POLITECNICO DI TORINO

MASTER's Degree in Biomedical Engineering

MASTER's Degree Thesis

Exploring the impact of Stress and Cognitive Workload on Eye Movements: A Preliminary Study

Supervisors

Prof. Danilo DEMARCHI

Prof. Irene BURAIOLI

Candidate

Eva DEL CARRETTO

DI PONTI E SESSAME

December 2023

Summary

The evolution of technology used in the service of human work has led to the development of human-machine interaction (HMI) systems. They enable the performance of certain functions and the achievement of specific goals through collaboration between humans and machines. These systems are incredibly innovative and can be extremely useful for humans; moreover, the HMI systems can be used mainly in two ways. In some cases, they are specifically designed to relieve the subject of several tasks carried out directly by the machine. In other cases, the operator must fulfill several duties simultaneously because of the interaction with the HMI device. These situations generate high stress and mental workload (MWL) levels in the subject, leading him or her to perform incorrect actions and risky situations in which accidents are more likely to occur. Hence, HMI systems must be created to avoid these dangerous situations for the person interacting with the machine and those around him. These safety mechanisms are also based on measuring and controlling the subjects' stress and cognitive load, preventing them from performing tasks or functions if these two altered states are too pronounced. In the context of this study, reference is made to aviation pilots who increasingly interact with machines during their work. There is also a tendency to move from having two pilots per aircraft to only one by replacing the copilot with an HMI system; the so-called Single Pilot Operation SPO. HMI technologies have the task of supervising the pilots and, in case these systems notice that they are overloaded, giving support by simplifying the charges they must fulfill or even replacing the pilots themself. Various biosignals have been seen to visualize changes in these two altered states; the most studied are the f-NIRS, the EDA, the ECG, the respiration, the body temperature, and the ocular signal. The last one is the signal analyzed in this thesis work, which gets more attention in the field of research, especially in the aeronautical one. Therefore, this study aims to relate stress and cognitive load variations to the eye signal. The goal was achieved through initial research of the state of the art regarding the features and their trends that can be extracted from the ocular signal and related to variations in stress and MWL. Subsequently, a test session was carried out with 64 volunteer participants who were asked to perform two kinds of tests, called Stroop and N-Back tests, explicitly designed

to generate feelings of stress and MWL in the subjects. The signal was acquired through the gold standard in eye-tracking Tobii Glasses 3. Finally, from the ocular signal, different types of eye movements were identified, and from them, various features were extracted. Some metrics were already analyzed in the literature, but others were never investigated; among them, a few are coming from a frequency analysis. The results showed that 83In conclusion, promising results have been found in this thesis study that push us to continue research in this field to achieve the development of increasingly reliable and safe HMI systems.

Table of Contents

List of Tables VII

List of Figures VIII

Acronyms XIII

1 Introduction 1
1.1 Objectives of the thesis project 3

2 Background 4
2.1 Mental Workload 4
2.2 Stress 5
2.3 Stress and MWL relationship 5
2.4 Methods for stress and mental workload assessment 6
2.4.1 Stroop Test 6
2.4.2 N-Back Test 7
2.5 Eyes 8
2.5.1 Eye Movements 8
2.6 State-of-the-art of eye movements assessment 10
2.7 Eye-Tracking: history, devices and applications 13
2.8 Data set acquisition 15
2.8.1 Biosignals acquired 17
2.8.2 Tobii Glasses 3 17
2.8.3 Tobii Pro Lab 20
2.8.4 Structure of the trial 24
2.9 Different kinds of normalization 28
2.10 Comparison Tobii Pro Lab - Algorithm 29
2.11 Statistical Analysis 31
2.11.1 Kruskal-Wallis H Test and Mann-Whitney U Test 31

