Acceleration	100%	3.83e-284
Mouse Velocity	100%	1.02e-126
Average Excess of Distance	94.4%	1.36e-13
Distance Between Clicks	94.4%	2.83e-07
Click Durations	94.4%	4.46e-143
Errors per key	77.7%	6.54e-21

In order to understand the effects of fatigue during the day it is useful to look at the evolution of the features. Figures 3 and 4 depict, respectively, the evolution of the velocity of the mouse and the key down time during the day, for a particular participant. They essentially depict the same trend of decreasing efficiency: the velocity of the mouse decreases from nearly 0.9 pixels/ms to 0.6 pixels/ms while the key down time increases from 86 ms to more than 90 ms. Both pictures belong to the same participant and the same day, between 9:27 AM to 5:50 PM.

Both the statistical differences and the trend identified show that there is indeed a gradual loss of performance over the day. Moreover, it is also proved that this loss of performance can be detected through non-invasive methods, specifically by looking at how the user interacts with the computer.

3.5 Classification Results

Having proved the existence of a change on the behaviours over the day due to fatigue, as well as the loss of performance in the use of mouse and keyboard, the final step is to train a classifier that can accurately quantify the level of fatigue of an individual.

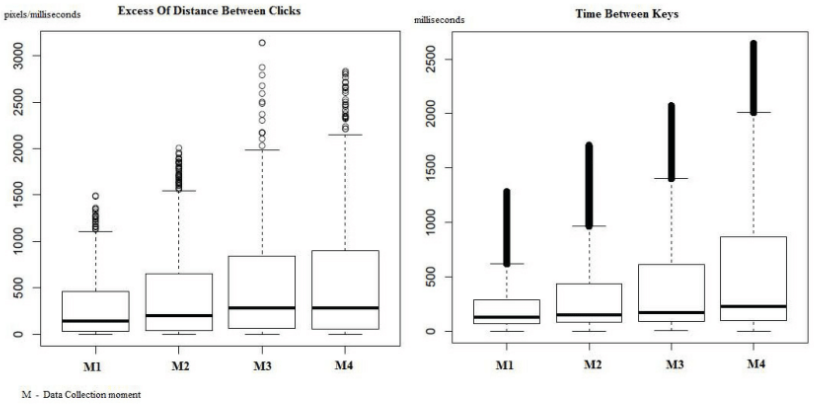


Fig. 2. The effects of fatigue on the excess of distance (left) and on the time between keys (right) for a participant, in each of the four moments

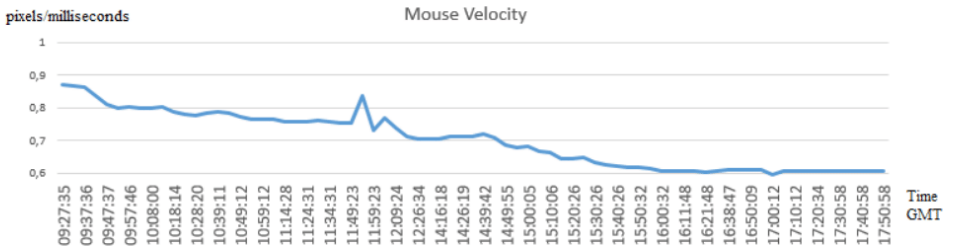


Fig. 3. Fatigue causes a gradual and consistent decrease in the mouse velocity over the day

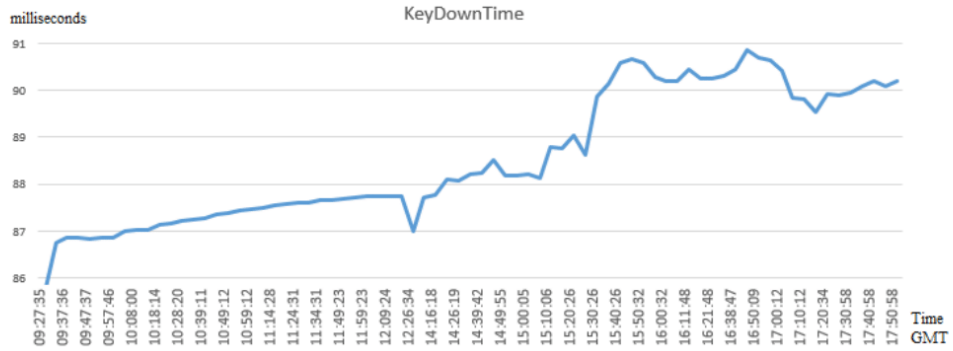


Fig. 4. Fatigue causes a gradual and consistent increase in the key down time over the day

Table 2. Results of several classifiers on WEKA

data set	Correctly Classified	Incorrectly Classified	Kappa s	MAE	ROC Area	RMSE
MultilayerPerceptron	92,97%	7.03%	0.83	0.09	0.939	0.240
KNN (k=10)	**96,85%**	**3.15%**	**0.92**	**0.04**	**0.992**	**0.155**
BayesNet	88,68%	11.31%	0.71	0.13	0.887	0.292
Regression	96,89%	3.10%	0.92	0.07	0.981	0.167

To this aim, the performance of several classifiers using the data collected was assessed. Table 2 depicts the performance of these classifiers, with the KNN holding the best results although all of them behave fairly good. These tests were performed with a dataset with a total of 2514 instances and 10-fold cross method.

To build the dataset it was used part of the data collected in the case study described. Specifically, we selected instances from the two most different distributions: instances resulting from periods where the performance is lower(M4) were labelled as *Fatigue*, and instances resulting from periods of better performance (M1) were labelled as *Normal*. The dataset describes 10 attributes: key down time, time between keys, error per key, mouse acceleration, mouse velocity, distance between clicks, click duration and average excess of distance. These attributes were selected to represent the features that revealed, in the overall, to be more significantly affected by fatigue in the moments considered.

Given this, we are currently using the k-Nearest Neighbor algorithm (k-NN) for classifying the interaction of a user with the computer in terms of fatigue. This is a non-parametric method that consists of a majority vote of its neighbours, with the object being assigned to the most common class among its k nearest neighbours. In this specific case, a value of k = 10 was used.

4 Conclusion

This paper presented an approach for measuring the level of fatigue of an individual through the analysis of his performance. This is achieved by analysing the interaction patterns of the individual with the mouse and keyboard. Thus being, it relies on behavioural analysis rather on traditional physiological sensors or questionnaires, being non-invasive, non-intrusive and objective.

The results achieved from the implementation of the case study detail prove that it is indeed possible to analyse and quantify human performance through the use of the mouse and keyboard, and this is a valid way to measure fatigue. It was also observed that the trend is towards the decrease in the performance of the interaction, visible through a generalized slower use of both the mouse and the keyboard or an increased distance between consecutive clicks.

This work opens the door to the development of real-time systems for fatigue monitoring through performance indicators. These can contribute to the development of healthier habits and even improve the quantity and quality of the

work produced. All in all, such approaches can have a positive effect on both the worker and the institution. Nonetheless, we acknowledge that more accurate approaches can be developed when considering more specific context information not yet considered, such as the type of the task or workload, as well as environment information. This will be essential to develop accurate approaches to very specific domains in which the interaction with the peripherals is necessarily different than the one we studied in this paper. Moreover, we will also validate this approach against other existing ones, namely self-report mechanisms and EEG.

Acknowledgments. This work was developed in the context of the project CAMCoF - Context-aware Multimodal Communication Framework funded by ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-028980. The work of Davide Carneiro is supported by an FCT post-doctoral grant (PTDC/EEI-SII/1386/2012).

References

1. Williamson, R.J., Purcell, S., Sterne, A., Wessely, S., Hotopf, M., Farmer, A., Sham, P.C.: The relationship of fatigue to mental and physical health in a community sample. Social psychiatry and psychiatric epidemiology 40(2), 126–132 (2005)
2. Tucker, P.: The impact of rest breaks upon accident risk, fatigue and performance: a review. Work & Stress 17(2), 123–137 (2003)
3. Jaber, M.Y., Neumann, W.P.: Modelling worker fatigue and recovery in dual-resource constrained systems. Computers & Industrial Engineering 59(1), 75–84 (2010)
4. Woźniak, M., Graña, M., Corchado, E.: A survey of multiple classifier systems as hybrid systems. Information Fusion 16(0), 3–17 (2014); Special Issue on Information Fusion in Hybrid Intelligent Fusion Systems
5. Borrajo, M., B.B.C.E.B.J.C.J.: Hybrid neural intelligent system to predict business failure in small-to-medium-size enterprises. International Journal of Neural Systems 21(04), 277–296 (2011); PMID: 21809475
6. Harris, W.: Fatigue, circadian rhythm, and truck accidents. In: Vigilance, pp. 133–146. Springer, Heidelberg (1977)
7. Folkard, S., Tucker, P.: Shift work, safety and productivity. Occupational Medicine 53(2), 95–101 (2003)
8. Smith, A.P., Miles, C.: The effects of lunch on cognitive vigilance tasks. Ergonomics 29(10), 1251–1261 (1986)
9. Hart, S.G., Staveland, L.E.: Development of nasa-tlx (task load index): Results of empirical and theoretical research. Advances in Psychology 52, 139–183 (1988)
10. Reid, G.B., Nygren, T.E.: The subjective workload assessment technique: A scaling procedure for measuring mental workload. Advances in Psychology 52, 185–218 (1988)
11. Turner, R.: The handbook of operator fatigue. Ergonomics 56(9), 1486–1486 (2013)

12. McClernon, C.K., Miller, J.C.: Variance as a measure of performance in an aviation context. The International Journal of Aviation Psychology 21(4), 397–412 (2011)
13. Morris, T., Miller, J.C.: Electrooculographic and performance indices of fatigue during simulated flight. Biological Psychology 42(3), 343–360 (1996)
14. Tanabe, S., Nishihara, N.: Productivity and fatigue. Indoor Air 14(s7), 126–133 (2004)
15. Kahneman, D.: Remarks on attention control. Acta Psychologica 33, 118–131 (1970)
16. Pimenta, A., Carneiro, D., Novais, P., Neves, J.: Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. In: Pan, J.-S., Polycarpou, M.M., Woźniak, M., de Carvalho, A.C.P.L.F., Quintián, H., Corchado, E. (eds.) HAIS 2013. LNCS, vol. 8073, pp. 222–231. Springer, Heidelberg (2013)

2.3 A neural network to classify fatigue from human–computer interaction

Title	**A neural network to classify fatigue from human–computer interaction**
Authors	A Pimenta, D Carneiro, J Neves, P Novais
Publication Type	Journal
Publication Name	Neurocomputing
Publisher	Elsevier
Volume	172
Number	
Pages	413–426
Year	2016
Month	January
Online ISSN	0925-2312
Print ISSN	
Url	http://dx.doi.org/10.1016/j.neucom.2015.03.105
State	Published
Scimago journal ranking	IF 1,202 (Q1 - COMPUTER SCIENCE APPLICATIONS)
ISI JCR	IF 2.392 (Q1 - COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE)

ARTICLE IN PRESS

Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A neural network to classify fatigue from human–computer interaction

André Pimenta *, Davide Carneiro [1], José Neves [2], Paulo Novais [3]

CCTC/Department of Informatics, University of Minho, Campus of Gualtar, 4710 - 057 Braga, Portugal

ARTICLE INFO

Article history:
Received 15 September 2014
Received in revised form
8 January 2015
Accepted 18 March 2015

Keywords:
Mental fatigue
Ambient intelligence
Neural networks
Human computer interaction

ABSTRACT

Fatigue, especially in its mental form, is one of the most worrying health problems nowadays. It affects not only health but also motivation, emotions and feelings and has an impact both at the individual and organizational level. Fatigue monitoring and management assumes thus, in this century, an increased importance, that should be promoted by private organizations and governments alike. While traditional approaches are mostly based on questionnaires, in this paper we present an alternative one that relies on the observation of the individual's interaction with the computer. We show that this interaction changes with the onset of fatigue and that these changes are significant enough to support the training of a neural network that can classify mental fatigue in real time. The main outcome of this work is the development of non-invasive systems for the continuous classification of mental fatigue that can support effective and efficient fatigue management initiatives, especially in the context of desk jobs.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In most of the so-called developed countries people have nowadays increasingly busier lifestyles. This makes them stretch their limits to find time for work, leisure, family and friends. This necessary extra time is frequently obtained at the expense of smaller periods of sleep or rest, and with a cost in terms of pressure and stress. This, together with other issues such as increasing workloads or harsher work conditions, results in the emergence of fatigue as one of the most worrying health conditions of this century [1,2]. Although the effects of fatigue may not be readily visible, they have consequences at many levels other than health, including emotions, productivity, performance, social behaviour, among others [3].

Generally, the term fatigue is used to describe a series of manifestations that range from drowsiness or lack of concentration to lack of physical strength or agility [4]. Thus, the concept of fatigue is very broad and subjective. It can be seen as a combination of symptoms, including impaired performance (loss of attention, slowed reaction and response times, impaired decision making, and poor performance on tasks that usually reflect good performance), and subjective feelings of sleepiness and tiredness [5,6]. It can also be divided in two main types [7]: mental fatigue (when it is our cognitive abilities that are decreased) and physical fatigue (when we become physically impaired due to an excess of a specific physical activity). In this paper we focus on mental fatigue, although a relationship between the two types, that sometimes interact, is acknowledged.

Fatigue can be seen all around us, on an everyday basis: students approaching an exam that sleep less in order to have more time to study, office or industrial workers working long, night or consecutive shifts, healthcare professionals subject to long hours of highly wearing conditions, drivers of heavy machinery, or specialized personnel such as the military or fire fighters. Either in trivial tasks or in more specialized ones, fatigue is regarded as one of the main causes of human error [2]. It may lead, especially in the aforementioned scenarios, to serious life-or-death situations, when people working in risky jobs have their performance, attention or motivation impaired due to fatigue.

Fatigue can occur at any time during the day and is not necessarily related to the circadian rhythm [8] alone but results of more complex set of factors that include time of rest, motivation, physical or cognitive impairments or social environment [9]. Depending on its duration and intensity, fatigue can make the carrying out of daily tasks increasingly hard or even impossible [10]. In severe or prolonged cases it can cause illnesses such as depression or chronic fatigue syndrome, with an impact at a personal, social and economic level. Given all this, it is nowadays clear that the detection, monitoring and management of fatigue should

* Corresponding author. Tel.: +351 253604435.
E-mail addresses: apimenta@di.uminho.pt (A. Pimenta), dcarneiro@di.uminho.pt (D. Carneiro), jneves@di.uminho.pt (J. Neves), pjon@di.uminho.pt (P. Novais).
URLS: http://isrt.di.uminho.pt/davide (A. Pimenta), http://isrt.di.uminho.pt/davide (D. Carneiro), http://www3.di.uminho.pt/~jmn/ (J. Neves), http://www.di.uminho.pt/~pjn/, http://islab.di.uminho.pt/ (P. Novais).
[1] Tel.: +351 253604435.
[2] Tel.: +351 253604466, +351 253604430; fax: +351 253604471.
[3] Tel.: + 351 253604437; fax: + 351 253604471.

http://dx.doi.org/10.1016/j.neucom.2015.03.105
0925-2312/© 2015 Elsevier B.V. All rights reserved.

Please cite this article as: A. Pimenta, et al., A neural network to classify fatigue from human–computer interaction, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.03.105

be considered as a priority in the healthcare domain, with organizations having a special responsibility in their human resources management strategies.

This paper deals with the issue of fatigue monitoring, with the aim of providing a non-intrusive, reliable and easy to use tool that can be used freely in organizations, without changing or interfering with the established work routines. Specifically, we look at desk jobs and others similar, in which people spend long hours interacting with the computer.

In preliminary work we have established that our interaction patterns with the computer, measured in terms of the use of the keyboard and mouse, change when under fatigue [11]. In this paper we take this work a step forward by presenting a service-based architecture, suited for the collection of data for the purpose of classifying fatigue. We detail the whole process, from the definition of the architecture to the real-time collection of data. Finally, we describe the process through which soft computing techniques are used to classify fatigue, in real-time, from the interaction with the computer. Specifically, we train a neural network that takes as input features that describe the user's interaction patterns and provides as output an estimation of the user's mental fatigue.

The paper is structured as follows. Following this introductory section we analyse related work in what concerns the analysis and detection of fatigue, both using subjective and objective measures. As we do this, we analyse existing approaches and compare them with the one proposed in this paper. The paper then moves on to describe the service-based architecture developed to implement the collection of data. The following section is devoted to a specific case-study, in a real environment, concerning the detection of fatigue. It details the process of data collection and analysis. Finally, the paper ends with the analysis and discussion of the results attained with this work, and with the concluding remarks and analysis of future directions for this line of research.

2. Related work

From the existing literature it is possible to conclude that there are three main types of approach to the problem of mental fatigue detection and monitoring. The first relies in quantifications of cognitive performance, often measured as the combination of the speed and success or accuracy of an individual when carrying out a task. The second approach consists on subjective ratings of mental workload and mental fatigue. Finally, physiological indicators can also reveal subtle but precise aspects of mental fatigue that are often not revealed by performance measures or subjective ratings.

2.1. Performance measures

Task performance may be defined as the relative competence of the individual's interpretation of task-associated stimuli, and the effectiveness of his responses on the task. Excessive mental and physical workload and very low workload (observed in monotonous tasks) are frequent scenarios in which performance decrements are observed. Perhaps the most widely used indicator of fatigue is the degradation of task performance over time or in response to changes in task demands. Welford [12] distinguished four types of performance changes that may occur, namely:

- impairment of sensory or perceptual functions;
- slowing of sensory-motor performance;
- irregularity of time; and
- the disruption of performance.

Performance decrement due to fatigue may be considered as the total quantity and/or quality of work performed as a function of time on task [6]. Cognitive performance is often measured as the combination of the accuracy per response time of an individual when carrying out a task. This combination is often called throughput [13]. Response speed is the inverse of response time and is more normally distributed that response time in data collected from fatigued operators. Both speed and accuracy are usually acceptable when the individual has normal cognitive skills, and one or the other or both are decreased when these skills are impaired or diminished by fatigue. When mental fatigue takes place, human performance is not consistent, with changes in speed, accuracy and variability [14]. These changes allow us to measure the performance of an individual on a task and are the base of the work proposed in this paper further ahead.

2.2. Subjective measures

Subjective measures of mental workload are often used to assess the mental workload associated with a task [15,16]. Obtaining mental workload levels during task performance can be a difficult procedure. However, estimations of subjective workload levels can help isolate task characteristics that affect performance. The two instruments of this type most often used in research were developed in parallel in the 1980s, one at the NASA-Ames Research Center in California and the other in the U.S. Air Force Human Factors Research Group in Wright-Patterson AFB, Ohio.

The NASA Task Load Index (NASA-TLX) is a multidimensional assessment tool [15]. The main seven-point scale is:

Overall performance: How successful were you in performing the task? How satisfied were you with your performance?

The TLX has five seven-point sub-scales that help identify difficult task characteristics. The sub-scales are as follows:

- *Mental demand*: How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?
- *Physical demand*: How much physical activity was required? Was the task easy or demanding, slack or strenuous?
- *Temporal demand*: How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?
- *Frustration level*: How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?
- *Effort*: How hard did you have to work (mentally and physically) to accomplish your level of performance?

The Subjective Workload Assessment Technique (SWAT) uses a three-dimensional model of workload: time load, mental effort load, and psychological stress load. There are three rating levels for each dimension: low, medium and high. Usually, subjects perform a pre-sort of the 27 possible combinations of the dimensions and their sort is transformed into an interval scale of workload ranging from 0 to 100. For the actual task, each activity or event is rated by assigning a value of 1–3 on each of the three dimensions. The pre-sort scale value associated with this combination becomes the workload value for that activity or event.

A seven-point mental workload scale was created by the Crew Performance Branch of the USAF School of Aerospace Medicine in the late 1970s, and then re-examined, linearized, and verified by the Human Factors Engineering Branch of the Air Force Flight Test Center [17]. This is the easiest rating instrument to use in high workload environments, but it provides less insight than the TLX or SWAT instruments. Individuals choose one of seven sets of statements describing their average mental workload during the preceding work period:

Please cite this article as: A. Pimenta, et al., A neural network to classify fatigue from human–computer interaction, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.03.105

- Nothing to do; no system demands.
- Light activity; minimum demands
- Moderate activity; easily managed; considerable spare time.
- Busy; challenging but manageable; adequate time available.
- Very busy; demanding to manage; barely time enough.
- Extremely busy; very difficult; non-essential tasks postponed.
- Overloaded; system unmanageable; essential tasks undone; unsafe.

Other tools exist to estimate mental fatigue, namely the vigor-activity and fatigue-inertia factors of the profile of mood states (POMS). The POMS measures five aspects of affect or mood [18,19]. It consists of 65 adjectives describing feelings and mood to which the subject responds according to a five-point scale ranging from "Not at all" to "Extremely".

Other similar tools exist, such as the Karolinska sleepiness scale (KSS) [20,21], the Stanford sleepiness scale (SSS) [22] or the Pittsburgh sleep quality index (PSQI).

2.3. Physiological measures

Several physiological variables can be used as indicators of fatigue. Electroencephalography (EEG) is a technique that has been used often for the detection of patterns of mental fatigue and also for the analysis of cognitive overload. Gevins has conducted EEG studies to analyse workload and fatigue [23,24]. Oculography and eye blink measures such as PERCLOS have been used for fatigue detection [25–27].

Wrist activity monitors (WAM) have been used to track sleep and sleep debt. The best-validated WAM is the ActiGraph, which uses the Cole–Kripke algorithm [28,29]. Other well-known physiological measures exist, such as heart rate (HR), heart rate variability (HRV) and the electrodermal response (EDR), that have however been shown to be somewhat unreliable for the detection of mental fatigue [30,31].

A relatively new discipline, called neuroergonomics, has arisen in the domain of physiological measurement [32]. Its goals parallel the goal of the performance-based approach that is the focus of this brief literature review. Neuroergonomics takes advantage of modern tools that provide information about brain functions, including EEG, magnetoencephalography and functional magnetic resonance imaging (fMRI).

Parasuraman and Wilson reviewed neuroergonomics studies in workload and vigilance, adaptive automation, neuroengineering, and molecular genetics and individual differences [33]. They concluded that the neuroergonomics approach can enrich our understanding of the use of technology by humans and can inform technological design. They suggested that applications should include the assessment of operator workload and vigilance, the implementation of real-time adaptive automation, neuroengineering for people with disabilities, and design of selection and training methods.

2.4. Critical analysis of current approaches

On the face of so many different approaches to the problem, one must enquire whether it makes sense to propose yet another one. In this exercise let us first recall the aims of our research line: we aim to develop an approach to fatigue detection and classification that can be used continuously and non-intrusively in real-life scenarios such as the workplace.

From this point of view, physiological approaches present a main problem, despite their accuracy: their continuous use in work environments is somewhat impractical, perhaps with the exception of monitoring eye and eyelid movements remotely (although image-based approaches are also frequently declined by workers on the grounds of violating privacy or promoting work surveillance). In general terms these approaches require trained personnel to be present to place sensors and other hardware, which is also often expensive. Taking the EEG as example, it requires the placement of scalp sensors that usually lead to significant limitation on the individual being monitored, especially in what concerns movements, besides the time it takes for an accurate placement. Moreover, there is another caveat, especially in work environments, that must also be stated: some individuals and some workers' unions object physiological measures in the workplace under the grounds of violating privacy. They label them as "medical" measures, despite some of them not possessing medical precision, and claim that they collect information that can be used for many other purposes.

Concerning subjective instruments, their main problem is that they do not fully take into account inter-individual differences. Although there are some that help account for individual differences, such as the profile of mood states (POMS) and the State-Trait Anxiety Inventory (STAI) [34,35], their use in complex systems can prove to be complicated and confusing, due to the same problems that can be observed in subjective measures of fatigue detection [36–38]. Subjective measurements also have the problem of being made through questionnaires that are filled by the user himself, interfering with the primary task, asides from the traditional problems related to wording, question construction and interpretation or answer biasing.

The use of the decrement of performance as an indicator of mental fatigue is also not without criticism, the main issue being the inexactness in accounting for the decrement. Fatigue effects may not be disclosed when subjects compensate by exerting greater effort. Usually, substantial or excessive time on task is required to induce task-specific fatigue or acute fatigue. However, a very high workload can cause this same effect in less time. In both cases there are excessive task demands during the time required to perform a task. Excessive workload conditions can be considered as accelerating the build-up of fatigue, while normal workload results in a slower emergence of fatigue.

Performance decrement is also often not observed, even though performance requirements or workload are increased. This phenomenon may be explained by the concept of spare capacity, a concept developed originally for the assessment and modelling of humans as drivers and pilots [39,40]. The theory of spare capacity proposes that man rarely works at their maximal work capacity, and can thus temporarily absorb additional cognitive or physical workload while maintaining a standard level of performance, given that his spare capacity is not exceeded. Thus, the increase of fatigue is often viewed as decreasing in individual's spare capacity, and the concepts of workload and fatigue are considered to be interconnected.

The strengths and weaknesses pointed out and summarized in Table 1 motivate and support the development of the approach proposed in this paper: one that is non-intrusive and can thus be used continuously throughout the day. It will improve access to this kind of initiatives, with a positive impact in individual health, absenteeism and related costs, productivity and quality of work.

3. An architecture for distributed data acquisition

Prior to carrying out the research work aimed at understanding the influence of fatigue on the interaction with the computer, a suitable framework for data acquisition was idealized and developed. We clearly aimed at a framework in line with the ambient intelligence (AmI) philosophy [41], in which the technological component is hidden in the environment and the user is placed in the middle of the paradigm. There is also a focus on non-

Please cite this article as: A. Pimenta, et al., A neural network to classify fatigue from human–computer interaction, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.03.105

ARTICLE IN PRESS

Table 1
Summary of the strengths and weaknesses of the three main types of approaches.

Type	Strengths	Weaknesses
Performance	– Can be acquired through non-invasive methods –Strongly related to fatigue	–Sensitive to spare capacity effect
Subjective	–Easy and inexpensive to implement –Many validated questionnaires and other tools available	–Based on the individual's subjective perceptions –Interferes with work routine
Physiological	–Very accurate –Impossible to fake	–Expensive and complex approach –Often requires specific hardware and trained personnel –Impractical and intrusive

intrusiveness, with acquisition of information taking place without the need for explicit or conscious user interactions.

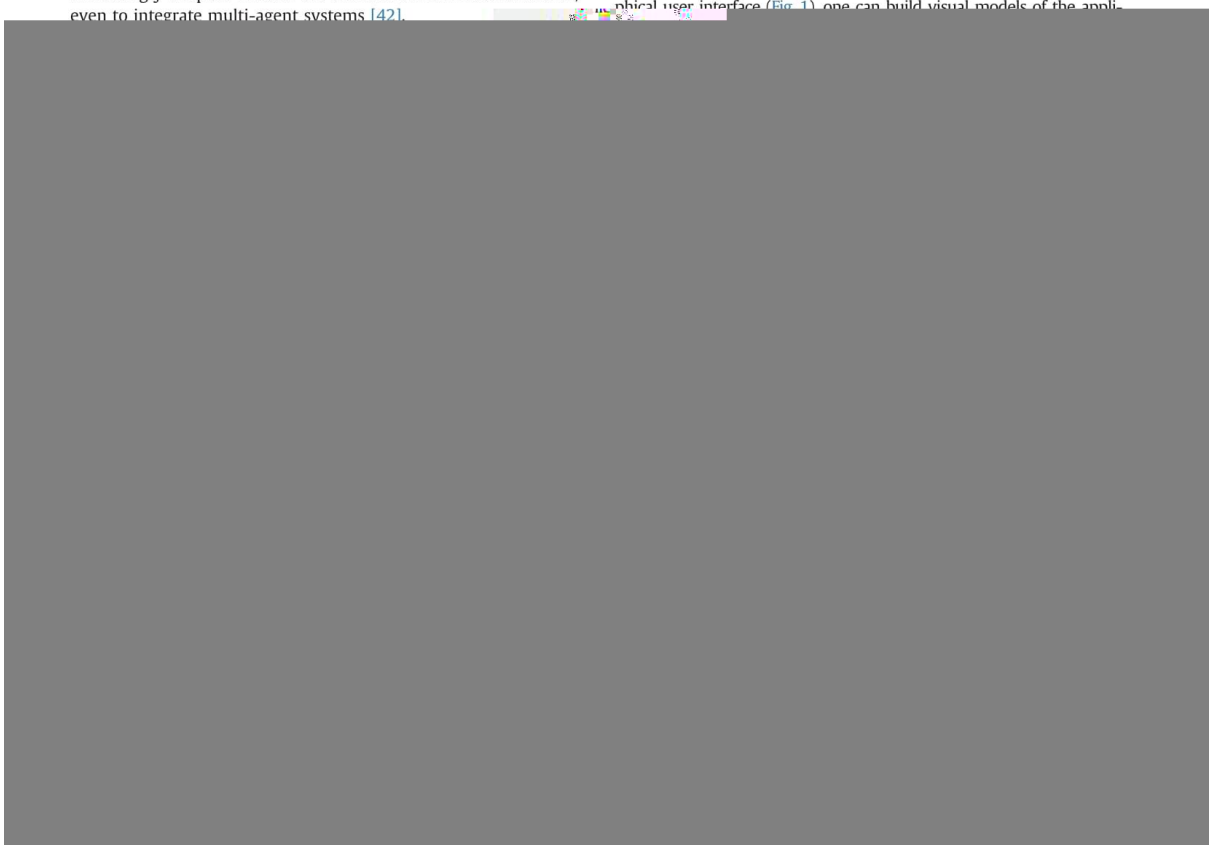
AmI systems are commonly described as electronic environments that seamlessly interact and adapt to human needs, in which people are surrounded by intelligent and intuitive interfaces embedded in all kinds of objects. To take full advantage of the information gathered ubiquitously from various sources in the environment there is a need for a software infrastructure that allows an easy integration, promotes interoperability, and focuses on extensibility. Considering these aspects, in this work we present an infrastructure to support an efficient approach for the development of AmI applications, following an approach based on service oriented architectures (SOAs). Indeed, SOAs are being increasingly adopted in both the academic and industrial arenas, even to integrate multi-agent systems [42].

debugged independently of the other components. This not only speeds up project implementation but, in the case of well-designed service components, also leads to significantly enhanced reuse of components.