3 Materials and Methods 32
3.1 Validation of Pixels - Degrees Conversion . . . 32
3.1.1 Dataset description for validation . . . 37
3.2 Algorithm for saccades identification . . . 37
3.2.1 Implemented algorithm description . . . 38
3.3 Tobii Pro Lab Use . . . 39
3.4 Eye movements and gaze points . . . 40
3.5 Preprocessing . . . 40
3.6 Processing . . . 47
3.6.1 Frequency Analysis . . . 49
3.7 Dataset description . . . 55

4 Results 56
4.1 Validation of Pixels-Degrees Conversion . . . 56
4.1.1 Discussion of validation of pixels-degrees conversion . . . 58
4.2 Stress and MWL assessment in results . . . 58
4.3 Steps for choosing the best normalization method . . . 58
4.3.1 Evaluation of different kinds of normalization . . . 59
4.4 Dataset Investigation . . . 61
4.4.1 Dataset distribution in classes . . . 62
4.4.2 Dataset investigation results . . . 63
4.5 Comparison with Tobii Pro Lab . . . 65
4.6 Statistical Analysis results . . . 74
4.6.1 Binary Classification Results . . . 75
4.6.2 Overall Statistical Results . . . 79
4.6.3 Resume of statistical analysis results . . . 82

5 Conclusions 83

A Pixels-degrees conversion validation 86

B Features boxplots 94

Bibliography 104

List of Tables

2.1 Summary table of characteristics of eye movements 10
2.2 Features trend showed by the literature 13
2.3 Tobii Glasses 3 characteristics of resolution and Field of View . . . 18

3.1 Summary of methods for saccades and fixations identification 38
3.2 Eye movement features and trends, if known in literature, are shown. 54

4.1 Resume of binary classification results 75
4.2 Resume of binary classification results with H statistic threshold . . 79
4.3 Resume of overall statistical results 80
4.4 Resume of overall statistical results with H statistic threshold . . . 82

List of Figures

1.1 Tobii Glasses 3 2

2.1 Stress and MWL relationship 6
2.2 Example of Stroop Test Interface 7
2.3 Example of N-Back Test Interface 7
2.4 Graphical description of fixations, saccades and SIs 9
2.5 EOG principle of operation 14
2.6 Trial setup: side shot 15
2.7 Trial setup: frontal shot 16
2.8 Tobii Glasses 3 setup and his additional devices 18
2.9 Tobii Glasses 3 coordinate system MCS 19
2.10 Tobii Glasses 3 coordinate system HUCS 20
2.11 Extract of raw data 20
2.12 Interface Tobii Pro Lab: 1) Data Export; 2) Metrics; 3) Recording information; 4) Snapshot; 5) Events list; 6) TOIs; 7) TOIs shown graphically; 8) Events shown graphically 21
2.13 TOIs and events shown graphically 21
2.14 Metrics file creation 22
2.15 Data Export file creation 22
2.16 Data Export from Tobii Pro Lab 23
2.17 TOIs creation 23
2.18 Events creation 23
2.19 Stroop Test interface 25
2.20 Dual N-Back Test interface 26
2.21 Questionnaire: personal data 27
2.22 Questionnaire: level of "difficulty" 27
2.23 Flowchart acquisition protocol 28
2.24 Confusion matrix 30

3.1 Reference systems of the Field of View of Tobii Glasses 3 and of the snapshot representing the screen of the PC 33

3.2 Reference systems explained of Diaz et Al. paper 33
3.3 Interface n.1 with all marker positions 34
3.4 Interface n.2 with all marker positions 34
3.5 Real setup of the experiment 35
3.6 Flowchart pixels-degrees conversion validation 36
3.7 Tokuda algorithm 39
3.8 Eye movements and gaze points 40
3.9 Raw data imported from Tobii Glasses 3 41
3.10 Eyes Not Found 43
3.11 Horizontal and vertical signals after "time axis filling" and pixels-degrees conversion 43
3.12 Tokuda algorithm applied to the horizontal signal 44
3.13 Horizontal and vertical signals with division in tests phases 45
3.14 Preprocessing flowchart 46
3.15 Processing flowchart 47
3.16 Fixation and saccade lasting 48
3.17 PSD of interpolated raw signal 49
3.18 Mask of Butterworth filter for tremor removal 50
3.19 Raw signal, filtered signal with moving median and filtered signal with Butterworth filter: PSD and signal 51
3.20 Mask of Butterworth filter for drift removal 51
3.21 Raw signal, filtered signal with moving median and filtered signal with Butterworth filter: PSD and signal 52
3.22 PSD of filtered signal in each macro-phase 53

4.1 Comparison of ideal and calculated angles of one recording of example 57
4.2 Mean error and standard deviation of the gaze location corresponding to each position of the marker with Assisted and Manual Mapping . 57
4.3 Stroop Test: histogram of standard deviations 59
4.4 Visual N-Back Test: histogram of standard deviations 60
4.5 Auditory N-Back Test: histogram of standard deviations 60
4.6 Dual N-Back Test: histogram of standard deviations 61
4.7 Distribution of classes 62
4.8 Data set investigation: Ratio of powers in horizontal signal 63
4.9 Data set investigation: single SI velocity over phases duration . . . 64
4.10 Data set investigation: Blinking duration 64
4.11 Data set investigation: SI duration 65
4.12 Confusion matrix for this use case 66
4.13 Accuracy, Specificity and Sensitivity of saccades detection 67
4.14 PPV and NPV of saccades detection 67
4.15 F1 score for saccade detection 68

4.16 Accuracy, Specificity and Sensitivity of fixations detection 68
4.17 PPV of fixations detection; TN is equal to 0 so NPV can't be shown 69
4.18 F1 score for fixation detection 69
4.19 Comparison: Duration fixations 71
4.20 Comparison: Duration fixations with SI and saccades considered as saccades 72
4.21 Comparison: Frequency fixations 73
4.22 Comparison: Frequency fixations with SI and saccades considered as saccades 73
4.23 Comparison: Frequency saccades 74
4.24 SI value in Stroop Test dataset, low H statistic 77
4.25 Fixation lasting in Stroop Test dataset, medium H statistic 77
4.26 Single SI velocity in Stroop Test dataset, high H statistic 78
4.27 Overall results: features resulting to have a feature significance >3 . 81

A.1 Comparison of ideal and calculated angles: Recording 2 86
A.2 Comparison of ideal and calculated angles: Recording 3 87
A.3 Comparison of ideal and calculated angles: Recording 4 87
A.4 Comparison of ideal and calculated angles: Recording 5 88
A.5 Comparison of ideal and calculated angles: Recording 6 88
A.6 Comparison of ideal and calculated angles: Recording 7 89
A.7 Comparison of ideal and calculated angles: Recording 8 89
A.8 Comparison of ideal and calculated angles: Recording 9 90
A.9 Comparison of ideal and calculated angles: Recording 10 90
A.10 Comparison of ideal and calculated angles: Recording 11 91
A.11 Comparison of ideal and calculated angles: Recording 12 91
A.12 Comparison of ideal and calculated angles: Recording 13 92
A.13 Comparison of ideal and calculated angles: Recording 14 92
A.14 Comparison of ideal and calculated angles: Recording 15 93

B.1 Duration fixations 94
B.2 Duration saccades 95
B.3 Frequency blinking 95
B.4 Frequency fixation 96
B.5 Frequency saccades 96
B.6 Frequency SI 97
B.7 X velocity 97
B.8 Y velocity 98
B.9 Single SI velocity 98
B.10 SI value 99
B.11 Left eye relative diameter 99

B.12 Right eye relative diameter 100
B.13 X HF bandwidth 100
B.14 Y HF bandwidth 101
B.15 X LF bandwidth 101
B.16 Y LF bandwidth 102
B.17 Interval blinking 102
B.18 SI velocity 103
B.19 Y ratio 103