- *Dynamic deployment and runtime modification/replacement.* Service components can be dynamically deployed to remote nodes at runtime, and components within a process can be easily replaced by new or updated components, further reducing the time taken to modify or change an existing process in response to business requirements.
- *Configuration management and version control.* Service components facilitate version control and dynamic configuration management, allowing fine-grained control over deployments across the enterprise.

SCA provides a good basis for AmI applications [43], it is in line with our architectural model and it fulfils major AmI deployment requirements by promoting late bindings at deploy time and runtime with the support of several relevant technologies including POJO, SOAP, REST, BPMN, BPEL, JMS, Camel or Rules services. But most of all it is currently supported by several major commercial and open source products such as Jboss Switchyard, IBM WebSphere or TRENTINO (C++).

From the several available implementations of SCA we have chosen JBoss SwitchYard since it is an open source solution in a relative mature state, and also enhances some of the SCA advantages. Specifically, SwitchYard advocates transparency when running a service during its whole lifecycle. Important aspects such as connectivity, orchestration and routing do exist in SwitchYard in a modular format, which means one can deploy them in an independent way. Using a SwitchYard graphical user interface (Fig. 1), one can build visual models of the appli-



ARTICLE IN PRESS



neucom.2015.03.105 http://dx.doi.org/10.1016/j.

ARTICLE IN PRESS

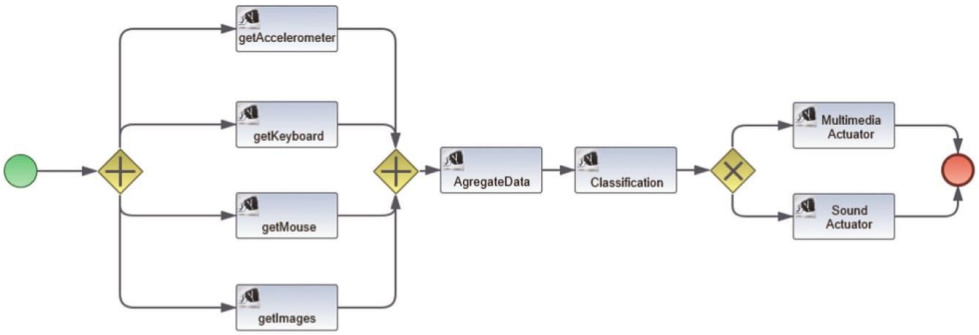


Fig. 3. The "camcof" process, viewed through BPMN.

- *Accelerometers* – Accelerometers placed on chairs, keyboards and mouses, measure how users are moving and how they are using their peripherals.
- *Mouse and keyboard* – These devices provide information about how the user interacts with the peripherals (e.g. velocity of the mouse, typing rhythm, and number of mistakes).
- *Microphone* – Microphones are used to measure the amount of noise in the vicinity of the user, allowing to perceive their social environment.

- To determine if mental fatigue does have a significant effect on the considered features.
- To assess the possibility of training a neural network to classify mental fatigue in real time.

Given these objectives, the main research question of the study is defined as follows: mental fatigue affects the interaction of an individual with the computer, and this can be accurately measured with a non-intrusive approach.

4.2. Data collection

For this particular study, and given the aforementioned objectives, we focused on the data collected from the mouse and the keyboard of the user's computers. The interaction of the user with the computer is monitored in terms of specific operating system events fired from the use of the computer's mouse and keyboard:

- MOV, timestamp, posX, posY – an event describing the movement of the mouse, in a given time, to coordinates (posX, posY) in the screen.
- MOUSE_DOWN, timestamp, [Left|Right], posX, posY – this event describes the first half of a click (when the mouse button is pressed down), in a given time. It also describes which of the buttons was pressed (left or right) and the position of the mouse in that instant.
- MOUSE_UP, timestamp, [Left|Right], posX, posY – an event similar to the previous one but describing the second part of the click, when the mouse button is released.
- MOUSE_WHEEL, timestamp, dif – this event describes a mouse wheel scroll of amount dif, in a given time.
- KEY_DOWN, timestamp, key – identifies a given key from the keyboard being pressed down, at a given time.
- KEY_UP, timestamp, key – describes the release of a given key from the keyboard, in a given time.

From these events, that fully describe the interaction of the user with the mouse and keyboard, we extract a set of features, based on notions of behavioural biometrics, that will be used to monitor fatigue:

Key down time (KDT) – the timespan between two consecutive KEY_DOWN and KEY_UP events, i.e., for how long was a given key pressed.

Time between keys (TBK) – the timespan between two consecutive KEY_UP and KEY_DOWN events, i.e., how long did the individual took to press another key.

Mouse velocity (MV) – the distance travelled by the mouse (in pixels) over the time (in milliseconds). The velocity is computed for each interval defined by two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively, in the coordinates $(x1, y1)$ and $(x2, y2)$, that took place, respectively, in the instants $time_1$ and $time_2$. Let us also assume two vectors *posx* and *posy*, of size *n*, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The velocity between the two clicks is given by $r_dist/(time_2 - time_1)$, in which *r*_dist represents the distance travelled by the mouse and is given by the following equation:

$$r_dist = \sum_{i=0}^{n-1} \sqrt{(posx_{i+1} - posx_i)^2 + (posy_{i+1} - posy_i)^2} \tag{1}$$

Mouse acceleration (MA) – the velocity of the mouse (in pixels/milliseconds) over the time (in milliseconds). A value of acceleration is computed for each interval defined by two consecutive MOUSE_UP and MOUSE_DOWN events, using the intervals and data computed for the Velocity.

Time between clicks (TBC) – the timespan between two consecutive MOUSE_UP and MOUSE_DOWN events, i.e., how long did it took the individual to perform another click.

Double click duration (DCD) – the timespan between two consecutive MOUSE_UP events, whenever this timespan is inferior to 200 ms. Wider timespans are not considered double clicks.

Average excess of distance (AED) – this feature measures the average excess of distance that the mouse travelled between each two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively, in the coordinates $(x1, y1)$ and $(x2, y2)$. To compute this feature, first it is measured the distance in straight line between the coordinates of *mup* and *mdo* as $s_dist = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$. Then, the distance actually travelled by the mouse by summing the distance between each two consecutive MOUSE_MV events is measured. Let us assume two vectors *posx* and *posy*, of size *n*, holding the coordinates of the consecutive MOUSE_MV events between *mup* and *mdo*. The distance actually travelled by the mouse, *real*_dist is given by Eq. (1). The average excess of distance between the two consecutive clicks is thus given by r_dist/s_dist.

Average distance of the mouse to the straight line (ADMSL) – in a few words, this feature measures the average distance of the mouse to the straight line defined between two consecutive clicks. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively, in the coordinates $(x1, y1)$ and $(x2, y2)$. Let us also assume two vectors *posx* and *posy*, of size *n*, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The sum of the distances between each position and the straight line defined by the points $(x1, y1)$ and $(x2, y2)$ is given by 2, in which *ptLineDist* returns the distance between the specified point and the closest point on the infinitely-extended line defined by $(x1, y1)$ and $(x2, y2)$. The average distance of the mouse to the straight line defined by two consecutive clicks is this given by

$$s_dists = \sum_{i=0}^{n-1} ptLineDist(posx_i, posy_i) \tag{2}$$

Distance of the mouse to the straight line (DMSL) – this feature is similar to the previous one in the sense that it will compute the *s_dists* between two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, according to Eq. (2). However, it returns this sum rather than the average value during the path.

Signed Sum of Angles (SSA) – with this feature the aim is to determine if the movement of the mouse tends to "turn" more to the right or to the left. Let us assume three consecutive MOUSE_MOVE events, *mov*1, *mov*2 and *mov*3, respectively, in the coordinates $(x1, y1)$, $(x2, y2)$ and $(x3, y3)$. The angle α between the first line (defined by $(x1, y1)$ and $(x2, y2)$) and the second line (defined by $(x2, y2)$ and $(x3, y3)$) is given by $degree(x1, y1, x2, y2, x3, y3) = \tan(y3 - y2, x3 - x2) - \tan(y2 - y1, x2 - x1)$. Let us now assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*. Let us also assume two vectors *posx* and *posy*, of size *n*, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The signed sum of angles between these two clicks is given by the following equation:

$$rCls_angle = \sum_{i=0}^{n-2} degree(posx_i, posy_i, posx_{i+1}, posy_{i+1}, posx_{i+2}, posy_{i+2}) \tag{3}$$

Absolute sum of angles (ASA) – this feature is very similar to the previous one. However, it seeks to find only how much the mouse "turned", independently of the direction to which it turned. In that sense, the only difference is the use of the absolute of the value returned by function $degree(x1, y1, x2, y2, x3, y3)$, as depicted in the following equation:

$$rCls_angle = \sum_{i=0}^{n-2} |degree(posx_i, posy_i, posx_{i+1}, posy_{i+1}, posx_{i+2}, posy_{i+2})| \tag{4}$$

Distance between clicks (DBC) – represents the total distance travelled by the mouse between two consecutive clicks, i.e., between each two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively, in the coordinates $(x1, y1)$ and $(x2, y2)$. Let us also assume two vectors *posx* and *posy*, of size *n*, holding the coordinates of the consecutive MOUSE_MOV

Please cite this article as: A. Pimenta, et al., A neural network to classify fatigue from human–computer interaction, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.03.105

ARTICLE IN PRESS

Table 2
Analysis of significance of the collected data. Column a depicts the average value of the Mann Whitney test. Column b depicts the percentage of individuals for which the differences observed in each feature when fatigued and rested were statistically significant.

Features	a	b (%)
Distance of the mouse to the straight line	0.022	79
Key down time	0.021	75
Mouse acceleration	0.050	67
Average distance of the mouse to the straight line	0.042	63
Mouse velocity	0.023	63
Average excess of distance	0.041	54
Time between clicks	0.020	54
Distance during clicks	0.026	54
Time between keys	0.271	46
Total excess of distance	0.258	46
Double click duration	0.003	33
Absolute sum of angles	0.336	25
Signed sum of angles	0.706	21
Distance between clicks	0.0238	20

events between *mup* and *mdo*. The total distance travelled by the mouse is given by Eq. (1).

4.3. Data analysis

Prior to the training of the ANN we conducted an analysis of the data to determine if there were statistically significant differences in the features when comparing data collected in different levels of fatigue.

To achieve this we divided the data collected in two groups: one containing the data collected while the users claimed to feel fatigued (values 1–3 of the USAFAM scale) and the other containing the remaining of the data, collected when users claimed to feel rested.

After this, we used the Mann–Whitney test to compare the distributions of the values of each feature in order to verify that there were statistically significant difference between the two groups of data (Table 2). The following features have shown the most statistically significant differences: key down time, mouse velocity, mouse acceleration, time between clicks, distance of the mouse to the straight line, distance during clicks, average excess of distance and average distance of the mouse to the straight line.

At this point, one of the concerns was to determine if the observed differences were not due to natural variations of the use of the peripherals. To this end we conducted an additional analysis, as follows. We took each group of data (fatigued versus rested) and divided each group into two subgroups of equal size, with the members of each group being selected randomly. We thus got four groups, two containing data collected when individuals were fatigued and other two when individuals were rested. We then used the Mann Whitney test to determine if there were differences between the pairs of subgroups that were extracted from the same group. We repeated this process 100 times and computed the average value of the Mann Whitney test. Table 3 summarizes these results. Only in the key down time feature were there found statistically significant differences, in average, and only when comparing the two rested subgroups. Given this, it is possible to conclude that the observed differences can indeed be attributed to consistently different behaviours that are related to the level of fatigue and not to natural variations.

However, and despite the observed differences, not every user revealed the same behaviours or trends when under fatigue. The average distance of the pointer to the line between clicks is an example of a metric in which the average values increased in half of the users and decreased in the other half, despite the statistically significant differences observed. This points out the importance of developing individual models, trained with data from each user. It also points out the subjectiveness of the effects of fatigue, which we had already previously analysed in the detection of stress following similar approaches [45].

Table 3
Results of testing the intra-group differences by subdividing each group into two subgroups, randomly, and comparing them. Features are represented by their initials.

Features	Average *p*-value (rested)	Average *p*-value (tired)
KDT	0.086	0.024
MV	0.196	0.773
MA	0.211	0.610
AED	0.515	0.994
TBC	0.413	0.098
DMSL	0.669	0.531
DDC	0.983	0.095
ADMSL	0.539	0.787

Table 4
The values of centrality and variance for each feature, represented by their initials, and each group of data. The three last columns represent the following trends: A – trend of the medians when comparing rested with fatigued; B – trend of the means when comparing rested with fatigued; and C – trend of the variance when comparing rested with fatigued. In these columns, a value of +70% should be interpreted as "increases in 70% of the cases".

Features	State	Variance	Median	Mean	A (%)	B (%)	C (%)
KDT	Rested	1177.6	121.8	139.1	+100	+50	+100
	Fatigued	4994.2	116.8	138.8			
MV	Rested	0.3710	0.182	0.191	+75	+50	+100
Fatigued	0.7344	0.188	0.199				
MA	Rested	0.4128	0.740	0.801	+60	−60	+90
Fatigued	0.7520	0.786	0.844				
TBC	Rested	1.3342	3539.0	3925.0	−51	−51	+100
Fatigued	3.2017	2794.0	2913.0				
DMSL	Rested	62 620.7	4232.0	4047.0	−50	−59	+77
Fatigued	39 489.9	3293.0	4366.0				
DDC	Rested	28 233.7	116.3	148.00	+77	+67	+77
Fatigued	55 839.0	87.99	93.5				
AFD	Rested	1.8564	2.080	1.991	+50	+60	+90
Fatigued	5.9635	1.743	1.818				
ADMSL	Rested	7765.5	28.100	29.5	−60	−50	−60
Fatigued	4311.1	27.040	30.5				

Given this, and following similar approaches [31,14], we centred the analysis in measures of variability, in an attempt to determine whether the variability of the two groups of data was different. We concluded that variability was significantly smaller in all features except for the average distance of the mouse to the straight line (Fig. 8) when users were rested, revealing that users were more consistent in the use of the peripherals not fatigued (Table 4). As we can see, the differences in the values are much more consistent in the case of variability (i.e., nearly all the features follow an increasing tendency with the onset of fatigue) than in the measures of centrality. Hence the decision of using variability for training the neural network.

These differences are also graphically visible. Figs. 4–8 depict how the differences in the distributions of the variance (rightmost images) are larger than the differences in the distributions of the average values.

Please cite this article as: A. Pimenta, et al., A neural network to classify fatigue from human–computer interaction, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.03.105

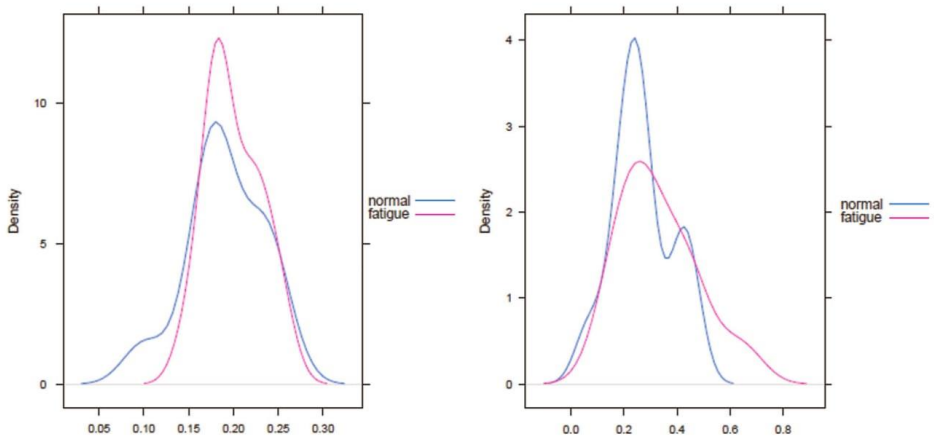
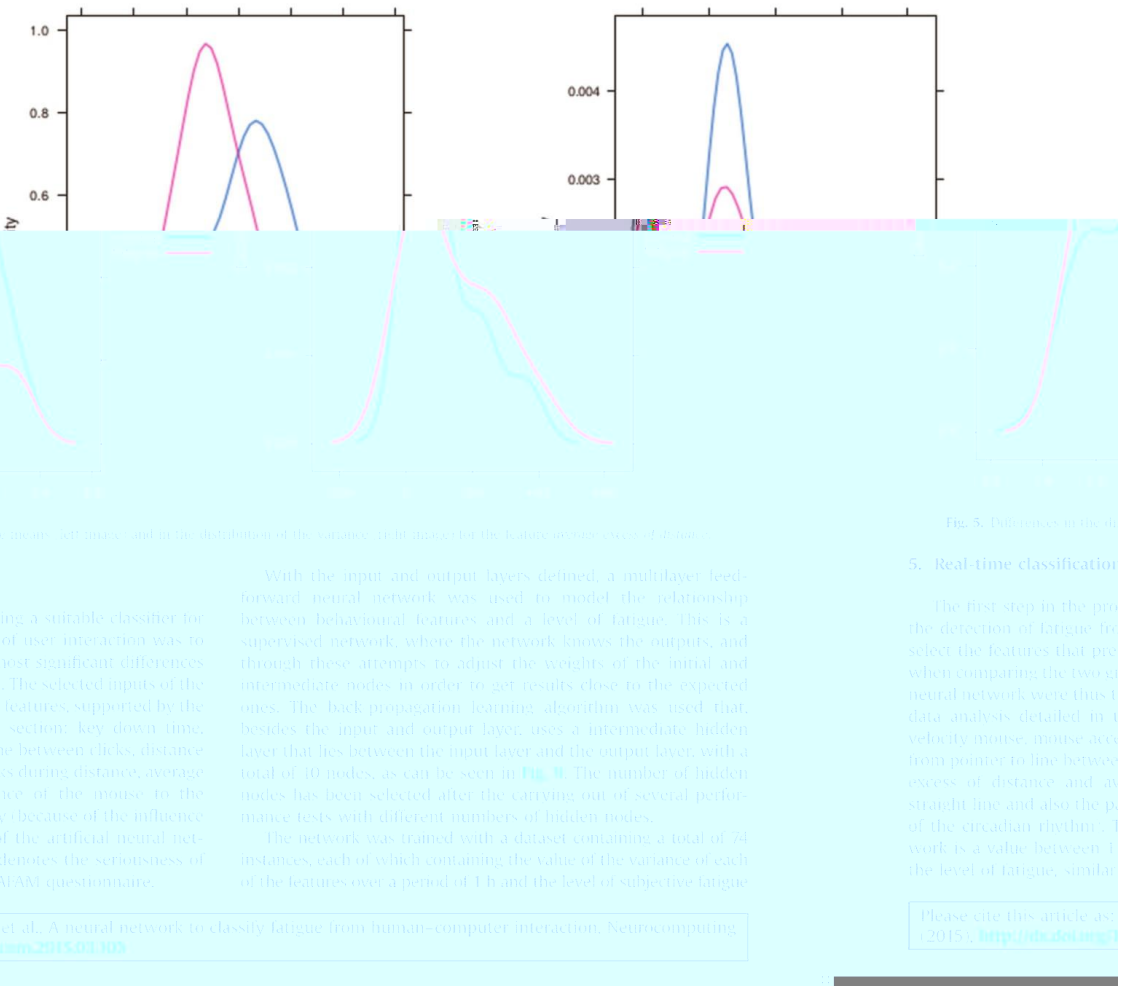


Fig. 4. Differences in the distribution of the means (left image) and in the distribution of the variance (right image) for the feature *mouse velocity*.



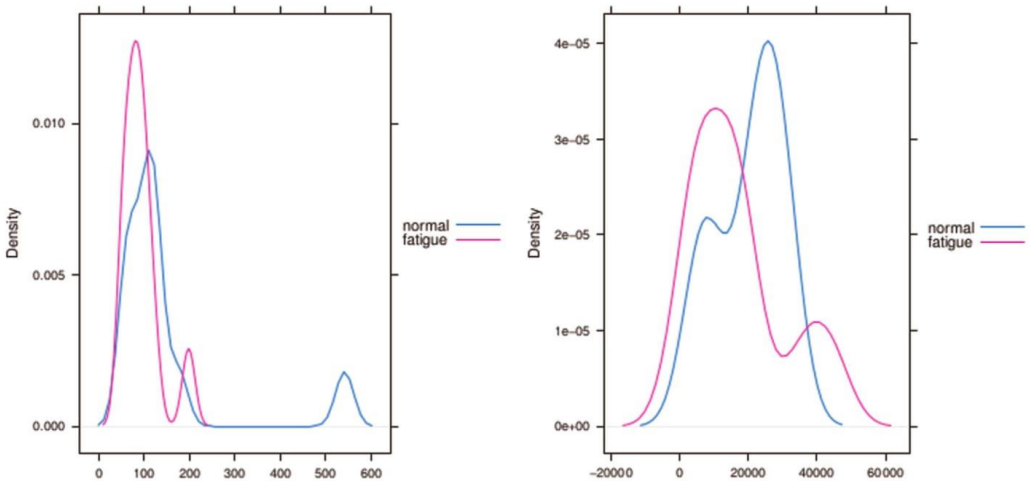


Fig. 6. Differences in the distribution of the means (left image) and in the distribution of the variance (right image) for the feature *distance during clicks*.

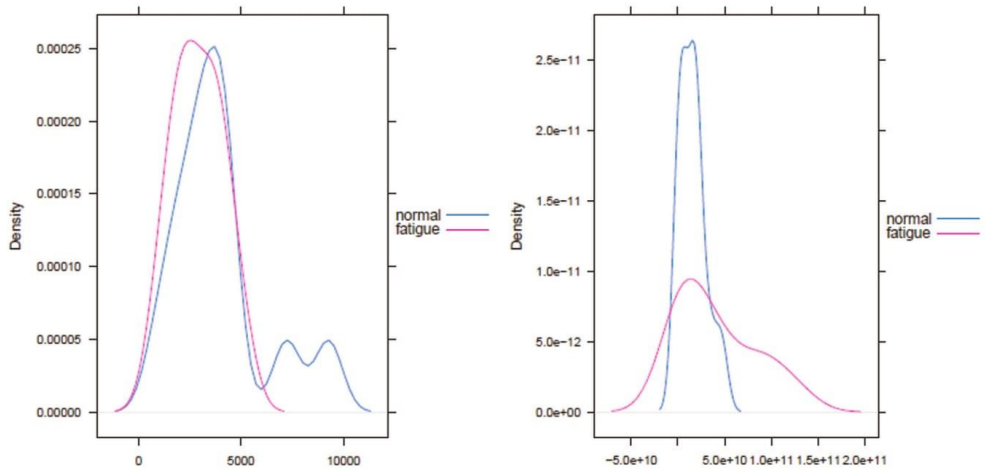


Fig. 7. Differences in the distribution of the means (left image) and in the distribution of the variance (right image) for the feature *time between clicks*.

provided by the users through the questionnaire, for that period. The network was trained during 125 iterations where it reached a minimum RMSE (root-mean-square error) as can be seen in Fig. 10.

The relative importance of the variables used as input was assessed through a method proposed by Garson [46]. This method identifies the relative importance of input variables to the final output on supervised neural network using the weights of each node of the model. The relative importance (or the strength of association) of an input variable to the output can be determined by identifying all connections between nodes' weighted interest. That is, all the weights of specific binding of the input node which pass through to the hidden layer output variable are identified. This process is repeated for all other input variables to obtain a list of all weights which are specific to each input variable. The connections are computed for each node of the input layer and balanced against all other entries. A single value is obtained for each input variable that describes the relationship with the model output.

Fig. 11 shows that x5 and x6 (mouse velocity and acceleration, respectively) have the strongest positive relationships, while the time between clicks and key down time have more negative relationships with the output. It is also visible that the shift/turn (morning, afternoon, and evening) in which the data was collected has little influence on the output, which indicates that the time of the day is less important than the other features when it comes to classify fatigue.

5.1. Results

The ANN trained and outlined above was tested and validated with data from the second week of the data collection period, in which each instance, as described above, contained the variance of each feature in the period of an hour and the subjective measure of the level of fatigue

Please cite this article as: A. Pimenta, et al., A neural network to classify fatigue from human–computer interaction, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.03.105

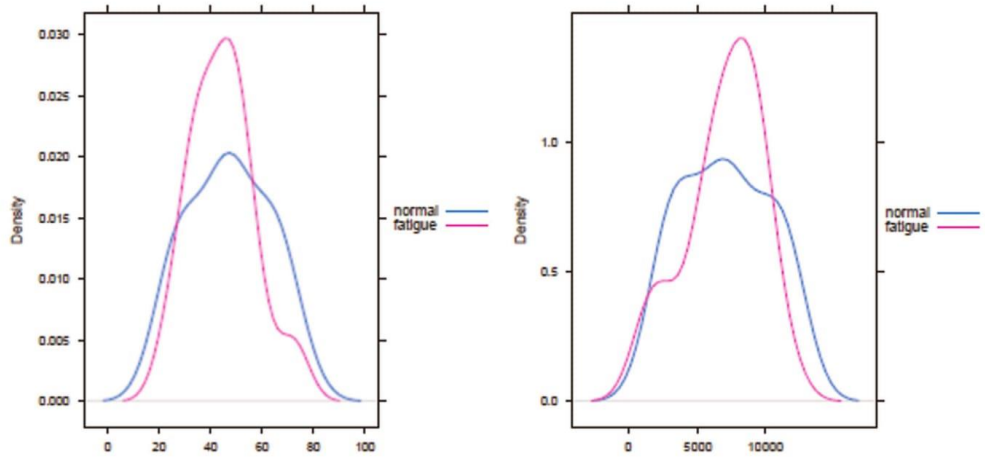
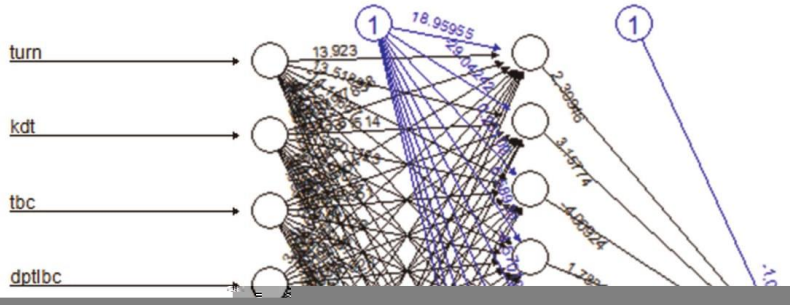


Fig. 8. Differences in the distribution of the means (left image) and in the distribution of the variance (right image) for the feature *average distance of the mouse to the straight line.*



ARTICLE IN PRESS

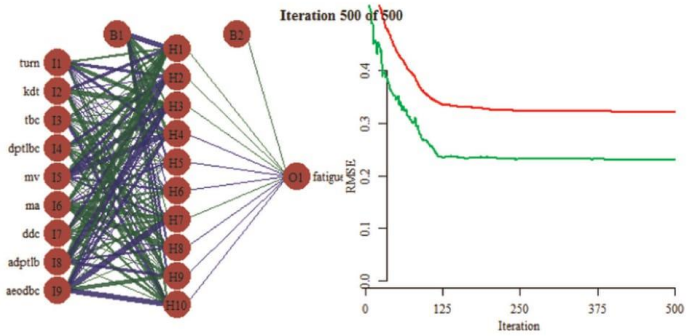


Fig. 10. Evolution of the RMSE during the evolution of the training of the ANN (green line) and during validation (red line). (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

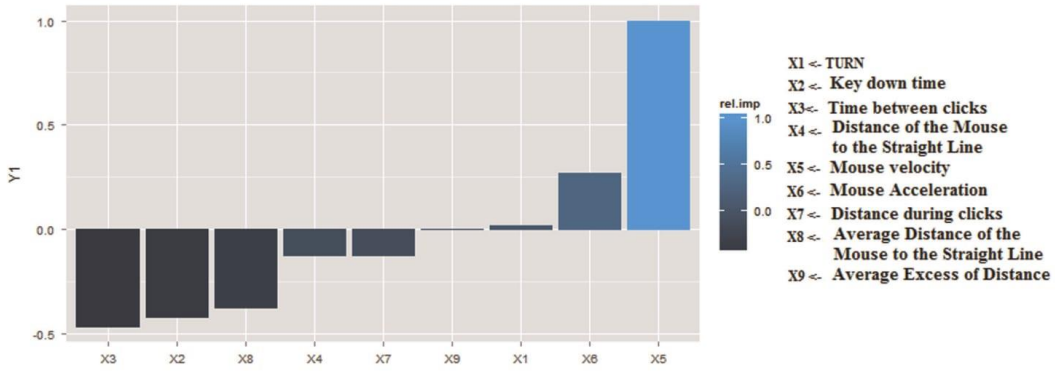


Fig. 11. Relative importance of different variables.

Table 5
Results of the validation of the ANN. 81% of the instances were correctly classified (bold). The 19% misclassified instances where nonetheless classified as neighbouring values (italic).

	PREDICTED							
		1	2	3	4	5	6	7
U	1	**9**	*2*	0	0	0	0	0
S	2	*3*	**13**	*3*	0	0	0	0
A	3	0	*2*	**15**	0	0	0	0
F	4	0	0	*1*	**8**	*1*	0	0
S	5	0	0	0	0	**8**	*1*	0
A	6	0	0	0	0	*1*	**6**	0
M	7	0	0	0	0	0	0	**1**

that can be used in real-time, without interfering in any way with the worker's routine. To this end, we developed an approach based on the monitoring of the user's performance. The main conclusion, as expected, is that when the users claim to be fatigued, their performance decreases. This is evidenced by less efficient interaction patterns (e.g. longer mouse clicks, larger distances travelled by the mouse, and longer key down times). This conclusion, which comes as no surprise, is nonetheless important in the sense that it allows, for the first time, to quantify and study these differences.

As analysed previously, there are some potential drawbacks associated to this kind of performance-based approaches, namely the difficulty in precisely accounting for changes in performance (e.g. they may not be entirely due to mental fatigue). Moreover, measuring performance is not always easy. The type of task being performed may significantly influence the interaction patterns. The workload is another issue to consider. Task standardization, which would to some extent solve these drawbacks, may be difficult to achieve. The motivation of the worker or feelings of boredom will also influence performance.

To address these issues, we are also working towards classifying the type of task (e.g. text typing and web browsing) from the interaction patterns, identifying workload and quantifying the level of task attention [47]. This kind of information, which to some extent describes the user's context, will enable the development of more accurate classifiers.

ARTICLE IN PRESS

Nonetheless, and despite these efforts, the accuracy of the approach will, evidently, never match that of other approaches such as the ones based on physiological sensors. But then again, that was not the main aim from the beginning.

The main aim was to develop an approach that could be used realistically in a common workplace environment, especially in those workplaces in which workers spend long hours using a computer. To this end, this research line accomplishes its objectives, as the results detailed above show.

Given that different effects due to fatigue could be found in different users, we are now working towards an increased personalization, in which each user has an ANN trained with their own data. Moreover, we are also working towards the validation of the ANN using electroencephalography in parallel with our approach, through a cooperation with researchers of the University of Groningen.

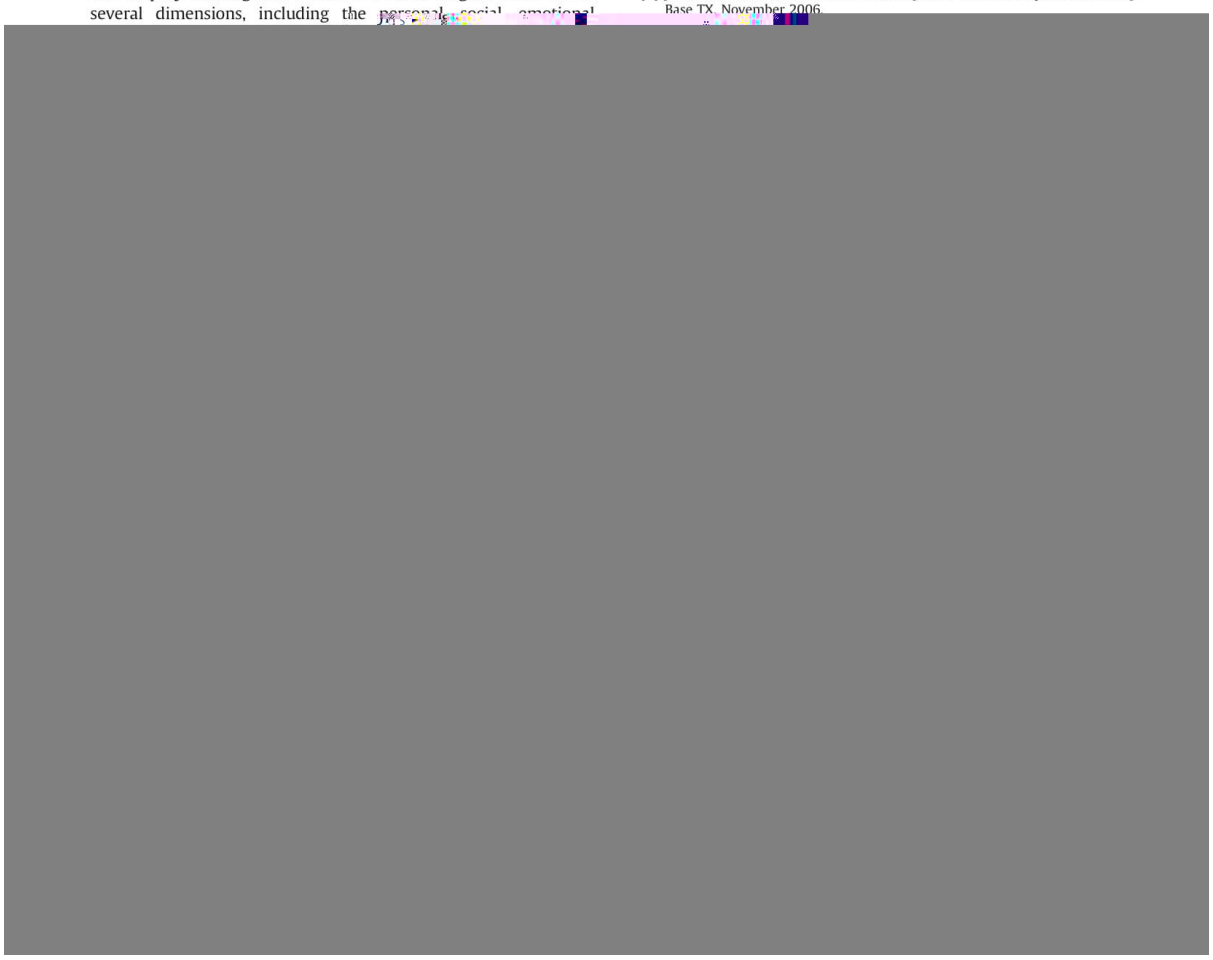
7. Conclusion

Fatigue can nowadays be seen as a major health issue. Mental fatigue is especially on the rise given the dramatic change in the labour market, from mostly physical jobs to very mentally demanding ones. Harsher work conditions, increased competitiveness or high unemployment add to the pressure of today's worker and amplify the negative effects of mental fatigue. These act on several dimensions, including the personal, social, emotional

programme for competitiveness) and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within projects FCOMP-01-0124-FEDER-028980 (PTDC/EEISII/1386/2012) and within the Project Scope UID/CEC/00319/2013.

References

[1] M.M. Mitler, J.C. Miller, Some practical considerations and policy implications of studies of sleep patterns, Behav. Med. 21 (4) (1996) 184–185.
[2] J.C. Miller, Fundamentals of shiftwork scheduling, Ergon. Des.: Q. Hum. Factors Appl. 16 (3) (2008) 13–17.
[3] S. Folkard, P. Tucker, Shift work, safety and productivity, Occup. Med. 53 (2) (2003) 95–101.
[4] D. van der Linden, P. Eling, Mental fatigue disturbs local processing more than global processing, Psychol. Res. 70 (5) (2006) 395–402, http://dx.doi.org/10.1007/s00426-005-0228-7, URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/15968553⟩.
[5] R.J. Williamson, S. Purcell, A. Sterne, S. Wessely, M. Hotopf, A. Farmer, P. C. Sham, The relationship of fatigue to mental and physical health in a community sample, Soc. Psychiatry Psychiatr. Epidemiol. 40 (2) (2005) 126–132, http://dx.doi.org/10.1007/s00127-005-0858-5, URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/15685404⟩.
[6] L. Perelli, Fatigue Stressors in Simulated Long-Duration Flight. Effects on Performance, Information Processing, Subjective Fatigue, and Physiological Cost, March 1977. URL ⟨http://www.oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA105484⟩.
[7] M.H. Bonnet, Sleep, performance and mood after the energy-expenditure equivalent of 40 hours of sleep deprivation, Psychophysiology 17 (1) (1980) 56–63.
[8] J.C. Miller, In Search of Circasemidian Rhythms, Technical Report, Brooks City-Base TX, November 2006.



[27] L.N. Orchard, J.A. Stern, Blinks as an index of cognitive activity during reading, Integr. Physiol. Behav. Sci.: Off. J. Pavlov. Soc. 26(2), 1991, 108–116 URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/1878317⟩.
[28] G. Jean-Louis, H.V. Gizycki, F. Zizi, A. Spielman, P. Hauri, H. Taub, The actigraph data analysis software: I. A novel approach to scoring and interpreting sleep–wake activity, Percept. Motor Skills 85 (1) (1997) 207–216, http://dx.doi.org/10.2466/pms.1997.85.1.207, URL ⟨http://www.amsciepub.com/doi/abs/10.2466/pms.1997.85.1.207?journalCode=pms⟩.
[29] N.K. Tang, A.G. Harvey, Correcting distorted perception of sleep in insomnia: a novel behavioural experiment? Behav. Res. Ther. 42 (1) (2004) 27–39, http://dx.doi.org/10.1016/S0005-7967(03)00068-8, URL ⟨http://www.linkinghub.elsevier.com/retrieve/pii/S0005796703000688⟩.
[30] M. Stork, Some methods systems and sensors which are possible for driver's drowsiness estimation, in: WSEAS.US, pp. 87–94 URL ⟨http://www.wseas.us/e-library/conferences/2012/Barcelona/CSCS/CSCS-14.pdf⟩.
[31] S.C. Segerstrom, L.S. Nes, Heart rate variability reflects self-regulatory strength, effort, and fatigue, Psychol. Sci. 18 (3) (2007) 275–281, http://dx.doi.org/10.1111/j.1467-9280.2007.01888.x, URL ⟨http://www.pss.sagepub.com/content/18/3/275.short⟩.
[32] C. Wickens, J. McCarley, Applied Attention Theory, CRC Press, Boca Raton, FL, 2007 URL ⟨http://www.prometei.de/fileadmin/prometei.de/veranstaltungen/2008-05-28-Wickens_AppliedAttentionTheory.pdf⟩.
[33] R. Parasuraman, G. Wilson, Putting the brain to work: neuroergonomics past, present, and future, J. Hum. Factors 50(3) (2008) 468–474 URL ⟨http://www.hfs.sagepub.com/content/50/3/468.short⟩.
[34] C. Spielberger, R. Gorsuch, R. Lushene, Manual for the State-Trait Anxiety Inventory URL ⟨http://www.ubir.buffalo.edu/xmlui/handle/10477/2895⟩.
[35] C. Spielberger, R. Gorsuch, R. Lushene, Self-Evaluation Questionnaire: STAI Form X-1 URL ⟨http://www.ubir.buffalo.edu/xmlui/handle/10477/1878⟩.
[36] S. Curran, Short form of the profile of mood states (POMS-SF): psychometric information, Psychol. Assess. 7 (1) (1995) 80–83, http://dx.doi.org/10.1037//1040-3590.7.1.80, URL ⟨http://www.doi.apa.org/getdoi.cfm?doi=10.1037/1040-3590.7.1.80⟩.
[37] K. Millar, M. Jelicic, B. Bonke, A.J. Asbury, Assessment of preoperative anxiety: comparison of measures in patients awaiting surgery for breast cancer, Br. J. Anaesth. 74 (2) (1995) 180–183 , URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/7696068⟩.
[38] P. Newcombe, G. Boyle, High school students' sports personalities: variations across participation level, gender, type of sport, and success, Int. J. of Sport Psychlo. (1995) 277–294 , URL ⟨http://www.epublications.bond.edu.au/hss_pubs/802/⟩.
[39] J. Casali, W. Wierwille, A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications, J. Hum. Factors 25 (6) (1983) 623–641 , URL ⟨http://www.hfs.sagepub.com/content/25/6/623.short⟩.
[40] W. Wierwille, J. Gutmann, Secondary task measurement of workload as a function of simulated vehicle dynamics and driving conditions, J. Hum. Factors 6 (19) (1977) 557–565 , URL ⟨http://hfs.sagepub.com/content/19/6/557.short⟩.
[41] C. Ramos, J.C. Augusto, D. Shapiro, Ambient intelligence—the next step for artificial intelligence, IEEE Intell. Syst. 23 (2) (2008) 15–18.
[42] J. Agüero, M. Rebollo, C. Carrascosa, V. Julián, Mdd-approach for developing pervasive systems based on service-oriented multi-agent systems, ADCAIJ: Adv. Distrib. Comput. Artif. Intell. J. 1 (6) (2013) 55–64.
[43] P. Giner, V. Pelechano, An architecture to automate ambient business system development, in: Ambient Intelligence, Springer, Berlin, 2008, pp. 240–257.
[44] A. Woodrow, Cognitive performance research at Brooks Air Force Base, Texas, 1960–2009, Aviat. Space Environ. Med.
[45] D. Carneiro, P. Novais, J. Neves, Conflict Resolution and its Context: From the Analysis of Behavioural Patterns to Efficient Decision-Making, vol. 18, 2014, Springer International Publishing, Switzerland.
[46] G.D. Garson, Interpreting neural-network connection weights, AI Expert 6 (4) (1991) 46–51.
[47] A. Pimenta, D. Carneiro, P. Novais, J. Neves, Detection of distraction and fatigue in groups through the analysis of interaction patterns with computers, In < i > Intelligent Distributed Computing VIII < /i > , Springer International Publishing, 2015 (pp. 29-39).



André Pimenta is a researcher at the CCTC (Computer Science and Technology Center), in the Department of Informatics, University of Minho, Braga, Portugal. He is currently enrolled in a doctoral programme in Computer Science in the University of Minho. He develops scientific research in the field of Artificial Intelligence and Machine Learning. Specifically, he steers his research work towards the development of noninvasive approaches for the classification of fatigue. He is the author of several papers in proceedings of conferences the fields of Soft Computing and Machine Learning.



Davide Carneiro is a researcher at the CCTC (Computer Science and Technology Center), in the Department of Informatics, University of Minho, Braga, Portugal. He is currently enrolled in a joint doctoral programme in Computer Science of three top Portuguese Universities. He develops scientific research in the field of Artificial Intelligence, namely Multi-Agent Systems, Genetic Programming and Case-Based Reasoning Applied to the Legal Field and Ambient Intelligence. He has participated in several research projects in the fields of Ambient Intelligence and Online Dispute Resolution. He is the co-author of several publications in his field of interest, including one book and multiple book chapters, journal papers and conference and workshop papers. He has been a member of the Program Committee of the 6th International Conference on Hybrid Artificial Intelligence Systems 2011 (HAIS 2011), Scientific Committee of the International Symposium on Distributed Computing and Artificial Intelligence (DCAI 2011), publicity and web chair of the Ambient Intelligence Environments (AmI Environments 2011), Local Organization Committee of the International Symposium on Ambient Intelligence (ISAmI 2010) and the International Workshop on Soft Computing Models in Industrial Applications (SOCO 2010). In 2008 he has been awarded the TLeIA08 – a National Award for Artificial Intelligence projects attributed by the Portuguese Artificial Intelligence Association and in 2009 he has been awarded an Academic Merit Scholarship by the Portuguese Government.



José Neves is a professor of Computer Science at Minho University, Portugal, since 1983. Jose Neves is the deputy director of the Division for Artificial Intelligence (AI). He received his PhD in Computer Science from Heriot Watt University, Scotland, in 1983. His current research interests relate to the areas of knowledge representation and reasoning and evolutionary intelligence, aiming to construct dynamic virtual worlds of complex symbolic entities that compete against one another in which fitness is judged by one criterion alone, intelligence, here measured in terms of a process of quantification of the quality of their knowledge, leading to the development of new procedures and other metaheuristics and their application in complex tasks of optimization and model inference in distinct areas, namely in the healthcare arena (e.g. machine learning in an intensive care unit environment and ambient intelligence).



Paulo Novais is an associate professor of Computer Sciences at the Department of Informatics, in the University of Minho, Braga, Portugal and researcher at the CCTC (Computer Science and Technology Center) in which he is the coordinator of the Intelligent Systems Lab (http://www.islab.di.uminho.pt/). From the same university he received a PhD in Computer Sciences in 2003 and his Habilitation in Computer Sciences (Agregação ramo de conhecimento em Informática) in 2011. He was the director of the Master in Informatics Engineering and the Master in Informatics, during 2012–2014 and he is the deputy director and co-founder of the Master in Law and Informatics at the University of Minho. He develops scientific research in the field of Artificial Intelligence, namely Knowledge Representation and Reasoning, Machine Learning and Multi-Agent Systems, with applications to the areas of Law and Ambient Intelligence. He has leaded and participated in several research projects sponsored by Portuguese and European public and private institutions and has supervised several PhD and MSc students. He is the co-author of over 150 book chapters, journal papers, conference and workshop papers and books. He was a vice-president of APPIA, the Portuguese Association for Artificial Intelligence, during 2006–2013.

Please cite this article as: A. Pimenta, et al., A neural network to classify fatigue from human–computer interaction, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.03.105

2.4 A multi-modal architecture for non-intrusive analysis of performance in the workplace

Title	**A multi-modal architecture for non-intrusive analysis of performance in the workplace**
Authors	Carneiro D., Pimenta A., Neves J., Novais P.
Publication Type	Journal
Publication Name	Neurocomputing
Publisher	Elsevier
Volume	231
Number	29
Pages	41-46
Year	2017
Month	March
Online ISSN	0925-2312
Print ISSN	
Url	http://dx.doi.org/10.1016/j.neucom.2016.05.105
State	Published
Scimago journal ranking	IF 1,202 (Q1 - COMPUTER SCIENCE APPLICATIONS)
ISI JCR	IF 2.392 (Q1 - COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE)

ARTICLE IN PRESS

Neurocomputing xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A multi-modal architecture for non-intrusive analysis of performance in the workplace

Davide Carneiro[a,b,*], André Pimenta[a], José Neves[a], Paulo Novais[a]

[a] ALGORITMI/Department of Informatics, University of Minho, Braga, Portugal
[b] CIICESI, ESTGF, Polytechnic Institute of Porto, Portugal

ARTICLE INFO

Keywords:
Performance
Monitoring
Neural network
Non-intrusive

ABSTRACT

Human performance, in all its different dimensions, is a very complex and interesting topic. In this paper we focus on performance in the workplace which, asides from complex is often controversial. While organizations and generally competitive working conditions push workers into increasing performance demands, this does not necessarily correlates positively to productivity. Moreover, existing performance monitoring approaches (electronic or not) are often dreaded by workers since they either threat their privacy or are based on productivity measures, with specific side effects. We present a new approach for the problem of performance monitoring that is not based on productivity measures but on the workers' movements while sitting and on the performance of their interaction with the machine. We show that these features correlate with mental fatigue and provide a distributed architecture for the non-intrusive and transparent collection of this data. The easiness in deploying this architecture, its non-intrusive nature, the potential advantages for better human resources management and the fact that it is not based on productivity measures will, in our belief, increase the willingness of both organizations and workers to implement this kind of performance management initiatives.

1. Introduction

The change in the job offers in the last decades, caused by technological evolution, brings along many significant and broad changes. Some of the most notorious ones can be pointed out by the emergence of indicators such as stress or mental fatigue which, in extreme cases, can endanger the life and well-being of the employees. In more moderate cases it will impair performance, general cognitive skills and productivity. In addition to these factors, many of these jobs are the so-called desk-jobs, in which people frequently sit for more than 8 h [1].

Until now, the performance of the employees has been evaluated through their productivity: the more one produces, the better the performance at work. While the true nature of this relationship is yet to be thoroughly studied (properly contextualized in each work domain), there are other issues that need to be addressed. First of all, the worst aspect about this approach is that it only points out a potential decrease of performance after a productivity loss. This means that the "damage" is already done and that it is most likely too late for the employee to cope with whatever caused the performance loss. An approach that could point out, in advance, upcoming breaks in performance (e.g. through the observation of behavioral patterns) could allow for preventive interventions rather than reactive [2].

Another major aspect that current approaches fail to consider are the side effects of productivity or performance monitoring in the workplace [3]. Indeed, as research in the last years has pointed out, this kind of approach might act as an additional stressor on the employee, which adds to the existing pressure in the workplace and to its negative consequences. In a study conducted in 1995 by researchers of the State University of New Jersey, it was analyzed the impact of electronic performance monitoring and its social context on the productivity and stress of employees [4].

Electronic Performance Monitoring (EPM) systems are one of the many technological developments employees face in today's workplaces. These systems provide managers a wide range of information about employees' routines including real-time information such as the pace of work, degree of accuracy, log-in and log-out times, and even the amount of time spent on bathroom breaks. This study examined how productivity and subjective experiences are affected by EPM systems and how the social context of the workplace moderates that influence.

In a survey involving the monitored workers, 81% of the respondents declared that electronic observation made their jobs more stressful [5]. Another study compared the behavior of monitored and non-monitored workers and found that monitored workers felt more

* Corresponding author at: ALGORITMI/Department of Informatics, University of Minho, Braga, Portugal.
E-mail addresses: dcarneiro@di.uminho.pt (D. Carneiro), apimenta@di.uminho.pt (A. Pimenta), jneves@di.uminho.pt (J. Neves), pjon@di.uminho.pt (P. Novais).

http://dx.doi.org/10.1016/j.neucom.2016.05.105
Received 15 November 2015; Received in revised form 23 March 2016; Accepted 2 May 2016
Available online xxxx
0925-2312/© 2016 Elsevier B.V. All rights reserved.

Please cite this article as: Carneiro, D., Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.05.105

ARTICLE IN PRESS

stressful [6].

The introduction of EPM systems can transform ordinary jobs into high-stress jobs. It can also reduce the opportunities for employees to socialize with each other at work, leading to a loss of social support, partially responsible for the stress associated with EPM [7,8].

In this paper we present a new approach on the problem of performance monitoring in desk jobs, in line with the Ambient Assisted Living view [9], that quantifies performance independently of the amount of work produced by the employee. Namely, we develop a multimodal approach that incorporates different sources of information to allow the extraction of behavioral and physical features to characterize the performance of the user. These features are extracted from the keyboard, the mouse and the chair of the user.

As the results point out, the selected features vary consistently throughout the day, showing a decrease in the performance of interaction as the day of work goes by and an increase in the movement of the chair, pointing out increasing discomfort. This multimodal approach allows a quantification of performance decrements without requiring productivity measures. In that sense, employees will be more prone to accept such an approach. It is also non-intrusive as it requires no specific actions by the employees: they simply need to carry out their regular work.

Our goal is that team managers make use of such information to implement better human resources management strategies, that take into account (possibly in real-time) the state of the employees, allowing the development of individualized working schedules, warnings when performance decreases or the implementation of automatized coping strategies. As an example, we developed a simple desktop application that produces a warning when significant decreases in performance are observed on the user.

2. Architecture

The architecture of the proposed system was developed as a Service-Component Architecture (SCA): a group of OASIS specifications that has become an industry standard. It is intended for the development of applications based on SOA, which defines how computing entities interact to perform work for each other. Originally published in November 2005, SCA is based on the notion that all the functions in a system should exist in the form of services that are combined into composites to address specific business requirements. In other words, it allows to build service-oriented applications as networks of service components. SCA is used for building service components, assemble components into applications, deploy to (distributed) runtime environments and reuse service components built from new or existing code using SOA principles.

SCA provides a good basis for applications under the umbrella of the Ambient Intelligence (AmI) field [10], such as this one, and it fulfills major AmI deployment requirements by promoting late bindings at deploy time and runtime with the support of several relevant technologies including POJO, SOAP, REST, BPMN, BPEL, JMS, Camel or Rules services. But most of all it is currently supported by several major commercial and open source products such as Jboss Switchyard, IBM WebSphere or TRENTINO (C++). From the several available implementations of SCA we have chosen JBoss SwitchYard since it is an open source solution in a relative mature state, and also enhances some of the SCA advantages.

A service-based approach was followed to develop an architecture logically divided into several packages that encapsulate a set of features and tasks. Fig. 1 pictorially depicts, from a high-level point of view, the proposed architecture.

Three main components can be identified. The first is the component that is on the user-area. It provides, in an non-intrusive way, features about the employees' behavior. Two of these features are extracted from two accelerometers placed in the chair, as detailed in Fig. 1. These accelerometers aimed to record the movements of the workers during the day (8 h workday), while sitting in front of the computer. Accelerometer 1 was placed at the level of the worker's back while accelerometer 2 was placed in one of the wheeled arms of the chair, with the aim to record acceleration generated by the moving of the chair. Specifically, Axivity's WAX3 wireless accelerometers were used.

The remaining features are extracted from the mouse and keyboard of the computer and fully characterize the employee's interaction with these peripherals. Although these features and the process of their extraction have been detailed in the past [11], we provide a brief overview. From the keyboard we extract features such as the typing speed or rhythm, the number of errors or the time of each key press. In previous studies we have noted that key presses, for example, may vary in duration from 80 ms at the beginning of a session, to as much as 100 ms at the end, depending on the user and on the type of task being performed. From the mouse we extract features such as velocity, acceleration, distance between clicks, duration of the click or excess of distance traveled, just to name a few. Once again, we consistently find decreases in performance on these features as the workday progresses [12].

Simultaneously with the acquisition of the behavioral features, employees answered a questionnaire about mental fatigue on a hourly basis (USAFSAM Fatigue Scale [13]). This was implemented with the aim of studying, in parallel, the daily evolution of mental fatigue given the well-known relationship between this indicator and performance [14].

The second component of the architecture is placed on the server side of the system. It is responsible for the continuous acquisition of behavioral data and its persistence in the database. Moreover, it also provides very important functionalities in the form of services. Namely, it allows for behavioral models to be trained based on individual user data and the results of the questionnaire (as depicted further below), namely using machine learning algorithms. It also allows for these models to be used in real-time for classifying user performance from behavioral data, which is essential given the aims of this research line.

Finally, the third major component of the architecture concerns the use of the performance models, in real-time, for improving performance throughout the day or preventing performance breaks. This considers both automatic and human-driven decision mechanisms. On the one hand, and as implemented in the developed prototype, the system can autonomously point out to the user or anticipate significant breaks in performance, providing a warning that encourages the user to

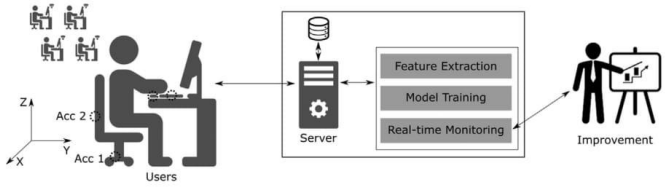


Fig. 1. High-level view of the architecture, highlighting the 4 sources of information in each user: the mouse, the keyboard and two accelerometers placed on the chair as well as their placement and directions of the axes.

take a break. Other more advanced context-dependent coping strategies can be implemented (e.g. taking a walk in a nearby park, grabbing a coffee, group activity with coworkers).

On the other hand, human decision makers can also make use of this kind of information, which may or may not be directly shown to the employees, in order to better manage their teams. Namely, managers can learn the best working schedule for each individual or each employee's resilience to working long hours. This will allow them to implement better strategies in what concerns the management of personnel, ultimately achieving working environments that are better for both organizations (e.g. increased performance/productivity) and collaborators (e.g. no stress from productivity measures, better social relationships, improved quality of life).

Summarizing, in an organizational context, the gathering and analysis of metrics describing people's behavior, and the providing of tools for visualization (particularly real time analytics) enables better decision making and data-driven actions that consider the state and well-being of each individual worker. Such initiatives can nowadays be scaled to hundreds or thousands of workers, through the use of Big Data tools and techniques, without compromising performance and availability.

The interaction features described above characterize the behavior of each individual while interacting with the computer. This behavior, as many others, is affected by factors that influence performance at work, including mental fatigue, stress level or emotional arousal. Specifically, each instance of the behavior is characterized by fifteen values (represented as doubles) that are a result of applying several data summarization techniques (e.g. aggregation of collected data by calculating values such as mean and variance on the very frequently collected values). Each of these instances also contains a timestamp.

Given that this data is stored in a MongoDB database, each record needs 136 bytes of storage space: 15 times 8 bytes (the MongoDB double size) plus 8 bytes for the timestamp, and 8 bytes for the two keys that describe the application being used and the user. A new record is produced every five minutes, for each user of the environment. Assuming that each individual is expected to work around 8 h per day, a production of around 12.75 Kbytes of data per worker is estimated. **Table 1** shows the expected data growth projections for different numbers of users and different time-spans.

3. Validation of the architecture

The developed architecture was used in a laboratory setting with the main aim to assess a potential relationship between performance and the features under study, namely the acceleration measured on the chair and the interaction with the peripherals. In what concerns the interaction with the peripherals, 24 participants (19 men) interacted with the computer while performing their regular tasks, with their interaction patterns being recorded. Of these, and due to hardware availability limitations, only 8 had accelerometers placed in their chairs, as depicted in **Fig. 1**. As would happen in a real-life environment, these participants were requested to come into the lab and perform their regular activities, without any restriction whatsoever.

Data was collected continuously and organized on an hourly basis.

Table 1
Projections of data growth for different number of users and time frames.