Acronyms

HMI
Human-Machine Interaction

MWL
Mental Workload

f-NIRS
Near-Infrared Spectroscopy

EDA
Electrodermal Activity

ECG
Electrocardiogram

EOG
Electroculogram

FOV
Field of View

TB
Time-Based

FB
Frequency-Based

LF
Low Frequency

HF

High Frequency

ENF

Eyes Not Found

MCS

Media Coordinate System

HUCS

Head Unit Coordinate System

TOI

Time of Interest

AOI

Area of Interest

PSD

Power Spectral Density

UM

Unity of Measurement

Chapter 1

Introduction

Increasingly used in a variety of technological applications, Human-Machine Interaction (HMI) systems are evolving rapidly. HMI technologies are systems where machines and humans collaborate for satisfying functions through interaction. They are gaining more and more popularity thanks to the rapid advancement of technology that can be used to support human work [1]. HMI systems are widely exploited in healthcare applications, in particular, in the fields of medical diagnostics and monitoring of patients to assess changes in vital signs [2]. Moreover, they are very widespread in aviation and automotive fields because they form specific information at the right time to ensure a proper reaction by the pilot in case of need [2][3]. In some cases, they are specifically designed to relieve the subject of several tasks carried out directly by the machine; in other cases, however, the subject finds himself having to fulfill several tasks simultaneously. If these situations take place, the mental workload (MWL) and stress levels of the subject increase [4][5]. In particular, the rising of a high level of these two cognitive conditions could lead to accidents and fatal errors. As key concepts in this context, stress and MWL must be studied in order to apply their measurement knowledge to the safety systems of HMI technologies. In general, they affect mental and emotional processes, thus hindering decision-making in many situations, even in people's daily lives. They are concepts that are very difficult to fix but must be studied and defined to better understand how to recognize and measure them [4].

MWL is a mental fatigue feeling; a person perceives a high level of MWL if she or he is performing different tasks at the same time [6]. Instead, stress is a feeling of discomfort; moreover, it is present in "fight or flight" contests [7]. It is possible to ensure greater reliability of human performance and, therefore, a higher level of safety, thanks to the knowledge of stress and MWL measurement and the use of different real-time control systems and synchronous human-machine communication [2]. This helps develop the most proper HMI systems.

There are several systems for measuring stress and MWL levels. Historically,

the most used have been subjective questionnaires and behavioral analysis, but since the former are not applicable in industrial environments, and the latter do not allow scalability of the approach, methods based on the analysis of physiological parameters are increasingly being developed in recent years. This diffusion is also due to the development of new technologies for sensors that made them wearable, more reliable, and less invasive [8].

Figure 1.1: Tobii Glasses 3

Speaking about the physiological approach, according to the literature, many biosignals can be related to the variation of stress and MWL; the most studied are the functional near infrared spectroscopy (f-NIRS), the electrodermal activity (EDA), the electrocardiogram (ECG), the respiration, the body temperature, and the ocular signal. The last one is the signal analyzed in this thesis study, which gets more attention in the field of research [9], especially in the aeronautical one, for the strong relationship with stress and MWL states shown. Despite this, the link between the two altered states and the ocular signal has not yet been well understood; that's the reason why this paper aims to investigate this relationship and understand how to extract information about stress and MWL levels from metrics derived from the signal itself.

To achieve the goal, a test session with 64 participants was carried out; they were asked to do two tests, the Stroop Test and the N-Back Test. They are two tests developed by researchers and widely used in literature, designed to generate different levels of these altered cognitive conditions in the participants. The ocular signals are acquired through eye-tracker Tobii Glasses 3, shown in Fig. 1.1, and

later processed for feature extraction. Some of these features were already analyzed in literature while others were never investigated before. In order to understand the significance of the features, an analysis of results, followed by a statistical analysis, has been performed.