Time	# Users			
	1	100	10 000	1 000 000
5 min	136 bytes	13.28 Kbs	1.297 Mbs	129.7 Mbs
1 day	12.75 Kbs	1.245 Mbs	124.5 Mbs	12.159 Gbs
1 week	89.25 Kbs	8.716 Mbs	871.6 Mbs	85.115
1 month	382.5 Kbs	37.354 Mbs	3.648 Gbs	364.8 Gbs
1 year	4.545 Mbs	454.5 Mbs	44.382 Gbs	4.438 Tbs

Each hour of data originates one instance, with each instance depicting the variance of each feature in [redacted] user, as well as a subjective evaluation of each user's level of fatigue, provided through a self-report mechanism. Specifically, we used the seven-point USAFSAM Mental Fatigue Scale created by Dr. William F. Storm and Captain (Dr.) Layne P. Perelli of the Crew Performance Branch of the USAF School of Aerospace Medicine, Brooks AFB, San Antonio, Texas, and then used in many field and laboratory tests [15].

Data from the accelerometers was organized in hourly series, so as to match the intervals at which questionnaires were collected. Using the Anderson–Darling test, it was determined that the distributions of the eight resulting datasets were not normal. Given this, the Kruskal Wallis test was used in the subsequent analysis. This test was used to compare the distributions of each of the three axes and of the two accelerometers. The maximum p-value observed was 2.2e−16, which demonstrates that the changes in acceleration that occur from one interval to the next are statistically significant.

The difference in the distributions can also be visually observed as in **Fig. 2**. This figure depicts violin plots, which in addition to the information provided by boxplots include the probability density for the different values. These plots show the evolution of the distributions of the values during the day.

Having carried out this preliminary analysis of the data, it was examined whether there is any relationship between the subjective feeling of mental fatigue and the acceleration measured on the chair, in both positions. To this end it was used the Pearson's test to determine the statistical correlation between the acceleration measured in each axis and the subjective level of fatigue.

Data show that the readings of accelerometer 1 have a positive correlation with the level of mental fatigue for the axes X and Y (0.67 and 0.53, respectively). Concerning accelerometer 2, a strong negative correlation with the subjective feeling of fatigue exists for the y-axis (−0.9).

The Mann–Whitney test was also used to compare the distributions of the values of each feature extracted from the mouse and the keyboard, in order to verify that there were statistically significant difference between the two groups of data (**Table 2**).

4. Results

User feedback was paired with the variance of the features considered in each hourly period to train the neural network and thus understand the relationship between the variables. In the preliminary analysis of the data of the accelerometers, a relationship between the subjective levels of mental fatigue and the movement of the chair in some of its axes in both accelerometers was found. For modelling accelerometer 1, which showed a strong negative correlation with fatigue for axes X and Y and no significant correlation for the z-axis, a linear regression was fit to the average acceleration values on these axes. While the z-axis is left out due to a lack of correlation, its absence is also positive in the sense that acceleration from when the user sits down or stands up (which happens mostly on this axis and is not the normal behaviour we are observing) will not influence the model.

The trained linear model contains residual values shown in **Fig. 3** where it is possible to verify that these values follow a normal distribution, and they have a median value close to 0. In this model we can still observe a Multiple R-squared: 0.8385, Ajusted R-squared: 0.8331 which confirms the correlation shown between the acceleration in the y-axis and the level of fatigue.

For the data recorded by the accelerometer 2, K-means clustering was used to verify that the acceleration of the chair movement in all three axes can be associated with the different levels of fatigue observed. The value of K was selected using the NbClust package from the R Software [16]. This package provides 30 indexes for determining the number of clusters and proposes the best clustering scheme from the different results obtained by varying all combinations of number of

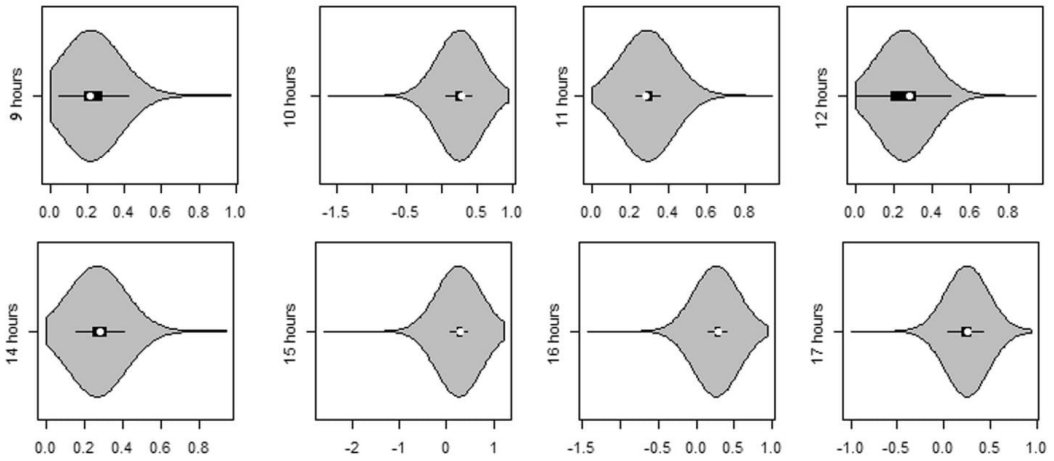


Fig. 2. Hourly violin plot of *x*-axis for accelerometer 2.

Table 2
Analysis of significance of the collected data. Column a depicts the average value of the Mann Whitney test. Column b depicts the percentage of individuals for which the differences observed in each feature when fatigued and rested were statistically significant.

clusters, distance measures, and clustering methods. The suggested value of K with the largest number of indexes chosen was three. Therefore a 3-cluster solution was proposed.

The accuracy of the resulting clusters was computed through the adjusted Rand index. The adjusted Rand index provides a measure of the agreement between two partiti

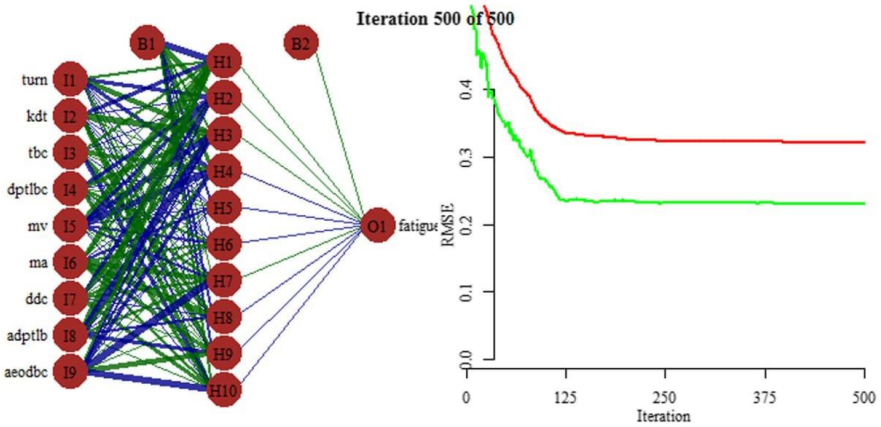


Fig. 4. Evolution of the RMSE during the evolution of the training of the ANN (green line) and during validation (red line). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

day (morning, afternoon, evening) to rule out the influence of the circadian rhythm.

As a classification algorithm we selected an Artificial Neural Network (ANN). This decision was taken after an analysis of existing work that showed the suitability of this kind of algorithms for modeling mental fatigue. Namely, existing works use electroencephalography signals [17], electromyography [18] and other features such as gaze detection or face pose [19] as inputs to neural networks. Existing related work is especially targeted at vehicle driving and operation of machines, which is a classical field of application of fatigue detection and performance monitoring [19]. To the extent of our knowledge, it is the first time that an ANN is used with performance features such as those put forward in this work. This is fundamental for domains such as the workplace or classrooms, in which the use of invasive approaches results impractical.

The previously mentioned group of features thus constitutes the input layer of the neural network. Its output is a value between 1 and 7 that denotes the degree of the level of fatigue, as in the USAFAM questionnaire.

With the input and output layers defined, a multilayer feed-forward neural network was used to model the relationship between behavioural features and a level of fatigue. The R software environment was used. A back-propagation learning algorithm was used which, besides the input and output layers, uses a intermediate hidden layer that lies between the input layer and the output layer, with a total of 10 nodes. The number of hidden nodes has been selected after the carrying out of several performance tests with different numbers of hidden nodes. In what concerns activation and error functions, the algorithm's default settings were used.

The network was trained with a dataset containing a total of 74 instances, each of which containing the value of the variance of each of the features over a period of one hour and the level of subjective fatigue provided by the users through the questionnaire, for that period. The network was trained during 125 iterations where it reached a minimum RMSE (Root-Mean-Square Error) as can be seen in Fig. 4.

The ANN trained and outlined above was tested and validated with data from the second week of the data collection period, in which each instance, as described above, contained the variance of each feature in the period of an hour and the subjective measure of the level of fatigue of the user. This approach allows us to compare the value pointed out by the user against the value provided as output by the ANN. The main result is that the trained ANN correctly classified 81% of the instances, i.e., producing as output the same value that the user provided in the questionnaire. The 19% that were misclassified where nonetheless classified as neighbouring values.

5. Conclusions

This paper presented a distributed architecture for the non-intrusive acquisition of interaction and behavioral features for the classification of human performance. We have shown that certain axes of accelerometers placed on specific points of a chair as well as some of the interaction features extracted from the mouse and keyboard do correlate with performance. The most positive aspect of this work is that it allows the classification of human performance without the use of features that are not positively regarded by workers, namely productivity measures.

In that sense, workers will be more prone to accept this kind of performance monitoring. From the point of view of the organization, it is a inexpensive approach and does not require any change to the workers' routines, which is positive. Moreover, it will provide valuable information to team managers, namely concerning each of the workers' natural rhythms (e.g. some work better in the morning, others in the afternoon) thus allowing for a more effective management of human resources.

In the overall, this will result in more sensitive and positive working environments, with expected positive impacts in productivity measures as well as the quality of the product and of the workplace.

The ANN developed in this work is a part of a wider performance model being developed by this research team which includes related aspects such as the performance of interaction and the degree of attentiveness. Ultimately, we aim at the development of a theoretical and practical multi-modal model that details how performance in the workplace is influenced by external factors and how it can be assessed non-intrusively.

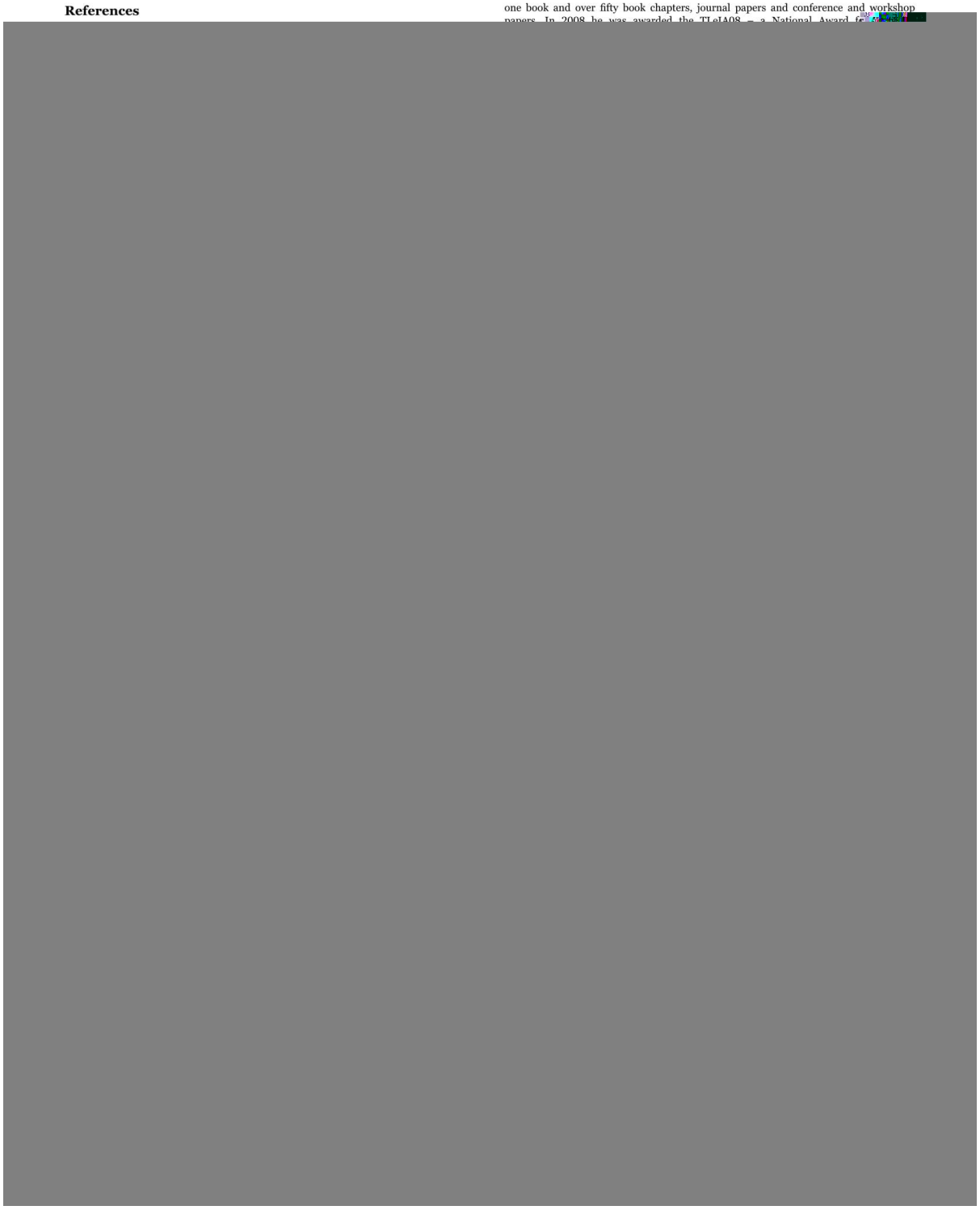
Acknowledgments

This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013. The work of Davide Carneiro is supported by a Post-Doctoral Grant by FCT (SFRH/BPD/109070/2015).

ARTICLE IN PRESS

References

one book and over fifty book chapters, journal papers and conference and workshop papers. In 2008 he was awarded the TLoIA08 – a National Award fo[illegible]



2.5 Non-intrusive quantification of performance and its relationship to mood

Title	**Non-intrusive quantification of performance and its relationship to mood**
Authors	Carneiro D., Pimenta A., Neves J., Novais P.
Publication Type	Journal
Publication Name	Soft Computing
Publisher	Springer Berlin Heidelberg
Volume	21
Number	17
Pages	4917-4923
Year	2017
Month	September
Online ISSN	1432-7643
Print ISSN	
Url	http://dx.doi.org/10.1007/s00500-016-2380-y
State	Published
Scimago journal ranking	IF 0.829 (Q2 - THEORETICAL COMPUTER SCIENCE)
ISI JCR	1.630 (Q2 - COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE)

Soft Comput
DOI 10.1007/s00500-016-2380-y

FOCUS

Non-intrusive quantification of performance and its relationship to mood

Davide Carneiro[1,2] · **André Pimenta**[1] · **José Neves**[1] · **Paulo Novais**[1]

© Springer-Verlag Berlin Heidelberg 2016

Abstract The number of jobs that takes place entirely or partially in a computer is nowadays very significant. These workplaces, as many others, often offer the key ingredients for the emergence of stress and the performance drop of its long-term effects: long hours sitting, sustained cognitive effort, pressure from competitiveness, among others. This has a toll on productivity and work quality, with significant costs for both organizations and workers. Moreover, a tired workforce is generally more susceptible to negative feelings and mood, which results in a negative environment. This paper contributes to the current need for the development of non-intrusive methods for monitoring and managing worker performance in real time. We propose a framework that assesses worker performance and a case study in which this approach was validated. We also show the relationship between performance and mood.

Keywords Worker performance · Mood · Statistical analysis · Distributed intelligence

Communicated by C. Analide.

✉ Davide Carneiro
dcarneiro@di.uminho.pt

André Pimenta
apimenta@di.uminho.pt

José Neves
jneves@di.uminho.pt

Paulo Novais
pjon@di.uminho.pt

[1] Algoritmi Centre/Department of Informatics, University of Minho, Braga, Portugal

[2] CIICESI, ESTG, Polytechnic Institute of Porto, Porto, Portugal

1 Introduction

We currently live in a society that moves at an unprecedented pace. People need to find a balance between work, family and leisure, and doing so is increasingly harder (Kobayashi and Demura 2006). In an attempt to achieve it, people stretch their limits and their days, usually at the expense of rest and relaxation time. This together with inadequate sleep patterns is leading causes of fatigue.

Mental fatigue may not be felt immediately or may not even be felt at all, but it has effects at many different levels, including health, performance, decision-making and even safety. These effects are often visible in students preparing for examinations, office or industrial workers, healthcare professionals, drivers, pilots or military personnel. Fatigue may even put people working in safety-sensitive jobs at risk, and any mistake on their part can lead to loss of lives (Tucker 2003; Jaber and Neumann 2010).

Mental fatigue can be seen as a state that involves a number of effects on a set of cognitive, emotional and motivational skills and usually results in overall discomfort, as well as the emergence of performance limitations (Perelli 1980). Some of these limitations imply that a tired person is often less willing to engage in tasks of effort, or perform the task of a conditioned form, well below their normal capacity (Lorist et al. 2000). Thus, mental fatigue may be characterized by a perception of a lack of mental energy. Persons who are affected by mental fatigue may feel like they have less energy than usual and are unusually tired and lethargic. Excessive activity and stimulation of the brain can cause a person to feel mentally exhausted, and the feeling is similar to what the body feels when a person is physically fatigued (Williamson et al. 2005).

Fatigue effects may occur at any moment and they may persist from only a few hours to several consecutive days.

Published online: 03 October 2016

Depending on its duration and intensity, fatigue may make the carrying out of daily tasks increasingly difficult or even impossible. In severe or prolonged cases, it can cause illnesses such as depression or chronic fatigue syndrome (Tucker 2003). The importance of detecting and managing fatigue, especially in the workplace, assumes thus an unquestionable importance nowadays.

This paper presents a framework for monitoring worker performance in real time and in a non-intrusive way. For this purpose, the framework acquires and analyzes worker performance indicators. Specifically, it considers the performance of the interaction with the computer: how the user interacts with the mouse and the keyboard. We describe a case study in which this framework was used to assess the performance of two groups of students: one taking classes in the morning and the other in the afternoon. We show that performance is significantly lower in the afternoon. We assessed the participant's fatigue in two different ways: through a questionnaire and using our framework. Both approaches achieve very similar results, which point out increased fatigue in the afternoon. Moreover, the afternoon group also evidences more negative states of mind.

The proposed framework may thus constitute an important step not only for monitoring and managing performance in the workplace but also to improve other interrelated aspects such as mood, well-being and workplace quality (Carneiro and Novais 2014).

1.1 Profiling and others factors

There are many factors, both internal and external, that modulate the onset and presence of fatigue. These include sleep deprivation, naps, noise, heat, mood, motivation, time of day, workload and of course the individual profile (Balkin and Wesensten 2011).

The user's profile provides valuable information with respect to the potential level of fatigue. It can be seen as a predicted base level of fatigue in the sense that it establishes a baseline, according to the lifestyle of the individual (Akerstedt et al. 2004). These aspects have been thoroughly studied, mostly by psychologists, and encompass:

- *Age* Defines the mental age of the individual. It is important to understand the expected cognitive abilities of the individual, which may have a tendency to degrade with old age.
- *Gender* The mental states are different between men and women.
- *Professional occupation* Many occupations are intrinsically more tiresome or exhausting than others.
- *Consumption of alcohol and drugs* The use of certain substances for short or prolonged periods may cause dependencies and other effects that lead to a state of mental fatigue.

Mental fatigue is also affected by a number of other external factors. They may or may not be directly related to the individual's behavior. They include:

- *The mood* of the individual may influence decisively his or her mental state, with a particular effect on his or her motivation to work. Although tired, the individual may overcome (even if only temporarily) the effects of fatigue with a positive mood and motivation.
- *Stress* may be defined as the demands placed upon the individual's mind or body by external stimuli, requiring the individual to acclimatize to the dynamic requirements of the environment. However, these processes of acclimatization require an additional effort from the brain which, when prolonged over long or intense periods, will result in mental fatigue.
- *Mental Workload* as a result of the relationship between the amount of mental processing capacity and the amount required by the task.
- *Sleepiness* is often mistaken for mental fatigue or generalized as such. A difference exists and must be pointed out. However, the mistake is understandable since sleepiness is a symptom that is strongly connected to mental fatigue: It is one of the methods our brain uses to tell us that it is running out of resources. Sleepiness often results in a general loss of the individual vitality.

These factors can be assessed using validated tools such as the USAFSAM fatigue scale (Samn and Perelli 1982) for mental fatigue states, or as the NASA TLX (Hart and Staveland 1988) in the case of mental workload.

However, these instruments, based on the individual's subjective interpretation of the symptoms, do not fully take into account inter-individual differences. There are instruments that help to account for individual differences, such as the Profile of Mood States (POMS) (Curran 1995) and the State-Trait Anxiety Inventory (STAI) (Spielberger 2010). However, their use in complex systems can prove to be complicated and confusing, due to the same problems that can be observed in subjective measures of fatigue detection (Newcombe and Boyle 1995; Curran 1995).

The POMS can also be used to estimate fatigue with the Vigor–Activity and Fatigue–Inertia factors. The POMS measures five aspects of affect or mood (McNair et al. 1971). It consists of 65 adjectives describing feeling and mood to which the subject responds according to a five-point scale ranging from "Not at all" to "Extremely." Results are reported as six mood factors, namely

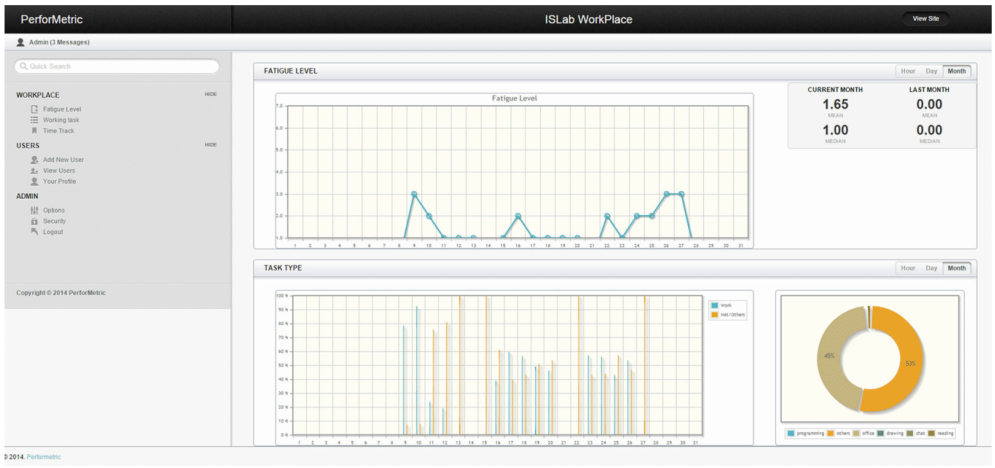


Fig. 1 Interface of the performance monitoring Web service

– *Tension–Anxiety* Heightened musculoskeletal tension including reports of somatic tension and observable psychomotor manifestation;
– *Depression–Defection* Depression accompanied by a sense of personal inadequacy;
– *Anger–Hostility* Anger and antipathy toward others;
– *Vigor–Activity* Vigorousness, ebullience and high energy;
– *Fatigue–Inertia* Weariness, inertia and low energy level; and
– *Confusion–Bewilderment* Bewilderment, muddleheadedness, appears to be an organized–disorganized dimension of emotion.

Because of its length, the POMS only results practical for occasional uses such as establishing a baseline and estimating the effects of excessive sleep deprivation or restriction.

2 Framework

In this paper, we propose a non-invasive, non-intrusive, real-time approach to assess worker performance through the analysis of keyboard and mouse interaction patterns. The analysis takes place directly from the usage of an individual's computer as within the context of the so-called desk jobs. We build on the fact that computers are nowadays used as major work tools in many workplaces to devise a non-invasive method based on the observation of the worker's interaction with the computer.

The main aim is to develop leisure and work context-aware environments that may improve quality of life, mental health and individual performance, as well as productivity in organizations. To achieve this purpose, we follow the guiding lines of Ambient Intelligence (AmI) (Ramos et al. 2008), in which the technological aspects are hidden in the environment and the user is placed in the middle of the paradigm. There is also a focus on non-intrusiveness, with acquisition of information taking place without the need for explicit or conscious user interactions.

Another major objective of the system is to support the decision-making processes of team managers or group coordinators. In this perspective, each element of a group/organization is seen as part of a whole which contributes to the general state of the group. Thus, the estimated performance level of the group is the average result of the fatigue level of its elements (Fig. 1).

Specifically, the framework acquires and compiles data regarding the following features of interaction:

– Absolute Sum of Degrees: Quantifies how much the mouse moves or not in straight lines. A higher value in this feature means that the movement of the mouse has many curves, which is a sign of decreased efficiency, i.e., the movement is more efficient if it moves in a more straight line between each two clicks (units: degrees);
– Average Distance of the Pointer to the Line: Quantifies the average distance of the mouse pointer to the straight line defined by each two consecutive clicks. A higher value denotes that the user is moving the mouse farther away from this line, which is a sign of decreased efficiency (units: pixels);

- Average Excess of Distance: The distance of the straight line defined by each two consecutive clicks is the shortest possible distance between those two clicks. This feature measures how much the mouse pointer traveled in excess (units: pixels);
- Click Duration: It measures the length of each click, not considering long clicks associated with mouse dragging nor double clicks. The longer the click, the smaller the efficiency (units: milliseconds);
- Distance Between Clicks: This feature quantifies the amount of distance traveled by the mouse between each two consecutive clicks. A larger distance between each two consecutive clicks is a sign of decreased performance (units: pixels);
- Distance of the Pointer to the Line: This feature is similar to the previously mentioned Average Distance of the Pointer to the Line. However, it does not compute the average value but rather the sum of the distances of each position of the mouse to the straight line, between each two consecutive clicks (units: pixels);
- Excess of Distance: It is similar to the previously mentioned Average Excess of Distance, but the excess of distance is not divided by the distance between the two consecutive clicks (units: pixels);
- Mouse Velocity: It quantifies the velocity of the mouse by dividing the distance traveled (in pixels) between each two consecutive clicks by the time it took (in seconds) (units: pixels/milliseconds);
- Time Between Clicks: This feature quantifies the time, in seconds, that passes between each two consecutive clicks. Higher values are sign of a slower working rhythm (units: milliseconds);

These features describe, in an extensive manner, the performance of the user's interaction with the mouse. Additional features are also compiled by the framework concerning the use of the keyboard, window switching patterns, among others. These are, however, outside the scope of this work.

Besides the data acquisition module, the framework is composed of the following key components:

- Dashboard: An online reporting tool to aggregate and compile information about the team's performance, aimed at supporting decision-making and team management;
- Database: To maintain and manage the large amounts of information that are acquired and compiled about the workers;
- Machine Learning: A group of tools and algorithms to train models that characterize the users' interaction with the computer and the behavioral changes caused by different levels of performance.

3 Case study

In order to test and validate the proposed system and approach, a case study was prepared using detection system models which were previously trained (Pimenta et al. 2015, 2013). To reach that goal, we used two groups of volunteers who used the tool while they were providing feedback of their state of fatigue through USAFSAM Fatigue scale, as well as their mood through the POMS questionnaire.

Feedback values recorded during the monitoring period were used to validate the system and used to analyze the variation of moods.

The participants in the case study, fourteen in total (10 men and 4 women), were students from the University of Minho in the field of physics. Their age ranged between 18 and 25.

3.1 Methodology

The methodology followed to implement the study was devised to be as minimally intrusive as the approach it aims to support. Participants were provided with an application for logging the previously mentioned events of the mouse and keyboard. This application, which maintained the confidentiality of the keys used, needed only to be installed in the participant's computer and would run in the background, starting automatically with the operating system. The only explicit interaction needed from the part of the user was the input of very basic information on the first run, including some personal identifying and profiling information.

As mentioned, two different groups of users were selected (7 participants in each group) who underwent the experience in different periods of the week. The first group was monitored on a Monday morning (starting at 9 a.m.), while the second did so on a Thursday in the afternoon (starting at 14 p.m.). Each session had a duration of 3 h. During the session, in addition to the collection of interaction features, user feedback regarding their state was collected hourly.

3.2 Data analysis

A statistical analysis of the interaction data collected was carried out before analyzing the results in terms of mental fatigue classification. The aims of this analysis are twofold: (1) to demonstrate that there are, indeed, significant changes in the participants' interaction with the computer in different moments of the day and (2) to demonstrate that the proposed framework is able to detect these changes based on worker performance.

Given the characteristic of the task attributed to the participants, that required mostly the use of the mouse, in this section we focus on the analysis of the features extracted from the mouse.

Table 1 Mean and median values of the distribution of the data for each feature and each group

Feature	Mean		Median		*p*-value
	Group 1	Group 2	Group 1	Group 2	
Absolute Sum of Degrees	4786.89	7603.2	2716.49	3745.91	5.15×10^{-27}
Average Distance of the Pointer to the Line	37.08	49.63	15.89	23.01	5.97×10^{-18}
Average Excess of Distance	4.07	8.02	1.46	1.78	4.06×10^{-18}
Click Duration	216.32	208.52	94	94	0.009
Distance Between Clicks	187.86	275.34	96.71	135.88	1.28×10^{-17}
Distance of the Pointer to the Line	8390.94	16663.6	1352.74	2412.89	2.00×10^{-20}
Excess of Distance	340.73	612.51	83.04	167.44	8.58×10^{-27}
Mouse Velocity	0.42	0.41	0.18	0.18	0.003
Time Between Clicks	16579.7	17876.3	2282	2547	0.018

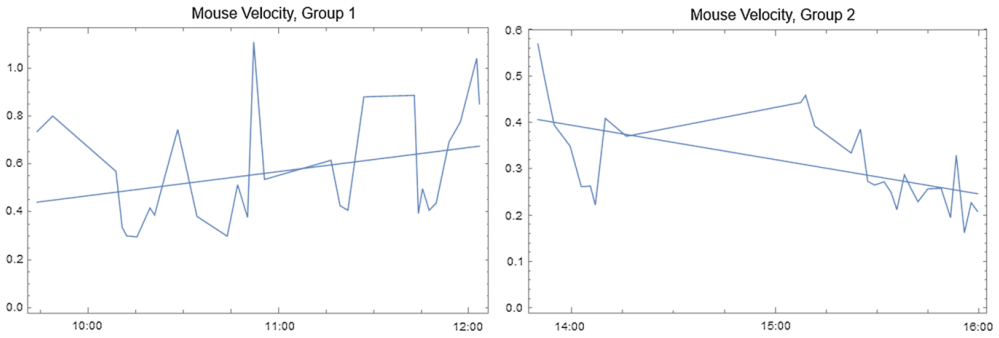


Fig. 2 Evolution of the velocity of the mouse during the experiment in the two groups

In this analysis, it is worth noting that both groups were carrying our similar tasks and that participants were randomly assigned to each group. The effects on performance that are shown here are thus potentially due to the effects of the circadian rhythm on the human body. Indeed, the performance of group 2 is significantly lower. This group carried out the assigned tasks in the afternoon.

Table 1 details, for each group and for each feature, the mean and median values as well as the *p*-value of the Mann–Whitney test, used to quantify the statistical differences of the distributions of the data for both groups. The mean values are higher in group 2 in seven of the features, while the median values are higher in seven of the features as well and equal in the other two. Moreover, all the distributions of the data are significantly different.

The data collected by the framework also allow more detailed and interesting analysis. For example, it is possible to analyze the evolution, over time, of any of the features for individual participants or for the group. Figure 2 shows the evolution of mouse velocity for two participants, one from each group. It is interesting to notice that at the beginning of the task both move their mouse with similar velocities (around 0.4 px/ms). However, as the task progresses, the evolution of mouse velocity is different for both groups: In group 1 it tends to increase (a sign of increased performance) while in group 2 it tends to decrease.

A similar trend can be observed in other features as well. Figure 3 shows the evolution of the excess of distance during the experiment, in the two groups. Once again, performance tends to increase during the morning and to decrease in the afternoon.

Similar trends can be observed in the remaining features. Essentially, the data show two different and equally interesting aspects: (1) Performance in the morning is significantly higher than in the afternoon and (2) these differences are statistically significant.

3.3 Results

During the two sessions, the performance monitoring system was used in real time. It allows not only recording the interaction patterns with the mouse and keyboard, but also calculating the estimated individual level of mental fatigue for each user and the estimated average level of each group.

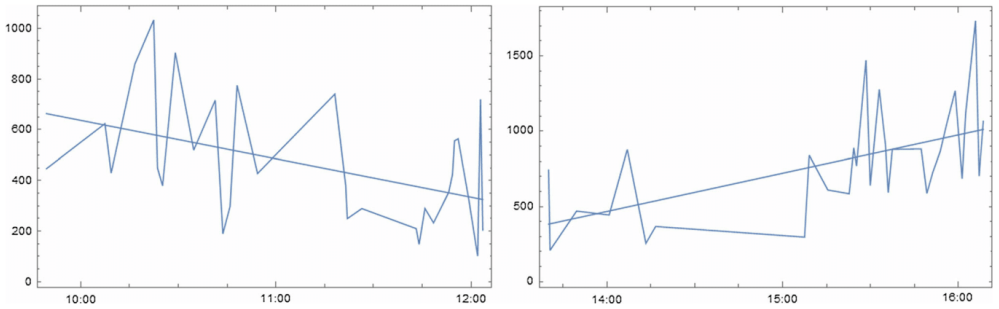


Fig. 3 Evolution of the excess of distance traveled by the mouse during the experiment in the two groups

Table 2 Average level of fatigue of the different groups in USAFSAM fatigue scale of the estimated values and feedback values

Group	Estimated fatigue level (SD)	Fatigue from feedback (SD)	RMSE
1	1.6 (0.7)	1.4 (0.8)	0.2
2	3.3 (0.8)	3.2 (1.2)	0.4

Table 3 Average vales of profile of mood states for group 1 and group 2, as well as the resulting p-value from the T-test

POMS factor	Group 1	Group 2	p-value
Tension–Anxiety	4.6	11.8	0.03
Depression–Dejection	4.4	9.1	0.02
Anger–Hostility	6.5	7.9	0.06
Vigor–Activity	19.4	7.6	0.04
Fatigue–Inertia	5.1	13.8	0.02
Confusion–Bewilderment	4.0	7.3	0.03

These values were compared with the feedback given by the participants.

Table 2 summarizes the results, showing that the first group evidences better performance than the second one. This can be confirmed through data collected during the session. It is also possible to check a root-mean-square error (RMSE) of 0.2 for group 1 and 0.4 for group 2. Taking into account the scale used, this is an acceptable error.

In addition to the subjective level of fatigue through the USAFSAM fatigue scale, it used the POMS factors in order to observe the influence of fatigue on the moods of the different groups and therefore the emergence of fatigue and loss of vigor.

Through Table 3 and Fig. 4, it is visible that the average values are different between the two groups. In addition, the T-test was used to validate the differences between the two groups, in order to determine whether the two sets of data are significantly different from each other, a fact that was confirmed.

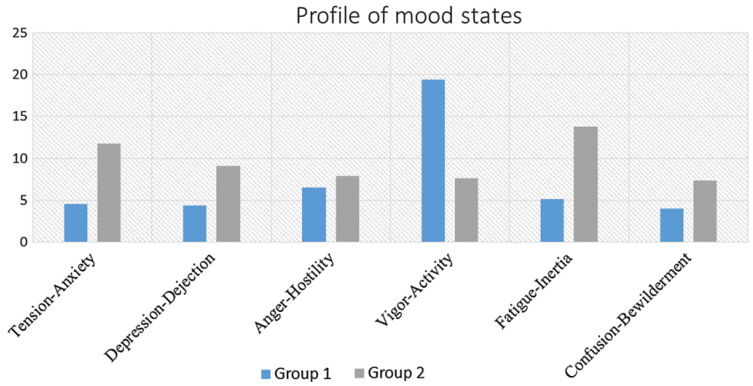


Fig. 4 Profile of mood states from groups 1 and 2: group 2 has more negative mood states

We can also observe that fatigue is higher in the second group and, on the contrary, vigor is lower in the second group. This is in accordance to the fatigue levels of each group.

4 Conclusion

This paper described a framework for monitoring the performance of individuals or groups while they interact with the computer, in real time. As addressed throughout the paper, information about individual performance may be extremely useful in workplaces. Namely, it may allow a far better management of each worker, taking into account each one's working rhythms, context and state. This is especially important in scenarios of critical tasks, which require individuals performing at their best.

The results of the study carried out demonstrate that it is possible to evaluate performance in a group of people using their interaction with the computer as input. The prototype used in the experiment uses a model trained in a previous study, with data of a different group of users. This also shows that the causes of performance changes on different users can be generalized, i.e., we all react similarly. The study carried out also showed that in assessing worker performance and overall work quality there are other important aspects such as worker mood.

In the context of academic environments, such as in the case study, this may allow to define better schedules for classes or examinations, for example, with the aim of improving student performance. In the workplace, similar strategies can be followed to improve worker performance. Moreover, and given that there are always inter-individual differences, personalized models can also be trained that shape each one's reaction. This will, undoubtedly, maximize the accuracy of fatigue classification and improve the success of performance management initiatives.

Acknowledgments This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT—Fundaçã para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013. The work of Davide Carneiro is supported by a Post-Doctoral Grant by FCT (SFRH/BPD/109070/2015).

Compliance with ethical standards

Conflict of interest None.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Akerstedt T, Knutsson A, Westerholm P, Theorell T, Alfredsson L, Kecklund G (2004) Mental fatigue, work and sleep. J Psychosom Res 57(5):427–433. doi:10.1016/j.jpsychores.2003.12.001

Balkin TJ, Wesensten NJ (2011) Differentiation of sleepiness and mental fatigue effects. Balkin, Thomas J.; Wesensten, Nancy J. Ackerman, Phillip L. (eds) Cognitive fatigue: multidisciplinary perspectives on current research and future applications. American Psychological Association, Washington, DC, pp 47–66. doi:10.1037/12343-002

Carneiro D, Novais P (2014) New applications of ambient intelligence. In: Ambient intelligence-software and applications, Advances in Intelligent Systems and Computing, vol. 291. Springer, pp. 225–232

Curran S (1995) Short form of the profile of mood states (POMS-SF): psychometric information. Psychol Assess 7(1):80–83. doi:10.1037//1040-3590.7.1.80

Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv Psychol 52:139–183

Jaber MY, Neumann WP (2010) Modelling worker fatigue and recovery in dual-resource constrained systems. Comput Ind Eng 59(1):75–84

Kobayashi H, Demura S (2006) Relationships between chronic fatigue, subjective symptoms of fatigue, life stressors and lifestyle in Japanese high school students. School Health 2:5

Lorist MM, Klein M, Nieuwenhuis S, De Jong R, Mulder G, Meijman TF (2000) Mental fatigue and task control: planning and preparation. Psychophysiology 37(5):614–625

McNair DM, Lorr M, Droppleman LF (1971) Manual for the profile of mood states. Educational and Industrial Testing Service, San Diego

Newcombe PA, Boyle GJ (1995) High school students' sports personalities: variations across participation level, gender, type of sport, and success. Int J Sport Psychol 26:277

Perelli LP (1980) Fatigue stressors in simulated long-duration flight. Effects on performance, information processing, subjective fatigue, and physiological cost. Tech rep, DTIC document

Pimenta A, Carneiro D, Novais P, Neves J (2013) Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. In: Hybrid artificial intelligent systems, Lecture Notes in Computer Science vol 8073, Springer, pp 222–231. doi:10.1007/978-3-642-40846-5_23

Pimenta A, Carneiro D, Novais P, Neves J (2015) Detection of distraction and fatigue in groups through the analysis of interaction patterns with computers. In: Intelligent distributed computing VIII, Studies in Computational Intelligence vol 570. Springer, pp 29–39. doi:10.1007/978-3-319-10422-5_5

Ramos C, Augusto JC, Shapiro D (2008) Ambient intelligence—the next step for artificial intelligence. IEEE Intell Syst 23(2):15–18

Samn SW, Perelli LP (1982) Estimating aircrew fatigue: a technique with application to airlift operations. Tech rep, DTIC document

Spielberger CD (2010) State-trait anxiety inventory. Wiley, Hoboken. doi:10.1002/9780470479216.corpsy0943

Tucker P (2003) The impact of rest breaks upon accident risk, fatigue and performance: a review. Work Stress 17(2):123–137

Williamson RJ, Purcell S, Sterne A, Wessely S, Hotopf M, Farmer A, Sham PC (2005) The relationship of fatigue to mental and physical health in a community sample. Soc Psychiatry Psychiatr Epidemiol 40(2):126–132

2.6 Improving User Privacy and the Accuracy of User Identification in Behavioral Biometrics

Title	**Improving User Privacy and the Accuracy of User Identification in Behavioral Biometrics**
Authors	Pimenta A., Carneiro D., Neves J., Novais P
Publication Type	Conference Paper
Publication Name	Portuguese Conference on Artificial Intelligence
Publisher	Springer – Lecture Notes in Artificial Intelligence
Volume	9273
Number	
Pages	15-26
Year	2015
Month	August
Online ISSN	978-3-319-23485-4
Print ISSN	978-3-319-23484-7
Url	https://doi.org/10.1007/978-3-319-23485-4_2
State	Published
Scimago journal ranking	IF 0.252 (Q3 – COMPUTER SCIENCE (MISCELLANEOUS)).
Indexed	Scopus, ISI Web of Science, DBLP

Improving User Privacy and the Accuracy of User Identification in Behavioral Biometrics

André Pimenta(✉), Davide Carneiro, José Neves, and Paulo Novais

Algoritmi Centre, Universidade Do Minho, Braga, Portugal
{apimenta,dcarneiro,jneves,pjon}@di.uminho.pt

Abstract. Humans exhibit their personality and their behavior through their daily actions. Moreover, these actions also show how behaviors differ between different scenarios or contexts. However, Human behavior is a complex issue as it results from the interaction of various internal and external factors such as personality, culture, education, social roles and social context, life experiences, among many others. This implies that a specific user may show different behaviors for a similar circumstance if one or more of these factors change. In past work we have addressed the development of behavior-based user identification based on keystroke and mouse dynamics. However, user states such as stress or fatigue significantly change interaction patterns, risking the accuracy of the identification. In this paper we address the effects of these variables on keystroke and mouse dynamics. We also show how, despite these effects, user identification can be successfully carried out, especially if task-specific information is considered.

Keywords: Mental fatigue · Machine learning · Computer security · Behavioral biometrics · Behavioral analysis

1 Introduction

In the last years there has been a significantly increase in jobs that are mentally stressful or fatiguing, in expense of otherwise traditional physically demanding jobs [1]. Workers are nowadays faced not only with more mentally demanding jobs but also with demanding work conditions (e.g. positions of high responsibility, competition, risk of unemployment, working by shifts, working extra hours). This results in the recent emergence of stress and mental fatigue as some of the most serious epidemics of the twenty first century [2,3]. In terms of workplace indicators, this has an impact on human error, productivity or quality of work and of the workplace. In terms of social or personal indicators, this has an impact on quality of life, health or personal development. Moreover, there is an increase in the loss of focus that leads people to be unaware of risks, thus lowering the security threshold.

Recent studies show the negative impact of working extra hours on productivity [4,5]: people work more but produce less. Stressful milieus just add to the

© Springer International Publishing Switzerland 2015
F. Pereira et al. (Eds.) EPIA 2015, LNAI 9273, pp. 15–26, 2015.
DOI: 10.1007/978-3-319-23485-4_2

problem. The questions is thus how to create the optimal conditions to meet productivity requirements while respecting people's well-being and health. Since each worker is different, what procedures need to be implemented to measure the level of stress or burnout of each individual worker? And their level of productivity? The mere observation of these indicators using traditional invasive means may change the worker's behavior, leading to biased results that do not reflect his actual state. Directly asking, through questionnaires or similar instruments, can also lead to biased results as workers are often unwilling to share feelings concerning their workplace with their coworkers.

Recent approaches for assessing and managing fatigue have been developed that look at one's interaction patterns with technological devices to assess one's state (e.g. we type at a lower pace when fatigued). Moreover, the same approaches can be used to identify users (e.g. each individual types in a different manner). This field is known as behavioral biometrics. In this paper we present a framework for collecting from users in a transparent way, that allows to perform tasks commonly associated to behavioral biometrics. Moreover, this framework respects user privacy. Finally, we show how including the user's state and information about the interaction context may improve the accuracy of user identification. The main objective of this work is to define a reliable and non-intrusive user identification system for access control.

1.1 Human-Computer Interaction

Currently there is a very large community of computer users. In fact, most of the jobs today require some type of computer usage [6]. Moreover, with services like home banking, tech support services and social networks, people start to interact more with computers than with other persons, even to take care of important aspects of their lives.

There is thus a new form of communication in which computers serve as intermediaries. Therefore, the perception of a conversation between human beings is somewhat lost. By using a computer, the people involved cannot perceive one of the most important aspects in communication which is body language. Other important aspects include speech, intonation or facial expressions, just to mention a few. To overcome this loss of information computer systems must adopt new processes to better perceive the human being [7].

In fact, Humans tend to show their personality or their state through their actions, even in an unconscious way. Facial expressions and body language, for example, have been known as a gateway for feelings that result in intentions. The resultant actions can be traced to a certain behavior. Therefore, it is safe to assume that a human behavior can be outlined even if the person does not want to explicitly share that information.

Human behavior can also be deemed complex as it is driven by internal and external factors, such as personality, culture, education, social roles, life experiences, among others. Accurately evaluating a behavior requires constant observation of all the elements that are able to provide useful information. Nonetheless, with our evolved social skills, we are often able to conceal certain emotions or

shielding them with others. Thus, multi-modal approaches should be considered for increased accuracy (e.g. relying solely on visual emotion recognition may be less accurate than including additional aspects such as tone of voice or speech rhythm). An potentially interesting process is to consider involuntary actions, that cannot consciously controlled by the individual. This often includes movement and posture, hand gestures, touches and interaction with objects the environment, among others. Therefore, it can be stated that the observation and evaluation of behaviors must consider not only the displayed emotions and actions, but also the nature of the interaction with the environment.

One particularly interesting source of these unconscious behaviors is our interaction with computers and other technological devices. In fact, we don't think about controlling the rhythm at which we type on a keyboard or the way we move the mouse when we become fatigued or stress, although we might want to hide our state from our colleagues or superiors. But the truth is that our interaction does change, as we have established in previous work [8,9]. Under different states we use the mouse and the keyboard differently. Moreover, we also interact differently with the smartphone. The case of the smartphone is still more interesting as it provides a range of sensors that are not available on other platforms and can provide very valuable information about interaction patterns, including a touch screen (that provides information about touches, their intensities, their area or their duration), gyroscopes, accelerometers, among others.

Human-Computer Interaction thus becomes a very promising field when it comes to reliable sources of information for characterizing one's state following behavioral approaches, the main advantage being that the individual generally does not or cannot consistently change such precise and fine-grained behaviors. Thus, while one can, to some extent, fake facial expressions and transmit a chosen emotion or state, doing so successfully through these behaviors results much harder.

2 Security Systems

As stated before, with the increase of Human-computer Interaction and the development of social-networking, people rapidly increased the rate at which they share information, even when it is sensitive personal information. Some of the current concerns are thus privacy, security and data protection.

Deemed as the most profitable crime of modern times, information theft is increasing at an alarming rate, the corporate sector being the most affected. This type of crime frequently includes the theft of personal data and often does a large damage in a person's life, as well as in companies. One way to improve security is to build robust authentication systems which prevent unauthorized access to machines. These systems may range from password-based authentication, in the simplest cases, to biometry systems in the most complex ones.

One of the possible ways to increase security in the context of user identification is to consider the user' behavior when interacting with a technological

device. As each individual has a particular way to walk, talk, laugh or do anything else, each one of us has also their own interaction patterns with technological devices. Moreover, most of the applications we interact with have a specific flow of operation or require a particular type of interaction, restricting or conditioning the user's possible behaviors to a smaller set. Maintaining a behavioral profile of authorized users may allow to identify uncommon behaviors on the current user that may indicate a possible unauthorized user. This is even more likely to work when behavioral information for particular applications is used.

Such systems are known as Behavioural Biometrics: they rely on the users' behavioral profiles to establish the behavior of authorized users. Whenever, when analyzing the behavior of the current user, a moderate behavioral deviation is detected from the known profiles, the system may take action such as logging off, notifying the administrator or using an alternative method of authentication.

Such systems can also include behaviors other than the ones originated from keyboard and mouse interaction patterns [10]. In fact, any action performed on the technological device can be used as threat detection. For example, if the system console is started and the authorized user of the device never used the console before, a potential invasion may be taking place. Similar actions can be taken on other applications or even on specific commands (e.g. it is unlikely that a user with a non-expert profile suddenly starts using advanced commands on the console).

To implement behavioral biometrics, distinct procedures that can be adopted, such as:

- Biometric Sketch: this method uses the user drawings as templates for comparison [11,12]. The system collects patterns from the user drawing and compares it to others in a database. Singularity is assured by the number of possible combinations. The downside is that the drawings must be very precise, which in most of the cases is quite difficult, even more by using a standard mouse to draw.
- GUI Interaction: this technique uses the interaction of the user with visual interfaces of the applications and compares it to the model present in the database. For every application that the user interacts with, a model must be present. Thus, both the model and the application must be saved in the database. This method requires that every action per application is saved, resulting in a large amount of information to be maintained. Moreover, each new application or update must be trained and modeled. Therefore, this method is very strict and complex to implement and maintain.
- Keystroke Dynamics: this method uses the keyboard as input and is based on the user's typing patterns. It captures the keys pressed, measuring time and pressing patterns, extracting several features about the typing behavior. This is a well established method, as it relies solely on the user's interaction, allowing to create simple and usable models.
- Mouse Dynamics: this method consists in capturing the mouse movement and translating it into a model. All the interactions are considered, such as movement and clicking. This method is similar to the GUI Interaction method but simpler; although it suffers the same context problem. The main

purpose of this method is to accompany the Keystroke Dynamics, providing additional data to the main model, thus increasing accuracy and decreasing false positives.
- Tapping: this method focuses on the pulse wave resulting from a touch sensor. The pulse duration and tapping interval are the proprieties considered for the analysis. The way a user acts with a smartphone or computer can be monitored, and a model can be extracted. Conceptually, it is a solid method, as tapping follows the same concept of the keystroke dynamics, but in reality the usage context must also be considered, as tapping changes according to the application being used.

In this work, Keystroke Dynamics and Mouse Dynamics are chosen as inputs. Their broader features and availability are the traits that suit the aim of the intended system. They are application independent and operating system independent, and are nowadays the most common input method when interacting with computers.

3 The Framework

The framework developed in the context of this work is a unified system, composed of two main modules: fatigue monitoring and security. The process of monitoring is implemented using an application that captures the keyboard and mouse inputs transparently. The features used are the same in both systems and are defined in more detail in [13]. The features extracted from the keyboard are:

- Key down time: total time that a key is pressed
- Time between keys: time between a key being released and the following key being pressed
- Writing speed: the rhythm at which keys are pressed
- Errors per key: quantification of the use of the backspace and delete keys

And the features extracted from the mouse are:

- Distance of the mouse to the straight line: the sum of the distances between the pointer and the line defined by each two consecutive clicks
- Mouse acceleration: the acceleration of the mouse
- Mouse velocity: the velocity of the mouse
- Average distance of the mouse to the straight line: the average distance at which the mouse pointer is from the line that is defined by each two consecutive clicks
- Total excess of distance: the distance that the mouse travels in excess between each two consecutive clicks
- Average excess of distance: the distance that the mouse travels in excess divided by the shortest distance between each two consecutive clicks
- Time between clicks: the time between each two consecutive clicks
- Distance during clicks: distance traveled by the mouse while performing a click

- Double click duration: the time between two clicks in a double click event
- Absolute sum of angles: the quantification of how much the mouse turns, regardless of the direction of the turn, between each two consecutive clicks
- Signed sum of angles: the quantification of how much the mouse turns, considering the direction of the turn, between each two consecutive clicks
- Distance between clicks: the distance traveled between each two consecutive clicks

The data gathered may be processed differently to extract the information related to each scope of the framework (fatigue monitoring and stress). An integrated system is beneficial due to the jointly nature of the data and to the fact that only one application is present locally, thus having a low footprint on computer resources. Furthermore, these are two areas that are intrinsically connected, and one can affect the other. Their joint analysis is fundamental to the achievement of the proposed objectives.

3.1 Providing Security and Safety in the Monitoring System

The use of Keystroke Dynamics and Mouse Dynamics for detecting behavior is extremely useful, especially since it allows the creation of non-intrusive and non-invasive systems.

However the use of behavioral biometrics, in particular the use of keystroke dynamics, can pose some risks to the data security and privacy of the user, especially when the data obtained from the mouse and keyboard are processed by remote or 3rd party Web Services. For this reason, and to ensure the users' privacy and data security, data must be encrypted. The most sensitive data is indeed the information about which key was pressed. However, this specific information is mostly irrelevant for behavior analysis, i.e., we are not interested in *which* keys the user presses but on *how* the user presses them.

We therefore propose the collection of events created by the mouse and the keyboard in the following way:

- MOV, timestamp, posX, posY - an event describing the movement of the mouse, in a given time, to coordinates (posX, posY) in the screen;
- MOUSE_DOWN, timestamp, [Left—Right], posX, posY - this event describes the first half of a click (when the mouse button is pressed down), in a given time. It also describes which of the buttons was pressed (left or right) and the position of the mouse in that instant;
- MOUSE_UP, timestamp, [Left—Right], posX, posY - an event similar to the previous one but describing the second part of the click, when the mouse button is released;
- MOUSE_WHEEL, timestamp, dif - this event describes a mouse wheel scroll of amount dif, in a given time;
- KEY_DOWN, timestamp, encrypted key - identifies a given key from the keyboard being pressed down, at a given time;
- KEY_UP, timestamp, encrypted key - describes the release of a given key from the keyboard, in a given time;

In this approach, the encrypted key replaces the information about the specific key pressed. It is therefore still possible, while hiding what the user wrote, to extract the previously mentioned features, thus guaranteeing the user's privacy. The encryption of the pressed keys is carried out through the generation of random key encryptions at different times. This is done as depicted in Algorithm 1, which exemplifies the developed approach for the case of the key down time feature. An example of the result of the algorithm is depicted in Table 1, where a record with and without encryption is depicted for different keys.

```
Data: Keyboard Inputs
Result: List of KeyDownTime records
while Keyboard Inputs have records do
    Get timestamp of a KEY_DOWN ;
    Get code of a KEY_DOWN ;
    while Keyboard Inputs have records do
        Get timestamp of a KEY_UP ;
        Get code of a KEY_UP ;
        if KEY_DOWN code and KEY_UP code are the same then
            Save the difference between KEY_UP timestamp and KEY_DOWN
            timestamp;
        end
    end
end
```

Algorithm 1. Key down time algorithm

Table 1. Example of a keyboard log with and without encryption.

No Encryption	Random Encryption Key 1
KD,63521596046072,A	KD,63521596046072,1COc0qNOOk=
KU,63521596046165,A	KU,63521596046165,1COc0qNOOk=
KD,63521596057943,v	KD,63521596057943,sMA0Wu0n3k=
KU,63521596058037,v	KU,63521596058037,sMA0Wu0n3k=
KU,63521596058084,a	KU,63521596058084,hb0s0lHEF8+sA==
KU,63521596058037,v	KU,63521596058037,sMA0Wu0n3k=

Hiding user input is just one part of the solution for the issue of user security. The other is to prevent intrusions. In this scope, behavioral biometrics security systems can run in two different modes [14]: identification mode and verification mode. In this system we use the identification mode instead of the verification mode to ensure a constant user identification in the monitoring system.

The identification mode is the process of trying to discover the identity of a person by examining a biometric pattern calculated from biometric data of

the person. In this mode the user is identified based on information previously collected from keystroke dynamics profiles of all users. For each user, a biometric profile is built in a training phase. When in the running phase, the usage pattern being created in real-time is compared to every known model, producing either a score or a distance that describes the similarity between the pattern and the model. The system assigns the pattern to the user with the most similar biometric model. Thus, the user is identified without the need for extra information.

4 Case Study

The system was analyzed and tested in four different ways. As a first step we used the records of 40 users registered in the monitoring system to train different models. The created models were then validated through 150 random system usage records, taken from the system in order to validate the models created. In a second step models were created using the type of task to be performed in addition to the biometric information, and in the third step, models were trained using the user's fatigue state. We finally created models that, in addition to using biometric data, also used the type of task and the user's mental state at the time of registration. Both the type of task as the user's mental state are provided by the monitoring system.

The participants, forty in total (36 men, 4 women) which are registered in the monitoring system. Their age ranged between 18 and 45. The following requirements were established to select, among all the volunteers, the ones that participated: (1) familiarity and proficiency with the use of the computer; (2) use of the computer on a daily basis and throughout the day; (3) owning at least one personal computer.

4.1 Results and Discussion

After training different models (Naive Bayes, KNN, SVM and Random Forest) with data from different users on the system, different degrees of accuracy have been obtained in user identification, as depicted in Figure 1. It is also possible to observe that the type of task and the level of fatigue have influence on the process of identifying the user, since these factors effectively influence interaction patterns. Taking this information into consideration allows the creation of more accurate models.

The type of task being carried out during the monitoring of the interaction patterns is particularly important, mainly due to the very nature of the task, as well as the set of tools available to perform the task. Figure 2 shows the values of features Key Down Time and Average Excess of Distance for five different types of applications: Chat, Leisure, Office, Reading and Programming. The way each different application conditions the interaction behavior is explicit. Such information must, therefore, absolutely be considered while developing behavioral biometrics system based on input behavior. Table 2 further supports this claim showing that data collected in different types of applications has statistically

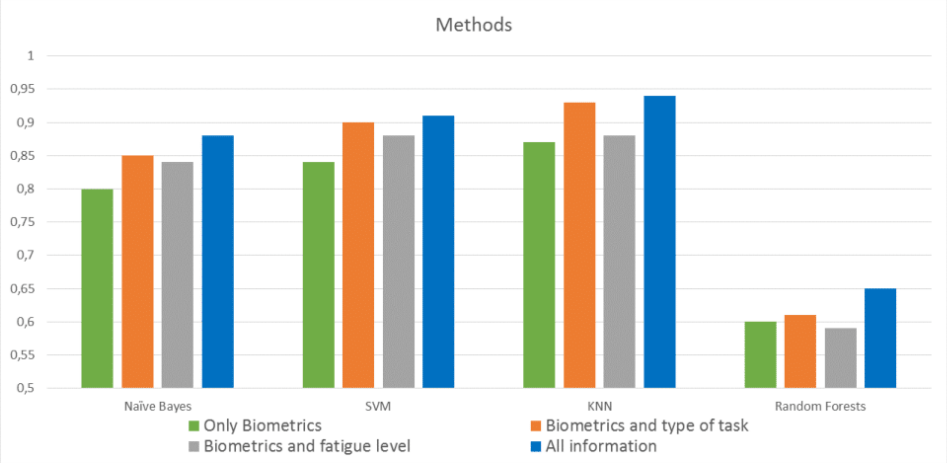


Fig. 1. Accuracy of the different algorithms and inputs considered for user authentication.