The thesis study is structured in four chapters; the first provides the background information needed to understand the results. All the information already present in the literature are shown in this chapter: they are about the definition of stress and MWL, the two tests performed by the participants, and the state-of-the-art of eye movement processing for stress and MWL detection. Moreover, they are described the devices used in the past and the present for ocular signal acquisition, and the device used in this study, and the related software. Finally, there is the description of the acquisition protocol. Chapter 3 describes the materials and methods used in this study, in particular, the preprocessing and the processing of the signal for feature extraction. In Chapter 4 results are described and discussed, while in Chapter 5 the conclusive observations are listed.

1.1 Objectives of the thesis project

The objectives of the thesis are the following ones:

- Definition of the best significant eye-tracking features to evaluate stress and mental workload variations.
- Definition and implementation of tests to detect mental workload variations to gather a significant dataset.
- Extraction of the selected features from the data gathered from the tests.
- Evaluation of the correlation between stress, mental workload variations, and eye tracking parameters variations.

Chapter 2

Background

In this section, background information are provided: the difference between MWL and stress is highlighted, the Stroop and N-Back Tests are described, and an overview of eye physiology and eye movements is carried out. Then, research about the state-of-the-art of eye movement processing for stress and MWL detection, the history of eye-tracking, and related devices and applications is performed. Then, the dataset acquisition, with the device and the software used, is explained. Finally, the theory part regarding the normalization, the comparison between Tobii Pro Lab and the algorithm, and the statistical analysis.

2.1 Mental Workload

To date, there is no single definition of 'Mental Workload' but different interpretations depending on the field of research. For example, Alsuraykh et al. [7] define Mental Workload (MWL) as the relationship between the demands on the subject and the resources he employs to cope with them. It is closely related to the processes of attention and the concept of effort; moreover, it is a person-specific and complex construct. It is the cognitive effort required to complete a task in a limited time [6], and it reflects not only task specificities but also the operator's abilities and performance. In fact, excessive cognitive load decreases both the subject's performance and his motivation by increasing the number of errors made in performing the assigned task [7]. Instead, too low cognitive workload levels may result in errors or accidents due to boredom. MWL is strongly related to the sympathetic nervous system, so that it can be recognized in many physiological signals. In this thesis study, the terms "mental workload" and "cognitive load" are used as synonymous, as explained in the review of Luzzani et al. [8].

2.2 Stress

Hans Selye [10], also known as the "father of stress", was the first researcher to define stress; according to him, it is "the nonspecific response of the body to any demand". For others, Selye's definition is too general and, it has to be taken into account that, according to the various disciplines, the term "stress" has different meanings. Richard Lazarus [11] has resumed them in these words: "For sociologists, it is a disequilibrium, that is disturbances in the social structure within which people live. Engineers conceive of stress as some external forces which produce strain in the material exposed to it. Physiologists deal with the physical stressors that include a wide range of stimulus conditions that are noxious to the body." According to the different meanings of stress just named, the definition that best fits, in this study, is the lack of balance between the ability of the subject and the environmental context in which it is located [7].

From the physiological point of view the concept of stress is closely related to the "fight or flight" mechanism. It is the body's reaction to a threat, allowing humans to survive situations of great stress and danger. It is regulated by the sympathetic nervous system, which leads to the secretion of adrenaline and stress hormones by the adrenal gland. The release of these hormones triggers a series of phenomena: the heartbeat accelerates and, with it, the blood pressure increases, the respiratory rate speeds up, digestion slows down and there is a diversion of blood flow to the main muscle groups to give the body a burst of energy and strength. The body becomes tense, ready to react, exactly to fight or fly. This stress response occurs automatically and involuntarily and often helps to perform better when people are under pressure and to increase their performance [12].

In recent years, the scientific contribution in the search for indicators that can identify variations in stress and cognitive load has greatly increased. Several biological signals have been detected, including the signal from eye movements.

2.3 Stress and MWL relationship

Stress and MWL are not two separate concepts with no relationship. In fact, as explained in Luzzani et al. [8], stress is a factor that, with the MWL, could influence activity performance. Analysing Fig. 2.1, the first block *Task Load* represents the tasks and the duties that a subject must perform; the second block *Mental Load* represents the amount of cognitive load necessary to carry out all the tasks with and high-performance level. If the third block, called *Deplation Factors* is added, then stress, fatigue and motivation are additional perceptions that the subject must manage that lead to the fourth block *Performance* that represents how a person is carrying out the tasks assigned.