Table 2. *p*-values of the Kruskal-Wallis test when comparing the data organized according to the types of application and to the level of fatigue. In the vast majority of the cases, the differences between the different groups are statistically significant.

Features	**Applications**	**Fatigue**
Writing Velocity	4.4e-15	0.43
Time Between Keys	<2.2e-16	<2.2e-16
Key Down Time	<2.2e-16	3.8e-14
Distance of the Mouse to the SL	0.01	<2.2e-16
Mouse acceleration	<2.2e-16	1.3e-09
Average Distance of the Mouse to the SL	0.03	<2.2e-16
Mouse velocity	<2.2e-16	<2.2e-16
Average Excess of Distance	0.04	1.6e-08
Time Between Clicks	<2.2e-16	<2.2e-16
Distance During Clicks	<2.2e-16	0.05
Total Excess of Distance	0.22	<2.2e-16
Double Click Duration	1.0e-4	<2.2e-16
Absolute Sum of Angles	8.1e-3	2.5e-13
Signed Sum of Angles	0.91	<2.2e-16
Distance between clicks	0.06	<2.2e-16

significant differences for most of the features. The same happens for different levels of fatigue.

Another extremely important aspect in user identification is the influence of mental states on interaction patterns. Previous studies by our research team [8,15] show that individuals under different states of stress, fatigue, high/low mental workload or even mood evidence significant behavior changes that impact interaction patterns with devices. They do, consequently, influence behavioral

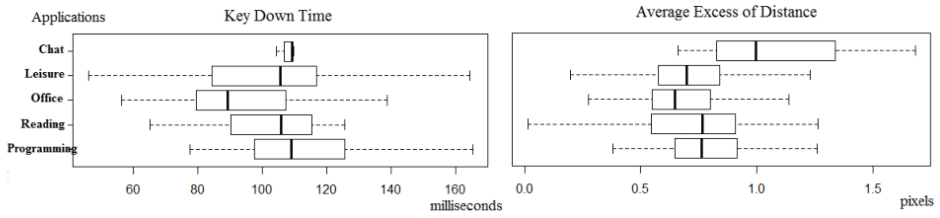


Fig. 2. Differences in the distributions of the data when comparing interaction patterns with different applications, for two interaction features.

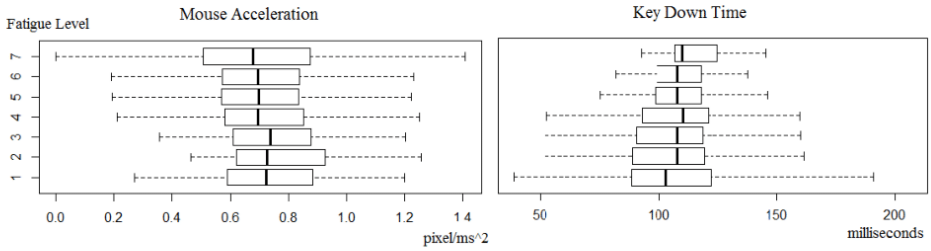


Fig. 3. Effects of different levels of fatigue on the interaction patterns, depicted for two different features.

biometric features. Figure 3 depicts this influence for two interaction features. Numbers in the y-axis represent the level of fatigue as self-reported by the individual using the seven-point USAFSAM Mental Fatigue Scale questionnaire [16]. Each value represents the following state:

1. Fully alert. Wide awake. Extremely peppy.
2. Very lively. Responsive, but not at peak.
3. Okay. Somewhat fresh.
4. A little tired. Less than fresh.
5. Moderately tired. Let down.
6. Extremely tired. Very difficult to concentrate.
7. Completely exhausted. Unable to function effectively. Ready to drop.

It is therefore possible to see how increased levels of fatigue result in generally less efficient interactions of the participants with the computer. For example, a higher value of the Mouse Acceleration depicts a more efficient interaction behavior in which the user is moving the mouse. The same happens with Key Down Time, where a shorter time corresponds to a more efficiency in the use of the keyboard.

This conclusion justifies the need for the inclusion of mental states on behavioral biometrics approaches and explains the increased accuracy of the presented approach concerning user identification when all modalities are used jointly: interaction patterns, user state and type of application, as depicted previously in Figure 1.

5 Conclusions and Future Work

This paper describes a non-invasive and non-intrusive approach to the field of behavioral biometrics. Specifically, we described a system that, transparently, acquires interaction features from keystroke dynamics and mouse dynamics, with the purpose of user identification. We analyzed the significant effect that mental states have on interaction patterns, particularly fatigue, to conclude that these aspects must be considered when developing identification systems based on interaction patterns. Likewise, we also show how user interaction significantly changes according to the application being used.

These notions support the claim put forward in this paper that accurate user identification approaches based on behavioral biometrics must absolutely include these aspects when building user profiles. That much is evidenced by our results. In all four classification algorithms used for user identification, the cases in which all these types of information are used together outperform the others.

Moreover, the proposed approach is also concerned with user privacy. Specifically, it is based not on *what* the user inputs but on *how* the user inputs. To this end it conceals each key under an encryption key that prevents remote services from knowing what the user typed. The data gathered from users thus respects their privacy while allowing to extract the necessary features for the system to carry out its task, i.e., to detect unauthorized accesses.

In future work we will focus on including additional user features, namely in what concerns user state. Specifically, we will include information regarding the user's level of stress, which we have already determined in previous work to significantly influence interaction patterns with technological devices. In the long-term we will include more potentially interesting features, namely the user's mood. It is our conviction that this kind of information, when available, can significantly increase the accuracy of user identification in behavioral biometrics approaches.

Acknowledgments. This work is part-funded by ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-028980 (PTDC/EEI-SII/1386/2012) and project PEst-OE/EEI/UI0752/2014.

References

1. Tanabe, S., Nishihara, N.: Productivity and fatigue. Indoor Air **14**(s7), 126–133 (2004)
2. Miller, J.C.: Cognitive Performance Research at Brooks Air Force Base, Texas, 1960–2009. Smashwords, March 2013

3. Wainwright, D., Calnan, M.: Work stress: the making of a modern epidemic. McGraw-Hill International (2002)
4. Folkard, S., Tucker, P.: Shift work, safety and productivity. Occupational Medicine **53**(2), 95–101 (2003)
5. Rosekind, M.R.: Underestimating the societal costs of impaired alertness: safety, health and productivity risks. Sleep Medicine **6**, S21–S25 (2005)
6. Beauvisage, T.: Computer usage in daily life. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 575–584. ACM (2009)
7. Pantic, M., Rothkrantz, L.J.: Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE **91**(9), 1370–1390 (2003)
8. Pimenta, A., Carneiro, D., Novais, P., Neves, J.: Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. In: Pan, J.-S., Polycarpou, M.M., Woźniak, M., de Carvalho, A.C.P.L.F., Quintián, H., Corchado, E. (eds.) HAIS 2013. LNCS, vol. 8073, pp. 222–231. Springer, Heidelberg (2013)
9. Carneiro, D., Castillo, J.C., Novais, P., Fernández-Caballero, A., Neves, J.: Multimodal behavioral analysis for non-invasive stress detection. Expert Systems with Applications **39**(18), 13376–13389 (2012)
10. Lee, P.M., Chen, L.Y., Tsui, W.H., Hsiao, T.C.: Will user authentication using keystroke dynamics biometrics be interfered by emotions?-nctu-15 affective keyboard typing dataset for hypothesis testing
11. Al-Zubi, S., Brömme, A., Tönnies, K.D.: Using an active shape structural model for biometric sketch recognition. In: Michaelis, B., Krell, G. (eds.) DAGM 2003. LNCS, vol. 2781, pp. 187–195. Springer, Heidelberg (2003)
12. Brömme, A., Al-Zubi, S.: Multifactor biometric sketch authentication. In: BIOSIG, pp. 81–90 (2003)
13. Pimenta, A., Carneiro, D., Novais, P., Neves, J.: Analysis of human performance as a measure of mental fatigue. In: Polycarpou, M., de Carvalho, A.C.P.L.F., Pan, J.-S., Woźniak, M., Quintian, H., Corchado, E. (eds.) HAIS 2014. LNCS, vol. 8480, pp. 389–401. Springer, Heidelberg (2014)
14. Shanmugapriya, D., Padmavathi, G.: A survey of biometric keystroke dynamics: Approaches, security and challenges (2009). arXiv preprint arXiv:0910.0817
15. Rodrigues, M., Gonçalves, S., Carneiro, D., Novais, P., Fdez-Riverola, F.: Keystrokes and clicks: measuring stress on E-learning students. In: Casillas, J., Martínez-López, F.J., Vicari, R., De la Prieta, F. (eds.) Management Intelligent Systems. AISC, vol. 220, pp. 119–126. Springer, Heidelberg (2013)
16. Samn, S.W., Perelli, L.P.: Estimating aircrew fatigue: a technique with application to airlift operations. Technical report, DTIC Document (1982)

2.7 Mental workload management as a tool in e-learning scenarios

Title	**Mental workload management as a tool in e-learning scenarios**
Authors	Pimenta A., Gonçalves S., Carneiro D., Riverola F., Neves J., Novais P
Publication Type	Conference Paper
Publication Name	Pervasive and Embedded Computing and Communication Systems (PECCS), 2015 International Conference on
Publisher	IEEE
Volume	
Number	
Pages	25-32
Year	2015
Month	February
Online ISSN	978-989-758-137-3
Print ISSN	978-989-758-084-0
Url	http://ieeexplore.ieee.org/document/7483728
State	Published
Scimago journal ranking	
Indexed	Scopus (Document Type: Conference Paper), DBLP

Mental Workload Management as a Tool in e-Learning Scenarios

André Pimenta[1], Sergio Gonçalves[2], Davide Carneiro[1], Florentino Fde-Riverola[2], José Neves[1] and Paulo Novais[1]

[1]*Departament of Informatics, University of Minho, Braga, Portugal*

[2]*Informatics Department, University of Vigo, Ourense, Spain*

apimenta@di.uminho.pt, sgoncalves@alumnos.uvigo.es, dcarneiro@di.uminho.pt, riverola@uvigo.es, {jneves, pjon}@di.uminho.pt

Keywords: Mental Workload, Mental Fatigue, Machine Learning, e-Learning, Fatigue Management, Human Performance.

Abstract: In our daily life, we often have a sense of being exhausted due to mental or physical work, together with a feeling of performance degradation in the accomplishment of simple tasks. This is in part due to the fact that the working capacity and the performance of an individual, either physical or mental, generally decrease as the day progresses, although factors like motivation also play a significant role. These negative effects are especially significant when carrying out long or demanding tasks, as often happens in an educational context. In order to avoid these effects, initiatives to promote a good management of the time and effort invested in each task are mandatory. Such initiatives, when effective, can have a wide range of positive effects, including on the performance, productivity, attention and even mental health. Seeking to find a viable and realistic approach to address this problem, this paper presents a non-invasive and non-intrusive way to measure mental workload, one of the aspects that affects mental fatigue the most. Specifically, we target scenarios of e-learning, in which the professor may not be present to assess the student's state. The aim is to create a tool that enables an actual management of fatigue in such environments and thus allows for the implementation of more efficient learning processes, adapted to the abilities and state of each student.

1 INTRODUCTION

In our day-to-day we often feel a sense of tiredness during mental or physical work, generally known as fatigue. The term *fatigue* is used to describe a series of manifestations that range from drowsiness or loss of concentration to lack of physical strength or agility (van der Linden et al., 2003). Thus, it is a very broad and subjective term that may include symptoms such as loss of performance (loss of attention, slowed reaction and response times, impaired decision-making, and poor performance on tasks that generally reflect the good performance) as well as more subjective ones such as sleepiness and tiredness (Williamson et al., 2005; Perelli, 1980).

In seeking to formalize it, fatigue may be defined as a degree of failure of physical or mental factors associated with loss of physical or mental performance, hindering the natural or spontaneous accomplishment of a usual activity. Bartlett provides one of the clearest definitions of fatigue with respect to day-to-day tasks (Bartlett, n.d.) :
Fatigue is a term used to cover all those determinable changes in the expression of an activity which can be traced to the continuing exercise of that activity under its normal operational conditions, and which can be shown to lead, either immediately or after delay, to deterioration in the expression of that activity, or, more simply, to results within the activity that are not wanted.
I.D. Brown, on the other hand, conceptualized mental fatigue as:
(...) the subjective experience of individuals who are obliged (...) to continue working beyond the point at which they are confident of performing their task efficiently (...) [Fatigue is] the subjectively experienced disinclination to continue performing the task at hand. [The] main effect of fatigue [is] a progressive withdrawal of attention from the task at hand. [This withdrawal] may be sufficiently insidious that [operators] are unaware of their impaired state and hence in no position to remedy it (Brown, 1994).

Beyond the well-known effects on mood or energy, fatigue is also the cause or partial cause of errors and accidents. Often, this happens because fatigued individuals are unaware of the degree of their

impaired mental state (Miller, 2013). These errors and accidents assume particular importance when we

course (Brunger et al., 2004).

Without the obligatory physical presence of a teacher, the process of e-learning is exposed to some deficiencies that may result in poor student learning. Specifically, the teacher is not able to observe the students in search of signs evidencing problems such as doubts, frustration, stress or fatigue, preventing teachers from taking action in such scenarios. The setting up of appropriate monitoring mechanisms in the context of e-learning is therefore very important in achieving an efficient learning process.

As shown in the literature, for an efficient monitoring of the student to take place, it is crucial that the e-learning system allows for a personalized study strategy and is able to show the needs and strengths of each student (Cantoni et al., 2004). Thus it becomes possible to track the progress of students, as well as improving their learning, by providing better personalized learning methods. The identification of learning problems and the cause of those problems is another advantage that can be achieved through the e-learning context, and via monitoring systems, such as the tool proposed in this paper.

ments of the course while at the same time focusing on getting feedback that may allow to better steer the

consider the domain of high risk jobs that involve operating vehicles as well as the military, firefighters or medical personal, just to name a few. Beyond these immediate problems, fatigue can also lead to health problems in the long term, such as chronic fatigue syndrome or depression.

The negative effects of fatigue are thus clear. Moreover, they are also broad in the sense that affect many of our cognitive abilities. Learning is one of the functions that becomes impaired when under fatigue. Hence the importance of addressing this issue, especially in a time in which the teacher and student are growing apart due to the increasing use of electronic tools for learning. Indeed, due to the separation imposed by technology, it results more and more difficult for teachers to be sensible to the state of their students, impairing their ability to adapt both the contents and the teaching strategy accordingly.

This work details a tool for fatigue management in e-learning scenarios, with or without the teacher's presence, through the assessment of mental workload quantified in terms of the interaction patterns of the user with the computer. Through the use of behavioural biometries, specifically Keystroke Dynamics and Mouse Dynamics, we analyse the type of task performed by each user, the time spent performing it, as well as the mental workload of the task. With this information we train classifiers that are able to distinguish situations in which users show signs of fatigue or high mental workload.

This approach can be considered both non-invasive and non-intrusive, since it is based solely on the observation of the use of the mouse and keyboard, which allow for an assessment of the user's performance. This approach opens the door to the development of fatigue management initiatives in the context of e-learning, allowing teachers to not only have a better notion about their student's state but also to more efficiently adapt and above all personalize teaching strategies.

1.1 The Need for Monitoring in the Context of e-Learning

Electronic instruction, more commonly known as e-learning, is increasingly used as a method of teaching. E-learning differs from classroom-based training in several ways. Thus, the transition from a traditional course to a course supported by e-learning can be complex and difficult. There is the need for a good

1.2 Including Subjective Measures of Workload

In this paper we look at the monitoring and managi of mental workload as a way to improve the qual of information of the e-learning environment, esp cially to improve the teacher's decision making abi ties. One of the important parts of this work is a pre ously developed approach, deemed non-invasive a non-intrusive, for the analysis of the students' int action patterns.

Indeed, it was established in preliminary work th one's patterns of interaction with the computer, me sured in terms of the use of the keyboard and mou change when under fatigue as well as in periods of creased mental workload or even stress. Moreover were also found behavioral differences in performi different kinds of tasks, allowing to analyze patter of attention in the students who participated in t experimental studies (Pimenta et al., 2015; Pimen et al., 2014).

However, the work developed so far has the sho

a important aspect when it comes to determining the actual level of fatigue. It is also important, for example, to distinguish between scenarios of boredom

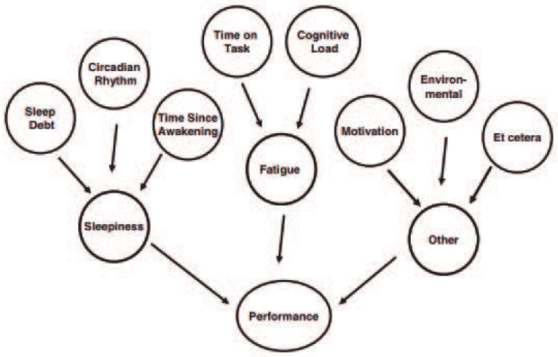


Figure 1: Conceptualisation of the set of factors that influence human performance and fatigue. (Balkin and Wesensten, 2011).

This aspect is now included in this paper, thus representing a step forward in the development of more accurate fatigue assessment approaches (Figure 1) that encompass the type of task, the time on task, and the mental workload of the task. Indeed, workload levels can help isolate the causes affecting performance at a given time, improving fatigue management initiatives. To this end, and besides the metrics derived from the use of the mouse and keyboard, subjective measures of mental workload are also used.

Obtaining mental workload levels during task performance may be a challenging procedure. Moreover, the workload level experienced by an individual can affect task performance twofold: either through excessive or reduced mental workload. To this end, subjective measures are often used, some of them detailed in (Reid et al., 1982).

The two instruments most often used in research were developed in parallel in the 1980s, one at the NASA-Ames Research Center in California and the other within the U.S. Air Force human factors research group at Wright-Patterson AFB, Ohio.

The NASA Task Load Index (NASA-TLX) is a multidimensional assessment tool (Hart and Staveland, 1988). The main seven-point scale is: Overall

- How successful were you in performing the task?
- How satisfied were you with your performance?

The TLX has five seven-point subscales

- Temporal Demand: How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?
- Frustration Level: How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?
- Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Although these measures require a manual data entry, they are extremely useful to validate the interaction-based performance. It must thus be made clear that we do not intend for a final version of a fatigue management tool to include such indicators. Nonetheless, at the moment we look at such indicators as a viable way to assess the validity of the developed approach, when both are used in parallel, as described further ahead in this paper.

2 FATIGUE MANAGEMENT AS A TOOL FOR IMPROVING LEARNING PERFORMANCE

As stated before, this paper presents a tool for monitoring fatigue in e-learning scenarios using non-

aimed at the development of better interaction mechanisms with computers.

The proposed approach distinguishes from existing work twofold: (1) the application area and (2) the features considered, as detailed further ahead in this section. The tool collects data bout the user's interaction with the computer and stores it in the form of a log. This log contains each particular interaction event, their timestamp and other important information such as coordinates or key code, when applicable.
The following events are considered:

- MOV, timestamp, posX, posY - an event describing the movement of the mouse, in a given time, to coordinates (posX, posY) in the screen;
- MOUSE_DOWN, timestamp, [Left—Right], posX, posY - this event describes the first half of a click (when the mouse button is pressed down), in a given time. It also describes which of the buttons was pressed (left or right) and the position of the mouse in that instant;
- MOUSE_UP, timestamp, [Left—Right], posX, posY - an event similar to the previous one but describing the second part of the click, when the mouse button is released;
- MOUSE_WHEEL, timestamp, dif - this event describes a mouse wheel scroll of amount dif, in a given time;
- KEY_DOWN, timestamp, key - identifies a given key from the keyboard being pressed down, at a given time;
- KEY_UP, timestamp, key - describes the release of a given key from the keyboard, in a given time;

From these events, that fully describe the interaction of the user with the mouse and keyboard, we extract a set of features, based on notions of behavioural biometrics:
KEY DOWN TIME - the timespan between two consecutive KEY_DOWN and KEY_UP events, i.e., for how long was a given key pressed.

TIME BETWEEN KEYS - the timespan between two consecutive KEY_UP and KEY_DOWN events, i.e., how long did the individual took to press another key.

VELOCITY - The distance travelled by the mouse (in pixels) over the time (in milliseconds). The velocity is computed for each interval defined by two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$, that took place respectively in the instants $time_1$ and $time_2$. Let us also assume two vectors *posx* and *posy*, of size *n*, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The velocity between the two clicks is given by $r_dist/(time_2 - time_1)$, in which *r_dist* represents the distance travelled by the mouse and is given by equation 1.

$$r_dist = \sum_{i=0}^{n-1} \sqrt{(posx_{i+1} - posx_i)^2 + (posy_{i+1} - posy_i)^2} \quad (1)$$

ACCELERATION - The velocity of the mouse (in pixels/milliseconds) over the time (in milliseconds). A value of acceleration is computed for each interval defined by two consecutive MOUSE_UP and MOUSE_DOWN events, using the intervals and data computed for the Velocity.

TIME BETWEEN CLICKS - the timespan between two consecutive MOUSE_UP and MOUSE_DOWN events, i.e., how long did it took the individual to perform another click.

DOUBLE CLICK DURATION - the timespan between two consecutive MOUSE_UP events, whenever this timespan is inferior to 200 milliseconds. Wider timespans are not considered double clicks.

AVERAGE EXCESS OF DISTANCE - this feature measures the average excess of distance that the mouse travelled between each two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. To compute this feature, first it is measured the distance in straight line between the coordinates of *mup* and *mdo* as $s_dist = \sqrt{(x2-x1)^2 + (y2-y1)^2}$. Then, it is measured the distance actually travelled by the mouse by summing the distance between each two consecutive MOUSE_MV events. Let us assume two vectors *posx* and *posy*, of size *n*, holding the coordinates of the consecutive MOUSE_MV events between *mup* and *mdo*. The distance actually travelled by the mouse, *real_dist* is given by equation 1. The average excess of distance between the two consecutive clicks is thus given by r_dist/s_dist.

AVERAGE DISTANCE OF THE MOUSE TO THE STRAIGHT LINE - in a few words, this feature measures the average distance of the mouse to the straight line defined between two consecutive clicks. Let us assume two consecutive MOUSE_UP and

MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. Let us also assume two vectors *posx* and *posy*, of size *n*, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The sum of the distances between each position and the straight line defined by the points $(x1, y1)$ and $(x2, y2)$ is given by 2, in which *ptLineDist* returns the distance between the specified point and the closest point on the infinitely-extended line defined by $(x1, y1)$ and $(x2, y2)$. The average distance of the mouse to the straight line defined by two consecutive clicks is this given by s_dists/n.

$$s_dists = \sum_{i=0}^{n-1} ptLineDist(posx_i, posy_i) \quad (2)$$

DISTANCE OF THE MOUSE TO THE STRAIGHT LINE - this feature is similar to the previous one in the sense that it will compute the s_dists between two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, according to equation 2. However, it returns this sum rather than the average value during the path.

SIGNED SUM OF ANGLES - with this feature the aim is to determine if the movement of the mouse tends to "turn" more to the right or to the left. Let us assume three consecutive MOUSE_MOVE events, $mov1, mov2$ and $mov3$, respectively in the coordinates $(x1, y1)$, $(x2, y2)$ and $(x3, y3)$. The angle α between the first line (defined by $(x1, y1)$ and $(x2, y2)$) and the second line (defined by $(x2, y2)$ and $(x3, y3)$) is given by $degree(x1, y1, x2, y2, x3, y3) = \tan(y3 - y2, x3 - x2) - \tan(y2 - y1, x2 - x1)$. Let us now assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*. Let us also assume two vectors *posx* and *posy*, of size *n*, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The signed sum of angles between these two clicks is given by equation 3.

$$s_angle = \sum_{i=0}^{n-2} degree(posx_i, posy_i, posx_{i+1}, posy_{i+1}, posx_{i+2}, posy_{i+2}) \quad (3)$$

ABSOLUTE SUM OF ANGLES - this feature is very similar to the previous one. However, it seeks to find only how much the mouse "turned", independently of the direction to which it turned. In that sense, the only difference is the use of the absolute of the value returned by function $degree(x1, y1, x2, y2, x3, y3)$, as depicted in equation 4.

$$s_angle = \sum_{i=0}^{n-2} \mid degree(posx_i, posy_i, posx_{i+1}, posy_{i+1}, posx_{i+2}, posy_{i+2}) \mid \quad (4)$$

DISTANCE BETWEEN CLICKS - represents the total distance travelled by the mouse between two consecutive clicks, i.e., between each two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. Let us also assume two vectors *posx* and *posy*, of size *n*, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The total distance travelled by the mouse is given by equation 1.

After the collection of the data, it is processed and converted into a set of behavioural biometric features, able to classify the behaviours of the user in terms of fatigue and level of attention to the task, described in detail in (Pimenta et al., 2015). This approach to assess performance has been developed, validated and used previously. However, after its development, we concluded that its most significant drawback is that it looks at performance of interaction alone. Fatigue is a complex phenomenon and performance measures by themselves may not be enough for an accurate measurement. In particular, and as stated before, mental workload is a very important aspect when it comes to characterizing fatigue.

We are thus now extending this approach with the acquisition of contextual information about the user, including the type of task being performed, the time spent on each different task, as well as the mental workload felt while performing the task (Figure 1). This means that this tool will now also be able to analyse the level of attention of a user to each task (e.g. distinguish between time spent on tasks related to e-learning activities against time spent on other tasks). In the overall, the new tool results in a more complete approach by including these important contextual factors.

Figure 2 depicts the process through which the system operates, where it is possible to observe the different classifications of information in order to allow, in the end, the management of fatigue. Initially the system captures the mouse and keyboard inputs. These data are further processed, stored and then used to calculate the values of the behavioural biometrics. In the learning phase the system shows a questionnaire in order to evaluate the subjective feeling of fatigue of the user, as well as the mental workload. When the system has a large enough dataset that allows to make classifications with precision, it will classify the inputs received into different mental fatigue and mental workload levels in real-time. At this

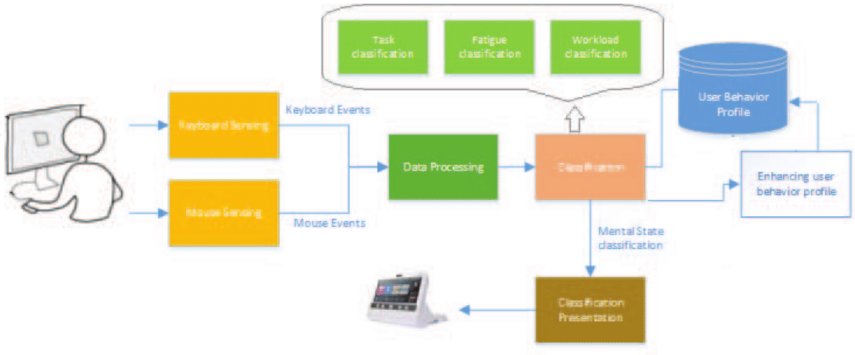


Figure 2: The flow of data in the fatigue management tool for e-learning scenarios.

point, the system can start to be used by the people involved, especially the teacher who can better adapt and personalize his teaching strategies.

3 CASE STUDY

In order to assess the validity of the approach described in the previous section a case study was implemented with the aim of collecting data over a period of time that encompassed different sessions of e-learning, and thus test if it is possible to monitor fatigue through the use of behavioural biometrics and mental workload.

For this purpose, twenty four students volunteered (19 men, 5 women), all students of the course of Physical Sciences at the University of Minho. Their age ranged between 18 and 30. Participants were provided with the application for logging the events of the mouse and keyboard during the duration of the class. This application started automatically in the background, when the Operating System started, requiring no specific action from the part of the students. The previously mentioned list of features was extracted from the use of the mouse and keyboard for the whole period.

3.1 Methodology

The methodology followed to implement the study was devised to be as minimally intrusive as the approach it aims to support. Participants were provided with an application for logging the previously mentioned events of the mouse and keyboard. This application, which maintained the confidentiality of the keys used, needed only to be installed in the participant's computer and would run in the background, starting automatically with the Operating System. The only explicit interaction needed from the part of the user was the input of very basic information on the first run, including the identification and age.

The course takes place physically in a classroom and comprises a teacher who is responsible for teaching a programming language (in this case MatLab) to a class of students. Each class has a duration of three hours, which always follows the same "protocol": some theoretical concepts are introduced at the beginning of the class and the rest of the session is spent practising and solving exercises using the computer and a specific IDE. During each session the system, while in a learning phase, presents the user with a questionnaire (Figure 3) based on the NASA TLX for measuring mental workload.

Thus, in each session all inputs resulting from the interaction of the user with the computer using the mouse and keyboard are collected, together with the subjective values of cognitive load, acquired from the NASA TLX.

3.2 Results

Using the data collected in the classroom over the two weeks, a classification model was trained based on the K-Nearest Neighbour (KNN) algorithm. It is a method of classification based on closest training samples in the feature space.

A model was built based on a dataset with 74 instances, each instance being constituted by the average values of all features during periods of one our. Each instance is also assigned a label, which represents the response of students to the questionnaire for measuring workload, provided while using the com-

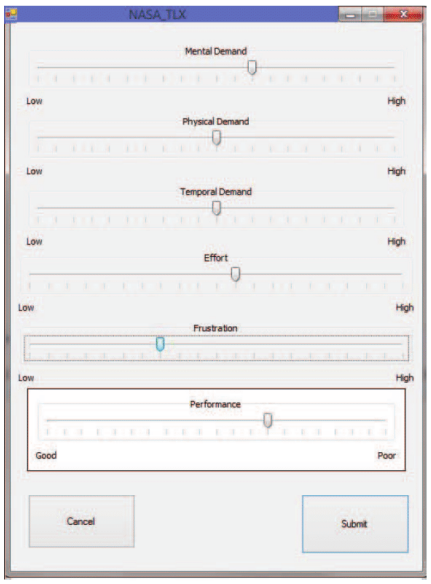


Figure 3: NASA TLX questionnaire used to collect information of mental workload.

puter during classes, thus in parallel with the collection process of the aforementioned data.

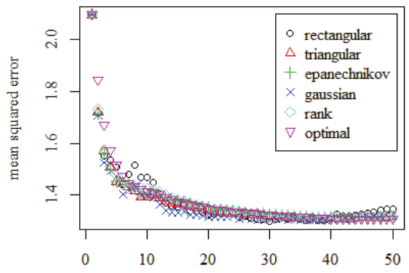


Figure 4: Results of different models trained with different kernels and number of neighbours (K).

Several tests with different numbers of neighbours (K) and with different heuristics to the distance between neighbouring (rectangular, triangular, epanachicov, gaussian, rank, optimal) were performed. With a maximum of 50 neighbours, the solution having a lower mean squared error (MSE), was found with K = 30 and using the rectangular kernel, as shown in Figure 4.

The trained model was then used to predict the mental workload in data collected in the second period, in a total of 78 instances resulting from the interaction of users with the computer. According to the classification carried out, 64 out of 78 (83%) of the instances were in accordance with the subjective opinion of the user about the mental workload of the task that was to perform, i.e., were correctly classified. It is also important to note that the remaining 14 instances (17%) were classified as adjacent values.

Table 1: Results of the validation of the classification model (KNN). 83% of the instances were correctly classified (green cells). The 17% misclassified instances where nonetheless classified as neighboring values (red cells).

		PREDICTED						
N		1	2	3	4	5	6	7
A	1	13	3	0	0	0	0	0
S	2	2	11	1	0	0	0	0
A	3	0	1	12	2	0	0	0
	4	0	0	2	15	0	0	0
T	5	0	0	0	0	7	1	0
L	6	0	0	0	0	1	6	0
X	7	0	0	0	0	0	0	1

4 CONCLUSIONS

This paper describes a prototype of a tool for managing fatigue. Its main innovative aspect is that, for the first time, it considers the mental workload of a user while performing a task as an important component of fatigue assessment. The main objective is to detect patterns of behavior at different levels of mental workload. Measurement of levels of mental workload are obtained through the NASA TLX instrument, which is based on a subjective self-evaluation. These subjective measures, paired with measures of performance and context of the task being performed by a user, allow to train a classifier as the one depicted which achieved fairly good results. In the described case study, the tool was used in several classes during the period of two weeks, which allowed not only to test it in a real scenario .

The results achieved from the implementation of the case study show that it is indeed possible to analyse and quantify mental workload through the use of the mouse and keyboard, and this allows not only to measure cognitive load but also to improve the process of monitoring mental fatigue.

Although at the moment we aim to support the teacher's decision making process, the long-term goal of this work is to develop environments that are

autonomous and take actions concerning their self-management. These actions will be guided by several objectives, one of them being to manage cognitive load, minimize fatigue and increase performance and well-being of an individual or group of individuals through an appropriate selection of tasks and task durations.

ACKNOWLEDGEMENTS

This work is part-funded by ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-028980 (PTDC/EEI-SII/1386/2012) and project PEst-OE/EEI/UI0752/2014.

REFERENCES

Balkin, T. and Wesensten, N. (2011). Differentiation of sleepiness and mental fatigue effects. (2004):47–66.

Brown, I. D. (1994). Driver fatigue. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36(2):298–314.

Cantoni, V., Cellario, M., and Porta, M. (2004). Perspectives and challenges in e-learning: towards natural interaction paradigms. *Journal of Visual Languages & Computing*, 15(5):333–345.

Card, S. K., Moran, T. P., and Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, 23(7):396–410.

Hamburg, I., Engert, S., Anke, P., Marin, M., and im IKM Bereich, E.-C. A. (2008). Improving e-learning 2.0-based training strategies of smes through communities of practice. *learning*, 2:610–012.

Hart, S. and Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*.

Miller, J. C. (2013). *Anatomy of a Fatigue-Related Accident*. Shiftwork, Fatigue and Safety, Book 3.

Perelli, L. (1980). Fatigue Stressors in Simulated Long-Duration Flight. Effects on Performance, Information Processing, Subjective Fatigue, and Physiological Cost. (March 1977).

Pimenta, A., Carneiro, D., Neves, J., and Novais, P. (2014). A non-invasive approach to detect and monitor acute mental fatigue. In Ali, M., Pan, J.-S., Chen, S.-M., and Horng, M.-F., editors, *Modern Advances in Applied Intelligence*, volume 8482 of *Lecture Notes in Computer Science*, pages 338–347. Springer International Publishing.

Pimenta, A., Carneiro, D., Novais, P., and Neves, J. (2015). Detection of distraction and fatigue in groups through the analysis of interaction patterns with computers. In Camacho, D., Braubach, L., Venticinque, S., and Badica, C., editors, *Intelligent Distributed Computing VIII*, volume 570 of *Studies in Computational Intelligence*, pages 29–39. Springer International Publishing.

Reid, G. B., Eggemeier, F. T., and Shingledecker, C. A. (1982). Subjective Workload Assessment Technique. Technical report.

van der Linden, D., Frese, M., and Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, 113(1):45–65.

Williamson, R. J., Purcell, S., Sterne, A., Wessely, S., Hotopf, M., Farmer, A., and Sham, P. C. (2005). The relationship of fatigue to mental and physical health in a community sample. *Social psychiatry and psychiatric epidemiology*, 40(2):126–32.

3 CONCLUSIONS

We have discussed different aspects of mental fatigue. We have also examined a number of methods used to detect fatigue as well as monitoring systems. Much of the work in this field has been conducted with the goal of helping critical tasks, particularly in highway driving and aviation. For example, fatigued driving is one of the most significant causes of traffic accidents, but physiological indicators such as PERCLOS appear to be potentially useful indicators of drowsiness.

Undeniably, the causes of mental fatigue can be as simple as overworking the brain or making untoward lifestyle choices. Indeed, continuously engaging in a difficult task for too long is a typical cause of mental fatigue, as well as not attaining enough sleep, lack of exercise, poor diet, and dehydration.

Most of the relevant research to date has been based upon a set of features that do not consider individual differences and the environment, and are often intrusive and/or invasive, and only few systems that take advantage of the use of AI, and those who use it, use for image processing or EEG signal processing. This was one of our research focuses throughout the doctoral program, the detection of mental fatigue using non-invasive methods and AI that resulted in a set of contributions in the development of techniques and mechanisms for the detection of fatigue.

In this chapter it is provided an analysis of one`s contributions in light with the objectives outlined initially. This chapter starts with a description on how the research hypothesis were formulated and how they were validated. There is also a summary of the activities that were undertaken for the dissemination of results. The chapter ends with final remarks regarding the work that was developed so far and with future work reflections.

3.1 Contributions

The present doctoral thesis is offered as a set of publications that encompass significant contributions in the use of machine learning algorithms to process human computer interactions in order to detect mental fatigue in a non-invasive way. It also reports the steps pursued in the creation of a system that is capable to detect mental fatigue and assess human performance based on the different forms that one may use when interacts with a computer.

As a complement to this section, Table 1 shows the thesis' objectives, as they were defined in Section 1.6, and the sections of Chapter 2 where they were achieved. The full list of contributions is now put forward, viz.

- **Identification of different methods for fatigue detection and existing solutions**

The first contribution, presented in Chapter 1, is the analysis and summary of all information related to the detection of fatigue, the different methods, as well as the respective advantages and disadvantages. A definition of fatigue and all its characteristics is also presented, since it is regarded by the scientific community as subjective.

It has become clear that most current systems use EEG or image analysis for intermittent processing. Other evidence is that the current focus is in areas where tasks are critical (e.g., mining industry, aviation) for the purpose of saving lives. These are some of the reasons why non-invasive techniques have not been explored in depth by the scientific community and industry.

- **Analysis of user interaction patterns related to different mental states and human performance on computer usage**

With a clear definition of what fatigue is and how it can be obtained, we used behavioral biometric information to assess human performance in computer usage, viewing mental fatigue as it is described in section 2.3. In order to reinforce and prove this statement, some studies were performed with fatigue and non-fatigue scenarios, a way that allow one to address this subject in depth.

This permitted us to prove that behavioral biometrics may be the key to assess mental fatigue and measure its impact on human behavior, and in this specific case on computer usage.

- **Creation of machine learning models capable of understand different interaction patterns, classifying them in fatigue or non-fatigue, as well as identifying the intensity of fatigue.**

Proved the possibility of measuring the loss of human performance through behavioral biometrics, in particular through Keystroke Dynamics and Mouse Dynamics with the evolution of our circadian rhythm or with the occurrence of fatigue, the studies presented in section 2.1 and 2.2 revealed the possibility to identify not only possible fatigue scenarios but also behavior changes.

This highlights the importance of the circumstance or context, where it was possible to verify that computer interaction patterns are exposed to large variations when we perform different types of tasks (e.g. writing a report or programming), which should be considered one of the most important aspects in the behavioral biometry analysis.

- **Implementation of a distributed intelligent architecture**

The process of capturing, collecting and processing large amounts of biometric information coming from different endpoints as raw data or processed pre-data and in real time was one of the challenges of this work. These were some of the requirements that were considered in the conception of the architecture described in section 2.4. It enables the use of modular

services in a simple way, i.e., it is very simple to add different data sources and computational models to process different data sources.

- **Development of a prototype capable of using machine learning models to detect mental fatigue in different contexts**

It was settled an archetype qualified to measure levels of mental fatigue on computer usage in real-time. It is based on the architect referred to above, and it served to model user behavior on computer usage, as well as to study different scenarios focused on the impact of mental fatigue management as presented in section 2.5 and 2.7.

- **Validation of the classification models for mental fatigue assessment.**

The models' analysis and validation were one of the most difficult and important tasks of this work. Due to its complexity, we chose to use questionnaires and statistical methods to validate the models created for the different scenarios and take advantage of the data to improve the supervised ones. Another important contribution was the fact that we have proved that fatigue has an impact on biometric identification methods for security purposes, as demonstrated the section 2.6.

3.2 Validation of the Research Hypothesis

In order to validate the hypothesis presented in section 1.5, it was necessary to demonstrate that it is possible to identify different mental states and assess human performance from the interaction patterns in the use of the computer. There are some techniques that have been used to identify the user in the computer usage in the security field that are based on the analysis of interaction patterns (biometric authentication). Based on these techniques, and on the premise that it is possible to create a user profile through

the behavioral biometrics, it was necessary to prove that such a signature suffers small changes with onset of fatigue and on different mental states.

Starting with the analysis of the human performance on the computer usage, we have been able to prove that during the day, the human behavioral change when performing different tasks or face different mental states. These differences are responsible for failures and less accuracy in some user authentication and on systems` identification based on behavioral biometrics. Mental fatigue is one of the causes of these failures, which led us to conclude that the onset and presence of mental fatigue is possible to observe through variations in the interaction patters with the computer on a deep level.

Considering these behavioral changes, the digital signature, and the characteristics of the machine itself, such as the layout of the computer or keyboard mechanism, machine learning methods were used to process the data from self-awareness questionnaires about the level of fatigue to prove that is possible to detect different levels of fatigue from the computer usage, and be specific for each user. It was therefore possible to create individual and general models with high accuracy. Taking this into account, it is believed that the research hypothesis was proven.

3.3 Dissemination of Results

3.3.1 Other Publications

Besides the publications included in the main body of the thesis, other articles were also published in international journals and scientific meetings in fields like Computer Science, Artificial Intelligence or Human Computer Interaction.

Carneiro D., Araújo D., Pimenta A., Novais P., Real Time Analytics for Characterizing the Computer User's State, Advances in Distributed Computing and Artificial Intelligence Journal (ADCAIJ), ISSN: 2255-2863, Volume 5, issue nº4, http://dx.doi.org/10.14201/ADCAIJ201654118, pp 1-18, 2016.

Carneiro D., Pimenta A., Gonçalves S., Neves J., Novais P., Monitoring and improving performance in human–computer interaction, Concurrency and Computation: Practice and Experience, ISSN: 1532-0634, Vol. 28, nº4, pp 1291-1309, 2016. http://dx.doi.org/10.1002/cpe.3635

Pimenta A., Carneiro D., Novais P., Neves J., Analysis of Mental Fatigue and Mood States in Workplaces, Intelligent Distributed Computing IX, Springer-Verlag - Studies in Computational Intelligence, Paulo Novais David Camacho, Cesar Analide, Amal El Fallah Seghrouchni and Costin Badica (Eds), Vol. 616, pp 111-119, ISBN: 978-3-319-25015-1, 2016. http://dx.doi.org/10.1007/978-3-319-25017-5_11

Araújo D., Pimenta A., Carneiro D., Novais P., Providing Wellness Services Using Real Time Analytics, Ambient Intelligence- Software and Applications – 7th International Symposium on Ambient Intelligence (ISAmI 2016), Springer - Advances in Intelligent Systems and Computing, Lindgren H. et al. (eds), Vol 476, ISSN 2194-5357, ISBN 9783319401133, pp 167-175, 2016. http://dx.doi.org/10.1007/978-3-319-40114-0_19.

Pimenta A., Carneiro D., Novais P., Neves J., A Discomfort-Sensitive Chair for Pointing Out Mental Fatigue, Ambient Intelligence - Software and Applications, 6th International Symposium on Ambient Intelligence (ISAmI 2015), Springer - Advances in Intelligent Systems and Computing, Vol 376, ISSN 2194-5357, ISBN 978-3-319-19694-7, pp 57-64, 2015. http://dx.doi.org/10.1007/978-3-319-19695-4_6

Pimenta A., Carneiro D., Novais P., Neves J., Detection of Distraction and Fatigue in Groups through the Analysis of Interaction Patterns with Computers, Intelligent Distributed Computing VIII, Springer-Verlag - Studies in Computational Intelligence, David Camacho, Lars Braubach, Salvatore Venticinque and Costin Badica (Eds) Vol. 570, pp 29-39, ISBN: 978-3-319-10421-8, 2014. http://dx.doi.org/10.1007/978-3-319-10422-5_5

Pimenta A., Carneiro D., Novais P., Neves J., A Non-invasive Approach to Detect and Monitor Acute Mental Fatigue, Modern Advances in Applied Intelligence - 27th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems - IEA/AIE 2014, Moonis Ali, Jeng-Shyang Pan, Shyi-Ming Chen, Mong-Fong Horng (Eds.), Springer-Verlag, Lecture Notes in Computer Science, Vol. 8482, pp 338-347, ISBN 978-3-319-07466-5, 2014. http://dx.doi.org/10.1007/978-3-319-07467-2_36

Gomes M., Carneiro D., Pimenta A., Nunes M., Novais P., Neves J., Improving Modularity, Interoperability and Extensibility in Ambient Intelligence Ambient Intelligence- Software and Applications – 5th International Symposium on Ambient Intelligence (ISAmI 2014), Carlos Ramos, Paulo Novais, Céline Nihan and Juan Corchado (Eds), Springer - Series Advances in Intelligent and Soft Computing, Vol. 291, ISBN 978-3 319-07595-2, pp 63-70, 2014. http://dx.doi.org/10.1007/978-3-319-07596-9_7

Carneiro D., Novais P., Catalão F., Marques J., Pimenta A., Neves J., Dynamically Improving Collective Environments through Mood Induction Procedures, Ambient Intelligence- Software and Applications – 4th International Symposium on Ambient Intelligence (ISAmI 2013), Ad van Berlo, Kasper Hallenborg, Juan M. Corchado, Dante I. Tapia, Paulo Novais (eds), Springer - Series Advances in Intelligent and Soft Computing, Vol 219, ISBN 978-3-319-00565-2, pp 33-40, 2013. http://dx.doi.org/10.1007/978-3-319-00566-9_5

3.3.2 Collaboration and Participation in Scientific Reviews and Events

Throughout the doctoral programme there were opportunities to participate in international scientific events. These opportunities were important in order to acquire new experiences and exchange ideas. Here is a list of the most significative ones, viz.

- 8th and 9th International Symposium on Intelligent Distributed Computing (IDC): IDC is a symposium focused on the emergent field of Intelligent Distributed Computing and the development of a new generation of intelligent distributed systems. The doctoral candidate participated in the 2015 edition of the symposium as a member of the organizing committee.
- 5th and 6th International Symposium on Ambient Intelligence the International Symposium on Ambient Intelligence and aiming to bring together researchers from various disciplines that constitute the scientific field of Ambient Intelligence to present and discuss the latest results, new ideas, projects and lessons learned.
- 17th Portuguese Conference on Artificial Intelligence (EPIA) is a well-established international conference on Artificial Intelligence, supported by Portuguese Association for Artificial Intelligence (APPIA). The purpose of the conference is to promote research in AI and scientific exchange among AI researchers, practitioners, scientists, and engineers in related disciplines.

Editorial Board Members and Collaborations:

- Invited Reviewer (2014)- IET Image Processing;
- Invited Reviewer (2015) - SOJ Nursing & Health Care (SOJNHC);
- Invited Reviewer (2015) – Journal of Computer Engineering and Information Technology (SciTechnol);
- Editorial Board Member (2015-2016) at International Journal of Computer & Software Engineering (IJCSE); and
- Invited Reviewer (2017) - Data in Brief.

3.3.3 Supervision of Students

During the doctoral project, the candidate successfully co-supervised master students at the University of Minho that developed work in the context of this thesis. All these works

were supervised by Professor Paulo Jorge Freitas de Oliveira Novais. The list of students and their respective dissertation titles are the following, viz.

- **Daniel Araujo (2015) –** Masters in Computer Science, Dissertation : “Real Time Intelligence”;
- **Ana Quintas (2016) –** Masters in Biomedical Engineering, Dissertation: **“**Unsupervised learning of human-computer interaction patterns”;
- **Priscila Gomes (2017) -** Masters in Biomedical Engineering, Dissertation: “Profiling Mental Fatigue through Keystrokes and Mouse Dynamics”;
- **Paulo Cardoso (2018)** - Masters in Computer Science, Dissertation: “Monitoring Attention and Performance on Critical Situations”.

3.3.4 Entrepreneurship

The result of all the research served as an incentive to entrepreneurship in order to develop the potential of the line of research followed during the doctoral program where the candidate was responsible for setting up a startup called Performetric - a real-time monitoring system that allows the management of mental fatigue in a non-invasive way that aims to develop leisure and work context-aware environments that may improve the quality of life, wellness, mental health and individual performance.

3.3.5 Awards and Recognition

During the development of this thesis the candidate saw his work being awarded, as a distinction of the quality of work, having got the distinctions, viz.

- Best Student Paper Award in the area of Mobile and Pervasive Computing with article: “Mental Workload Management as a Tool in e-Learning Scenarios” in the 5th

International Joint Conference on Pervasive and Embedded Computing and Communication Systems;

- Honorable Mention in the area of Intelligent Distributed and High-Performance Architectures with the article: “Analysis of Mental Fatigue and Mood States in Workplaces “in the 5th International Symposium on intelligent distributed Computing.

In addition to these awards, the work developed was cited and presented as a potential solution in different areas, from the military area, medicine or computer science, with emphasis on the following works, viz.

- “Workstation Analytics in Distributed Warfighting Experimentation: Results from Coalition Attack Guidance Experiment 3A” of Defense Research and Development Canada Ottawa (Ontario) Centre For Operational Research and Analysis;
- PRISM: A Data-Driven Platform for Monitoring Mental Health. From Stanford University;
- Characterizing and predicting mental fatigue during programming tasks from Microsoft and North of Carolina University.

3.4 Final Remarks and Future Work Considerations

With this doctoral thesis it was possible to investigate three distinct areas, which intersect when we look at the human being, and his daily life as the center of this research. The use of AI, the analysis of human-computer interaction, in order to analyze and classify a set of different mental states through non-invasive techniques.

This was a great challenge because, despite the increasingly common use of AI techniques and a large availability of sophisticated sensors, the community and the industry still use some old ones, such as the case of PERCELOS processing or EEG analysis. This happens because the detection of fatigue is a challenge due to its own structure, but also because it is a mental

state difficult to classify and sometimes not visible, and for this reason, in critical scenarios are used intrusive techniques, once they allow for the detection of fatigue with some precision, and due to complexion, non-intrusiveness is not a priority.

These challenges together with the possibility of creating systems capable of helping the human being's daily life were the great motivation behind this work, and that allowed to realize a set of experiences based on the use of machine learning models for the interpretation of HCI patterns affected by mental fatigue, as well as the creation of a system capable of improving the decision-making process through the knowledge of such state. Last but not least, it also served to alert the community to the potential of using HCI associated machine learning techniques in different existing processes which have not yet been explored.

However, as described throughout this thesis, fatigue is a subjective theme and is affected by innumerable different factors, such as sleep, mood, or even with the physical presence someone else. These were some of the difficulties encountered throughout this work and that require research in the field of psychology. For this same reason, a study was started with the University of Groningen that aims to analyze the patterns of brain activity recorded under fatigue scenarios, while correlating with some different behavioral biometrics, allowing, for example, identifying the weight / correlation coefficient of each biometric, but also to progress in the development of communication interfaces with the brain, an emerging area with enormous potential.

Finally, as a future work should be taken into account the use of information related to fatigue for building recommendation systems, since the monitoring may be important in cases already reported to the military scenario, or control rooms, but in common cases this type of information is difficult to read or extract for immediate action.

REFERENCES

Admi, H., Tzischinsky, O., Epstein, R., Herer, P., & Lavie, P. (2008). Shift Work in Nursing: Is it Really a Risk Factor for Nurses' Health and Patients' Safety? *Nursing Economic$*, *26*(4), 250–257. Retrieved from http://europepmc.org/abstract/MED/18777974

Ahonen, L., & Cowley, B. (2016). A short review and primer on electroencephalography in human computer interaction applications. *Foundations and Trends® in Human–Computer Interaction*, *9*(3–4), 151–308. https://doi.org/10.1561/1100000065

Åhsberg, E. (1998). Percieved fatigue related to work. *Arbete Och Hälsa*, *1998:19*.

Åhsberg, E., Gamberale, F., & Gustafsson, K. (2000). Perceived fatigue after mental work: An experimental evaluation of a fatigue inventory. *Ergonomics*, *43*(2), 252–268. https://doi.org/10.1080/001401300184594

Åkerstedt, T. (2003). Shift work and disturbed sleep/wakefulness. *Occupational Medicine*, *53*(2), 89–94. https://doi.org/10.1093/occmed/kqg046

Åkerstedt, T., & Folkard, S. (1997). The three-process model of alertness and its extension to performance, sleep latency, and sleep length. *Chronobiology International*, *14*(2), 115–123. https://doi.org/10.3109/07420529709001149

Åkerstedt, T., Folkard, S., & Portin, C. (2004). Predictions from the three-process model of alertness. *Aviation Space and Environmental Medicine*, *75*(SUPPL.1), A75-83. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15018267

Akerstedt, T., & Gillberg, M. (1990). SubjecAkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individABelyavin, a, & Wright, N. a. (1987). Changes in electrical activity of the brain with vigilance. Electroencephalography and Clinical Neurophysiology, 66(. *The International Journal of Neuroscience*, *52*(1–2), 29–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2265922

Ames, L. L., George, E. J., Air, E., & Base, F. (1993). Revision and Verification of a Seven-Point

Workload Estimate Scale. Retrieved from http://www.dtic.mil/docs/citations/ADA269194

Banbury, S., Dudfield, H., Hoermann, H. J., & Soll, H. (2007). FASA: Development and validation of a novel measure to assess the effectiveness of commercial airline pilot Situation Awareness training. *International Journal of Aviation Psychology*, *17*(2), 131–152. https://doi.org/10.1080/10508410701328557

Barr, L., Howarth, H., Popkin, S., & Carroll, R. J. (2009). An Evaluation of Emerging Driver Fatigue Detection Measures and Technologies. *Reproduction*, *TP 14620E*(June), 1–27. Retrieved from http://www.ecse.rpi.edu/~qji/Fatigue/fatigue_report_dot.pdf

Bartlett F. (1953). Psychological criteria of fatigue. *Symposium on Fatigue*, 1–5.

Bekiaris, E., Nikolaou, S., Panou, M., Van Wees, K., & Baten, G. (2002). System for effective Assessment of driver vigilance and Warning According to traffic risK Estimation Table of Contents. *Evaluation*, 1–264.

Belyavin, A. J., & Spencer, M. B. (2004). Modeling performance and alertness: The QinetiQ approach. *Aviation Space and Environmental Medicine*, *75*(SUPPL.1), A93-103-106. Retrieved from http://www.ingentaconnect.com/content/asma/asem/2004/00000075/A00103s1/art00016

Benoit, O. (1980). *The functions of sleep*. *Neuropsychologia* (Vol. 18). Yale University Press. https://doi.org/10.1016/0028-3932(80)90168-2

Bergasa, L. M., & Nuevo, J. (2005). Real-time system for monitoring driver vigilance. *IEEE International Symposium on Industrial Electronics*, *III*(1), 1303–1308. https://doi.org/10.1109/ISIE.2005.1529113

Bills, A. G. (1931). Blocking: A New Principle of Mental Fatigue. *The American Journal of Psychology*, *43*(2), 230. https://doi.org/10.2307/1414771

Bills, A. G. (1937). Blocking in mental fatigue and anoxemia compared. *Journal of Experimental Psychology*, *20*(5), 437–452. https://doi.org/10.1037/h0058001

Bittner, A. C., Carter, R. C., Krause, M., Kennedy, R. S., & Harbeson, M. M. (1983). Performance

evaluation tests for environmental research (PETER): Moran and computer batteries. *Aviation Space and Environmental Medicine*, *54*(10), 923–928. Retrieved from http://europepmc.org/abstract/MED/6651716

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology*, *72*(2), 123–132. https://doi.org/10.1016/j.biopsycho.2005.08.007

Bonnet, M. H. (1980). Sleep, Performance and Mood After the Energy-Expenditure Equivalent of 40 Hours of Sleep Deprivation. *Psychophysiology*, *17*(1), 56–63. https://doi.org/10.1111/j.1469-8986.1980.tb02460.x

Borbély, A. (1982). A two process model of sleep regulation. *Human Neurobiology*, *1*(3), 195–204. https://doi.org/10.1111/jsr.12371

Borbely, A. A., & Achermann, P. (1999). Sleep homeostasis and models of sleep regulation [see comments]. *J. Biol. Rhythms*, *14*(6), 557–568. https://doi.org/http://dx.doi.org/10.1016/B978-1-4160-6645-3.00037-2

Broughton, R., Dunham, W., Newman, J., Lutley, K., Duschesne, P., & Rivers, M. (1988). Ambulatory 24 hour sleep-wake monitoring in narcolepsy-cataplexy compared to matched controls. *Electroencephalography and Clinical Neurophysiology*, *70*(6), 473–481. https://doi.org/10.1016/0013-4694(88)90145-9

Broughton, R., & Mullington, J. (1992). Circasemidian sleep propensity and the phase-amplitude maintenance model of human sleep/wake regulation. *Journal of Sleep Research*, *1*(2), 93–98. https://doi.org/10.1111/j.1365-2869.1992.tb00017.x

Brown, I. D. (1964). The measurement of perceptual load and reserve capacity. *Occupational Medicine*, *14*(1), 44–49. https://doi.org/10.1093/occmed/14.1.44

Brown, S. W., & Swartz, T. A. (1989). A Gap Analysis of Professional Service Quality. *Journal of Marketing*, *53*(2), 92. https://doi.org/10.2307/1251416

Calfee, S. H., & Rowe, N. C. (2004). Multi-Agent Simulation of Human Behavior in Naval Air Defense. *Naval Engineers Journal*, *116*(4), 53–64. https://doi.org/10.1111/j.1559-3584.2004.tb00304.x

Caraty, M. J., & Montacié, C. (2014). Vocal fatigue induced by prolonged oral reading: Analysis and detection. *Computer Speech and Language*, *28*(2), 453–466. https://doi.org/10.1016/j.csl.2012.12.003

Casali, J. G., & Wierwille, W. W. (1983). A Comparison of Rating Scale, Secondary-Task, Physiological, and Primary-Task Workload Estimation Techniques in a Simulated Flight Task Emphasizing Communications Load. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *25*(6), 623–641. https://doi.org/10.1177/001872088302500602

Chaffin, D. B. (1973). Localized muscle fatigue-definition and measurement. *Journal of Occupational and Environmental Medicine*, *15*, 346–354. Retrieved from http://journals.lww.com/joem/Abstract/1973/15040/Localized_Muscle_Fatigue___Definition_and.4.aspx

Cohen, S. (1980). Aftereffects of stress on human performance and social behavior: A review of research and theory. *Psychological Bulletin*, *88*(1), 82–108. https://doi.org/10.1037/0033-2909.88.1.82

Cote, J. N., Feldman, A. G., Mathieu, P. a, & Levin, M. F. (2008). Effects of fatigue on intermuscular coordination during repetitive hammering. *Motor Control*, *12*(2), 79–92. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18483444

Curran, S. L., Andrykowski, M. A., & Studts, J. L. (1995). Short form of the Profile of Mood States (POMS-SF): Psychometric information. *Psychological Assessment*, *7*(1), 80–83. https://doi.org/10.1037/1040-3590.7.1.80

Cyganek, B., & Gruszczyński, S. (2014). Hybrid computer vision system for drivers' eye recognition and fatigue monitoring. *Neurocomputing*, *126*, 78–94. https://doi.org/10.1016/j.neucom.2013.01.048

Dawson, D., & Fletcher, A. (2001). A quantitative model of work-related fatigue: Background and definition. *Ergonomics*, *44*(2), 144–163. https://doi.org/10.1080/00140130119399

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, *17*(6), 652–655. https://doi.org/10.3758/BF03200977

Eddy, D. R., & Hursh, S. R. (2001). *Fatigue Avoidance Scheduling Tool* (No. 2001–140). *Phase I SBIR Final Report under Contract F41624-99-C-6041* (Vol. 5104). Brooks AFB, TX: Air Force Research Laboratory.

Eddy, D. R., & Hursh, S. R. (2006a). *Fatigue Avoidance Scheduling Tool (FAST) Phase II SBIR Final Report, Part 1* (No. 2006–15). Brooks City-Base, TX: Air Force Research Laboratory.

Eddy, D. R., & Hursh, S. R. (2006b). *Fatigue Avoidance Scheduling Tool (FAST) Phase II SBIR Final Report, Part 1* (No. 2006–0040, ADA452991). Brooks City-Base, TX: Air Force Research Laboratory.

Edenborough, N., Hammoud, R., Harbach, A., Ingold, A., Kisačanin, B., Malawey, P., … Zhang, H. (2005). Driver State Monitor from DELPHI. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 2, pp. 1206–1207). https://doi.org/10.1109/CVPR.2005.135

Edwards, R. H., Hill, D. K., Jones, D. A., & Merton, P. A. (1977). Fatigue of long duration in human skeletal muscle after exercise. *Journal of Physiology*, *272*(3), 769–78. https://doi.org/10.1113/jphysiol.1977.sp012072

Englund, C. E., Reeves, D. L., Shingledecker, C. A., Thorne, D. R., & Wilson, K. P. (1987a). *Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB). 1. Design and Specification of the Battery*.

Englund, C. E., Reeves, D. L., Shingledecker, C. A., Thorne, D. R., & Wilson, K. P. (1987b). *Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB). 1. Design and Specification of the Battery*.

Enoka, R. M., & Stuart, D. G. (1992). Neurobiology of muscle fatigue. *Journal of Applied Physiology*, *72*(5), 1631–1648. https://doi.org/0161-7567/92

Faber, L. G., Maurits, N. M., & Lorist, M. M. (2012). Mental Fatigue Affects Visual Selective Attention. *PLoS ONE*, *7*(10), e48073. https://doi.org/10.1371/journal.pone.0048073

Fletcher, A., & Dawson, D. (2001). A quantitative model of work-related fatigue: Empirical evaluations. *Ergonomics*, *44*(5), 475–488. https://doi.org/10.1080/00140130119824

Folkard, S., Lombardi, D. A., & Tucker, P. T. (2005). Shiftwork: Safety, Sleepiness and Sleep.

Industrial Health, *43*(1), 20–23. https://doi.org/10.2486/indhealth.43.20

Folkard, S., & Tucker, P. (2003). Shift work, safety and productivity. *Occupational Medicine*, *53*(2), 95–101. https://doi.org/10.1093/occmed/kqg047

Gartner, W. B., & Murphy, M. R. (1977). Pilot workload and fatigue: A critical survey of concepts and assessment techniques. *Applied Ergonomics*, *8*(4), 238. https://doi.org/10.1016/0003-6870(77)90183-1

Gevins, A., Leong, H., Du, R., Smith, M. E., Le, J., DuRousseau, D., … Libove, J. (1995). Towards measurement of brain function in operational environments. *Biological Psychology*, *40*(1–2), 169–186. https://doi.org/10.1016/0301-0511(95)05105-8

Gevins, A. S., Bressler, S. L., Cutillo, B. A., Illes, J., Miller, J. C., Stern, J., & Jex, H. R. (1990). Effects of prolonged mental work on functional brain topography. *Electroencephalography and Clinical Neurophysiology*, *76*(4), 339–350. https://doi.org/10.1016/0013-4694(90)90035-I

Gevins, A., & Smith, M. E. (1999). Detecting transient cognitive impairment with EEG pattern recognition methods. *Aviation Space and Environmental Medicine*, *70*(10), 1018–1024. Retrieved from http://www.eeg.com/Detecting Transient Cognitive.pdf

Gillberg, M., Kecklund, G., & Akerstedt, T. (1994). Relations between performance and subjective ratings of sleepiness during a night awake. *Sleep*, *17*(3), 236–241. https://doi.org/10.1093/sleep/17.3.236

Gilliland, K., & Schlegel, R. E. (1993). *Readiness to Perform Testing: A Critical Analysis of the Concept and Current Practices* (No. ADA269379). Norman OK: University of Oklahoma.

Goldstein, R., Bauer, L. O., & Stern, J. A. (1992). Effect of task difficulty and interstimulus interval on blink parameters. *International Journal of Psychophysiology*, *13*(2), 111–117. https://doi.org/10.1016/0167-8760(92)90050-L

Grace, R., Byrne, V. E., Bierman, D. M., Legrand, J.-M., Gricourt, D., Davis, B. K., … Carnahan, B. (1998). A drowsy driver detection system for heavy vehicles. In *17th DASC. AIAA/IEEE/SAE. Digital Avionics Systems Conference. Proceedings (Cat. No.98CH36267)* (Vol. 2, p. I36/1-I36/8). https://doi.org/10.1109/DASC.1998.739878

Grace, R., & Steward, S. (2001). Drowsy driver monitor and warning system. In *International driving symposium on human factors in driver assessment, training and vehicle design* (pp. 64–69). https://doi.org/10.1.1.21.3891

Greeley, H. P., Friets, E., Wilson, J. P., Raghavan, S., Picone, J., & Berg, J. (2007). Detecting fatigue from voice using speech recognition. *Sixth IEEE International Symposium on Signal Processing and Information Technology, ISSPIT*, 567–571. https://doi.org/10.1109/ISSPIT.2006.270865

Hammarskjöld, E., & Harms-Ringdahl, K. (1992). Effect of arm-shoulder fatigue on carpenters at work. *European Journal of Applied Physiology and Occupational Physiology*, *64*(5), 402–409. https://doi.org/10.1007/BF00625058

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*(C), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hartley, L., Horberry, T., Mabbott, N., & Krueger, G. P. (2000). *Review of Fatigue Detection and Prediction Technologies* (No. ISBN 0642544697). Australia: National Road Transport Commission. Retrieved from www.nrtc.gov.au

Hartman, B. O. (1967). Psychological factors in flying fatigue(Psychological factors of acute, cumulative and chronic flying fatigue). *In- Psychiatry in Aerospace Medicine. Edited By C. J. G. Perry.*

Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems*, *13*, 18–28. https://doi.org/10.1109/5254.708428

Hockey, G. R. J. (1983). *Stress and fatigue in human performance.* (Vol. 3). John Wiley & Sons Inc.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of Sleepiness: A New Approach. *Psychophysiology*, *10*(4), 431–436. https://doi.org/10.1111/j.1469-8986.1973.tb00801.x

Hursh, S. R., Balkin, T. J., Miller, J. C., & Eddy, D. R. (2004a). The Fatigue Avoidance Scheduling Tool: Modeling to Minimize the Effects of Fatigue on Cognitive Performance. *SAE Transactions*, *113*(1), 111–119. https://doi.org/10.4271/2004-01-2151

Hursh, S. R., Balkin, T. J., Miller, J. C., & Eddy, D. R. (2004b, June). *The Fatigue Avoidance Scheduling Tool: Modeling to Minimize the Effects of Fatigue on Cognitive Performance*. Rochester MI. https://doi.org/10.4271/2004-01-2151

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., … Eddy, D. R. (2004). Fatigue models for applied research in warfighting. *Aviation Space and Environmental Medicine*, *75*(SUPPL.1), A44-53-60.

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., … Storm, W. F. (2004, May). *The DOD Sleep, Acitivity, Fatigue, and Task Effectiveness Model*. *Aviation, space, and environmental medicine*. Mesa, AZ.

Jean-Louis, G., von Gizycki, H., Zizi, F., Hauri, P., Spielman, A., & Taub, H. (1997). The Actigraph Data Analysis Software: I. A Novel Approach to Scoring and Interpreting Sleep-Wake Activity. *Perceptual and Motor Skills*, *85*(1), 207–216. https://doi.org/10.2466/pms.1997.85.1.207

Jex, H. R., McDonnel, J. D., & Phatak, A. V. (1966). A “critical” tracking task for man-machine research related to the operator’s effective delay time. *NASA Contractor Report NASA CR*, *1*(ePub), 1689–1699.

Jex, H. R., McDonnell, J. D., & Phatak, A. V. (1966). A “Critical” Tracking Task for Manual Control Research. *IEEE Transactions on Human Factors in Electronics*, *HFE-7*(4), 138–145. https://doi.org/10.1109/THFE.1966.232660

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Epworth sleepiness scale. *Sleep*. https://doi.org/10.1093/sleep/14.6.540

Johns, M. W. (1992). Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep*, *15*(4), 376–381. https://doi.org/10.1093/sleep/15.4.376

Koh, A., Jones, G. R., Spencer, J. W., & Thomas, I. (2007). Chromatic analysis of signals from a driver fatigue monitoring unit. *Measurement Science and Technology*, *18*(3), 747–754. https://doi.org/10.1088/0957-0233/18/3/026

Kohler, K. J. (2006). Paradigms in experimental prosodic analysis: from measurement to function. *Methods in Empirical Prosody Research*, 123–152. Retrieved from http://books.google.com/books?hl=pt-

PT&lr=&id=1b5cebiOykAC&oi=fnd&pg=PA123&dq=prosodic+analysis&ots=HU-ncNaMOe&sig=iz8cm9wJscVN6ki8IeCwOJ37BfE

Kronauer, R. E., Czeisler, C. A., Pilato, S. F., Moore-Ede, M. C., & Weitzman, E. D. (1982). Mathematical model of the human circadian system with two interacting oscillators. *The American Journal of Physiology*, *242*(1), R3-17. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7058927

Kuo-An Hwang, & Chia-Hao Yang. (2009). Automated Inattention and Fatigue Detection System in Distance Education for Elementary School Students. *Educational Technology & Society*, *12*(2), 22–35. Retrieved from https://www.engineeringvillage.com/share/document.url?mid=inspec_10655dd128cbb50f29M49a22061377553&database=ins

Li, Z., Li, W., Yao, Z., & Li, Y. (2011). Speech-based five features extraction and hardware trade-off design for mental fatigue monitoring. *Proceedings - 2011 International Conference of Information Technology, Computer Engineering and Management Sciences, ICM 2011*, *1*, 68–71. https://doi.org/10.1109/ICM.2011.163

Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, *1129*, 305–322. https://doi.org/10.1196/annals.1417.002

Lorist, M. M., Klein, M., Nieuwenhuis, S., De Jong, R., Mulder, G., & Meijman, T. F. (2000). Mental fatigue and task control: Planning and preparation. *Psychophysiology*, *37*(5), 614–625. https://doi.org/10.1017/S004857720099005X

Mabbott, N. (2003). ARRB pro-active fatigue management system. *Road and Transport Research*, *12*(1), 56–62.

Mackie, R. R. (1977). *Vigilance:Theory, Operational Performance, and Physiological Correlates* (1st ed.). Springer.

May, J. F. (2011). Driver fatigue. *Handbook of Traffic Psychology*, *36*(2), 287–297. https://doi.org/10.1016/B978-0-12-381984-0.10021-9

McClernon, C. K., & Miller, J. C. (2011). Variance as a measure of performance in an aviation context. *International Journal of Aviation Psychology*, *21*(4), 397–412. https://doi.org/10.1080/10508414.2011.606765

McNair, D M, Lorr, M, Droppelman, L. F. (1971). Manual for the profile of mood states. *Educational and Industrial Testing Service, San Diego*, 27. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Manual+for+the+Profile+of+Mood+States#4

McRuer, D. T., & Hofmann, L. G. (1968). *New Approaches to Human-Pilot/Vehicle Dynamic Analysis*. *Government PROcurement*.

McRuer, D. T., & Jex, H. R. (1967). A Review of Quasi-Linear Pilot Models. *IEEE Transactions on Human Factors in Electronics*, *HFE-8*(3), 231–249. https://doi.org/10.1109/THFE.1967.234304

McRuer, D. T., Weir, D. H., Jex, H. R., Magdaleno, R. E., & Wade Allen, R. (1975). Measurement Of Driver–Vehicle Multiloop Response Properties With A Single Disturbance Input. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-5*(5), 490–497. https://doi.org/10.1109/TSMC.1975.5408371

Millar, K., Jelicic, M., Bonke, B., & Asbury, A. J. (1995). Assessment of preoperative anxiety: Comparison of measures in patients awaiting surgery for breast cancer. *British Journal of Anaesthesia*, *74*(2), 180–183. https://doi.org/10.1093/bja/74.2.180

Miller, J. C. (1996). Fit for Duty? *Ergonomics in Design*, *4*(2), 11–17.

Miller, J. C. (2005). Real-time Bio-Sensors for Enhanced C2ISR Operator Performance. In *Proceedings of SPIE -Biomonitoring for Physiological and Cognitive Performance during Military Operations* (Vol. 5797, pp. 8–13). https://doi.org/10.1117/12.602180

Miller, J. C. (2006). *In Search of Circasemidian Rhythms* (No. 2006–0074, ADA458153). Brooks City-Base TX: Air Force Research Laboratory.

Miller, J. C. (2008). Fundamentals of Shiftwork Scheduling. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *16*(3), 13–17. https://doi.org/10.1518/106480408X333109

Miller, J. C. (2012). An Historical View of Operator Fatigue. In G. Matthews, P. A. Desmond, C. Neubauer, & P. A. Hancock (Eds.), *The handbook of operator fatigue* (pp. 25–45). Surrey, England: Ashgate.

Miller, J. C. (2013). *Anatomy of a Fatigue-Related Accident*. *Ashgate*. Retrieved from https://www.smashwords.com/books/view/379949

Miller, J. C. (2013). *Fundamentals of Shiftwork Scheduling, 3rd Edition: Fixing Stupid*. Smashwords.

Miller, J. C., & Mackie, R. R. (1978). *Effects of Hours of Service Regularity of Schedules and Cargo Loading on Truck and Bus Driver Fatigue* (No. HFR-TR-1765-F, NTIS PB-290-957). *National Technical Information Service, Us Department of Commence*. Washington DC: National Highway Traffic Safety Administration.

Miller, J. C., Mackie, R. R., & Standards, C. for C. P. T. N. B. of. (1980). *Vigilance Research and Nuclear Security: Critical Review and Potential Applications to Security Guard Performance* (No. 2722; National Bureau of Standards contract NBS-GCR-80-201 for the Defense Nuclear Agency; available from the Electronics and Electrical Engineering Laboratory, Office of Law Enforcement Standards, National Institute of Standards and Technology, nist.) (Vol. NBS-GCR-80). Goleta CA: Human Factors Research Inc.

Miller, J. C., Wagner, R., & Bezos, A. P. (1993). *Advanced Locomotive Cab Alerter* (Federal Railroad Administration Small Business Innovation Research No. SBIR 92-FR1, Contract DTRS-57-92-C-00100). Rockville MD: Pulse Electronics, Inc.

Mitler, M. M., Carskadon, M. A., Czeisler, C. A., Dement, W. C., Dinges, D. F., & Graeber, R. C. (1988). Catastrophes, sleep, and public policy: Consensus report. *Sleep*, *11*(1), 100–109. https://doi.org/10.1093/sleep/11.1.100

Montgomery, L. D., Montgomery, R. W., & Guisado, R. (1995). Rheoencephalographic and electroencephalographic measures of cognitive workload: analytical procedures. *Biological Psychology*, *40*(1–2), 143–159. https://doi.org/10.1016/0301-0511(95)05117-1

Morris, A., & Ulieru, M. (2012). FRIEND: A brain-monitoring agent for adaptive and assistive systems. *Annual Review of CyberTherapy and Telemedicine*, *10*, 314–318. https://doi.org/10.3233/978-1-61499-121-2-314

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*, *42*(3), 343–360.

https://doi.org/10.1016/0301-0511(95)05166-X

Mosso, A. (1915). *Fatigue*. London: Allen & Unwin. Retrieved from https://www.amazon.com/Fatigue-Mosso/dp/B005HIAVXQ

Muscio, B. (1921). Is a Fatigue Test Possible? *British Journal of Psychology. General Section*, *12*(1), 31–46. https://doi.org/10.1111/j.2044-8295.1921.tb00036.x

Newcombe, P., & Boyle, G. (1995). High school students' sports personalities: Variations across participation level, gender, type of sport, and success. *International Journal of Sport …*, 277–294. Retrieved from http://epublications.bond.edu.au/hss_pubs/802/

O'Brien, R. (1998). An overview of the methodological approach of action Research. *University of Toronto*, *17*, 1–15.

Orchard, L. N., & Stern, J. A. (1991). Blinks as an index of cognitive activity during reading. *Integrative Physiological and Behavioral Science*, *26*(2), 108–116. https://doi.org/10.1007/BF02691032

Palmer, M., Linde, M., & Pons, G. X. (2004). Correlational patterns between invertebrate species composition and the presence of an invasive plant. *Acta Oecologica*, *26*(3), 219–226. https://doi.org/10.1016/j.actao.2004.05.005

Parasuraman, R., & Wilson, G. F. (2008). Putting the Brain to Work: Neuroergonomics Past, Present, and Future. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 468–474. https://doi.org/10.1518/001872008X288349

Perelli, L. P. (1980). *Fatigue stressors in simulated long-duration flight. Effects on performance, information processing, subjective fatigue, and physiological cost*. *Sam-Tr* (Vol. (186). Retrieved from http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/734/CN-00179734/frame.html

Perez, W. A., Masline, P. J., Ramsey, E. G., & Urban, K. E. (1987). *Unified Tri-services cognitive performance ASSESSMENT BATTERY : REVIEW AND METHODOLOGY*.

Picard, R., & Klein, J. (2002). Computers that recognize and respond to user emotion: theoretical and practical implications. *Interacting with Computers*, *14*(Spitz 1945), 141–

169. Retrieved from http://iwc.oxfordjournals.org/content/14/2/141.short

Picard, R. W. (2003). Affective computing: Challenges. *International Journal of Human Computer Studies*, *59*(1–2), 55–64. https://doi.org/10.1016/S1071-5819(03)00052-1

Picard, R., & Wexelblat, A. (2002). Future Interfaces: Social and Emotional. *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, 698. https://doi.org/10.1145/506443.506552

Pimenta, A., Carneiro, D., Novais, P., & Neves, J. (2013). Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8073 LNAI, pp. 222–231). Springer. https://doi.org/10.1007/978-3-642-40846-5_23

Reeves, D. L., Winter, K. P., Bleiberg, J., & Kane, R. L. (2007). ANAM genogram: historical perspectives, description, and current endeavors. *Arch Clin Neuropsychol*, *22 Suppl 1*, S15-37. https://doi.org/10.1016/j.acn.2006.10.013

Reid, G. B., Eggemeier, F. T., & Shingledecker, C. A. (1982). Subjective workload assessment technique. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADP001142

Reilly, T., & Edwards, B. (2007). Altered sleep-wake cycles and physical performance in athletes. *Physiology and Behavior*, *90*(2–3), 274–284. https://doi.org/10.1016/j.physbeh.2006.09.017

Roach, G. D., Fletcher, A., & Dawson, D. (2004). A Model to Predict Work-Related Fatigue Based on Hours of Work. *Aviation Space and Environmental Medicine*, *75*(3), A61-69-74. https://doi.org/10.1016/S1369-8478(01)00015-8

Robert, G., & Hockey, J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1–3), 73–93. https://doi.org/10.1016/S0301-0511(96)05223-4

Roscoe, S. N. (1974). Assessment of Pilotage Error in Airborne Area Navigation Procedures. *Human Factors: The Journal of Human Factors and Ergonomics Society*, *16*(3), 223–228.

https://doi.org/10.1177/001872087401600304

Rosekind, M. R. (2005). Underestimating the societal costs of impaired alertness: Safety, health and productivity risks. *Sleep Medicine*, *6*(SUPPL. 1), S21–S25. https://doi.org/10.1016/S1389-9457(05)80005-X

Safety Net. (2009). Fatigue. *Transport*.

Samn, S., & Perelli, L. (1982). *Estimating Aircrew Fatigue: A Technique with Application to Airlift Operations* (No. 82–21, ADA125319). Brooks AFB, TX: USAF School of Aerospace Medicine. https://doi.org/SAM-TR-82-21

Segerstrom, S. C., & Nes, L. S. (2007). Heart rate variability reflects self-regulatory strength, effort, and fatigue. *Psychological Science*, *18*(3), 275–281. https://doi.org/10.1111/j.1467-9280.2007.01888.x

Serber, M. L., Kaminski, M. a, Payton, G. M., Moreland, K. L., Ph, D., Estes, S. L., & Ii, T. a N. (2010). *Human Performance and Fatigue Research for Controllers* (No. MTR100316). *Distribution*. Mitre Corp. https://doi.org/MTR100316

Seret, A., Maldonado, S., Weber, R., & Baesens, B. (2013). Knowledge based feature selection for unsupervised learning: theory and application. In *Status: Published*. Retrieved from https://lirias.kuleuven.be/handle/123456789/438975

Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2011). Profile of Mood States (POMS). *STOP, THAT and One Hundred Other Sleep Scales*, 285–286. https://doi.org/10.1007/978-1-4419-9893-4_68

Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2012). *STOP, THAT and one hundred other sleep scales*. *STOP, THAT and One Hundred Other Sleep Scales* (2012 editi). New York: Springer. https://doi.org/10.1007/978-1-4419-9893-4

Sharit, J. (2003). Perspectives on computer aiding in cognitive work domains: Toward predictions of effectiveness and use. *Ergonomics*, *46*(1–3), 126–140. https://doi.org/10.1080/00140130303533

Shen, K. Q., Ong, C. J., Li, X. P., Hui, Z., & Wilder-Smith, E. P. V. (2007). A feature selection method for multilevel mental fatigue EEG classification. *IEEE Transactions on Biomedical*

Engineering, *54*(7), 1231–1237. https://doi.org/10.1109/TBME.2007.890733

Shiomi, K., Itano, K., & Suzuki, A. (2010). Development and Evaluation of the Fatigue and Drowsiness Predictor. *Icas-Proceedings.Net*. Retrieved from http://www.icas-proceedings.net/ICAS2012/PAPERS/920.PDF

Smith, A. P., & Miles, C. (1986). The effects of lunch on cognitive vigilance tasks. *Ergonomics*, *29*(10), 1251–1261. https://doi.org/10.1080/00140138608967238

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacob, G. A. (1983). Manual for the State-Trait Anxiety Inventory. https://doi.org/10.1007/978-1-4419-9893-4

Stein, A. C., Parseghian, Z., & Allen, R. W. (1990). the Development of a Low-Cost Portable System for the Detection of Truck Driver Fatigue (pp. 523–536).

Stern, J. A., Walrath, L. C., & Goldstein, R. (1984). The Endogenous Eyeblink. *Psychophysiology*, *21*(1), 22–33. https://doi.org/10.1111/j.1469-8986.1984.tb02312.x

Stork, M. (2012). Some Methods Systems and Sensors which are Possible for Driver's Drowsiness Estimation. *Proceedings of the 3rd International Conference on Circuits, Systems, Control, Signals*, 87–94. Retrieved from http://www.wseas.us/e-library/conferences/2012/Barcelona/CSCS/CSCS-14.pdf

Sumner, W. 2nd, & Xu, J. Z. (2002). Modeling fatigue. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, 747–751. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2244328/

Tang, N. K. Y., & Harvey, A. G. (2004). Correcting distorted perception of sleep in insomnia: A novel behavioural experiment? *Behaviour Research and Therapy*, *42*(1), 27–39. https://doi.org/10.1016/S0005-7967(03)00068-8

Terman, M., Rifkin, J., & White, T. (2001). Morningness-eveningness questionnaire. https://doi.org/https://doi.org/10.1007/978-1-4419-9893-4_54

Tin Kam Ho. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* (Vol. 1, pp. 278–282). IEEE Comput. Soc. Press. https://doi.org/10.1109/ICDAR.1995.598994

Tucker, P. (2003). The impact of rest breaks upon accident risk, fatigue and performance: A

review. *Work and Stress*, *17*(2), 123–137. https://doi.org/10.1080/0267837031000155949

Turnage, J. J., Kennedy, R. S., Smith, M. G., Baltzley, D. R., & Lane, N. E. (1992a). *Development of microcomputer-based mental acuity tests* (No. NASA-CR-185607). *Ergonomics* (Vol. 35). National Aeronautics and Space Administration. https://doi.org/10.1080/00140139208967393

Turnage, J. J., Kennedy, R. S., Smith, M. G., Baltzley, D. R., & Lane, N. E. (1992b). Development of microcomputer-based mental acuity tests. *Ergonomics*, *35*(10), 1271–1295. https://doi.org/10.1080/00140139208967393

Turner, R. (2013). *The handbook of operator fatigue*. *Ergonomics* (Vol. 56). Ashgate Publishing, Ltd. https://doi.org/10.1080/00140139.2013.819674

Valdez, P., García, A., Ramírez, C., & García, A. (2012). Circadian rhythms in cognitive performance: implications for neuropsychological assessment. *ChronoPhysiology and Therapy*, *Volume 2*, 81. https://doi.org/10.2147/CPT.S32586

Van Der Linden, D., & Eling, P. (2006). Mental fatigue disturbs local processing more than global processing. *Psychological Research*, *70*(5), 395–402. https://doi.org/10.1007/s00426-005-0228-7

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. https://doi.org/10.1016/S0001-6918(02)00150-6

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance Requires Hard Mental Work and Is Stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 433–441. https://doi.org/10.1518/001872008X312152

Welford, A. T. (1968). Fundaments of skill. Retrieved from http://psycnet.apa.org/record/1968-35018-000

Wickens C. D., M. J. S. (2008). *Applied attention theory. Ergonomics* (Vol. 1). Boca Raton FL: CRC Press. https://doi.org/10.1080/00140130802295564

Wierwille, W. (1999). Historical perspective on slow eyelid closure: Whence PERCLOS?, Ocular

Measures of Driver Alertness. *Technical Conference Proceedings, Herndon, Va*, 31–53. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Historical+perspective+on+slow+eyelid+closure:+Whence+PERCLOS#0

Wierwille, W. W., & Gutmann, J. C. (1978). Comparison of Primary and Secondary Task Measures as a Function of Simulated Vehicle Dynamics and Driving Conditions. *Human Factors: The Journal of Human Factors and Ergonomics Society*, *20*(2), 233–244. https://doi.org/10.1177/001872087802000211

Wilkinson, R. T., & Houghton, D. (1982). Field test of arousal: a portable reaction timer with data storage. *Human Factors*, *24*(4), 487–493. https://doi.org/10.1177/001872088202400409

Williamson, R. J., Purcell, S., Sterne, A., Wessely, S., Hotopf, M., Farmer, A., & Sham, P. C. (2005). The relationship of fatigue to mental and physical health in a community sample. *Social Psychiatry and Psychiatric Epidemiology*, *40*(2), 126–132. https://doi.org/10.1007/s00127-005-0858-5

Woodrow, A. (2014). Cognitive Performance Research at Brooks Air Force Base, Texas, 1960-2009. *Aviation, Space, and Environmental Medicine*, *85*(4), 482. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Cognitive+Performance+Research+at+Brooks+Air+Force+Base,+Texas,+1960-2009#0

Wooldridge, M. (2009). *An Introduction to MultiAgent Systems [Paperback]*. *Chichester, England*. https://doi.org/10.1108/eb050656

X-, S. F. (1968). Self-evaluation questionnaire stai form x-1, 1968. Retrieved from http://ubir.buffalo.edu/xmlui/handle/10477/1878

Yang, J., Frangi, A. F., Yang, J. Y., Zhang, D., & Jin, Z. (2005). KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(2), 230–244. https://doi.org/10.1109/TPAMI.2005.33

Yannis, G., Richter, T., Ruhl, S., Dragomanovits, A., Graham, D., Laiou, A., … Domenichini, L. (2017). Road traffic accident prediction modelling : a literature review. *Proceedings of*

the Institution of Civil Engineers, *170*, 245–254. https://doi.org/http://dx.doi.org/10.1680/jtran.16.00067

Zhang, C., Zheng, C. X., & Yu, X. L. (2008). Evaluation of mental fatigue based on multipsychophysiological parameters and kernel learning algorithms. *Chinese Science Bulletin*, *53*(12), 1835–1847. https://doi.org/10.1007/s11434-008-0245-1