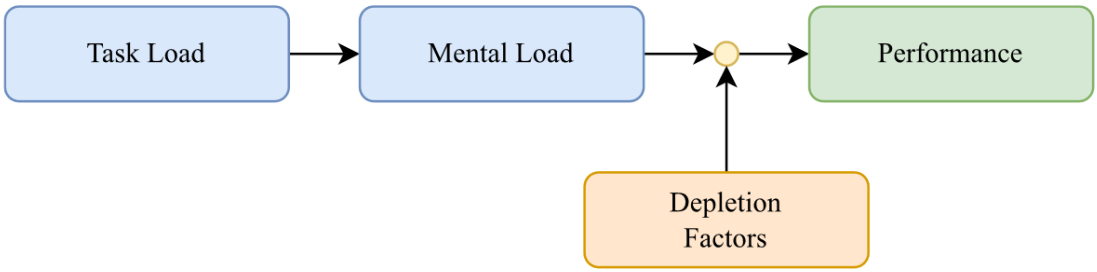

Figure 2.1: Stress and MWL relationship

2.4 Methods for stress and mental workload assessment

This study is concerned with measuring stress and MWL variations; that is why it is necessary to use tests to generate these states of alteration. Several tests have been developed depending on the aim and context in which they take place; for example, for aviation pilots, researchers developed different types of tasks, pursuit and tunnel tasks, to be performed in a flight simulator [13]. Always in the aeronautic field, air traffic controllers performed a Terminal Radar Approach Control [14]. All these aeronautical tests are explained in section 2.6. In more general areas, other types of tests have been developed; for example, in the study of Herten et al. [15] the Trier Social Stress Test (TSST) was assigned. The participants have the task of making a speech in public, which is a well-known stressful contest, in conditions of control, so with a "friendly" committee, or not. In section 2.6, other tests applied are explained: the Button Operation Experiment [16] and Digit Span Task [17]. Moreover, another way for generating altered states in the subjects is giving them cognitive tasks to carry out within a virtual environment [18] or not [19]. Two tests broadly applied for generating states of stress and high MWL are the Stroop Test [20] and the N-Back Test [21]; they are also the two tests used in this research.

2.4.1 Stroop Test

The Stroop Test is designed to create stress conditions in the participant. In the Stroop Test it is shown on the screen a word that means a color, colored any color. The participant must type the color of the word and not the color representing the meaning of the word. In Fig. 2.2 an example of Stroop Test interface is shown.

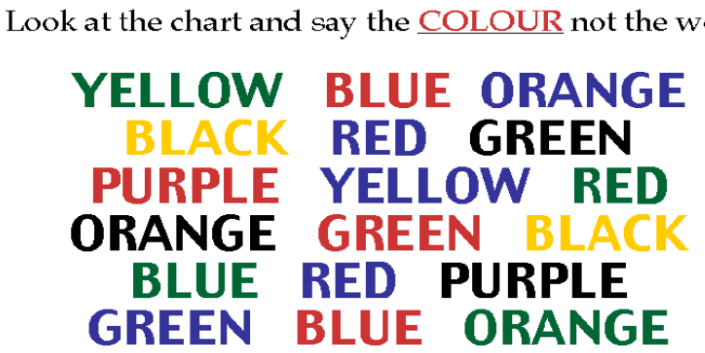

Figure 2.2: Example of Stroop Test Interface

2.4.2 N-Back Test

The N-Back Test is designed to create conditions with different levels of MWL in the participant. In the literature can be found several examples of this test applied to stress different perceptual channels: auditory, visual, etc. In all these kinds of tests the aim is to remember as many previous steps as possible. Fig. 2.3 an example of N-Back Test interface is shown.

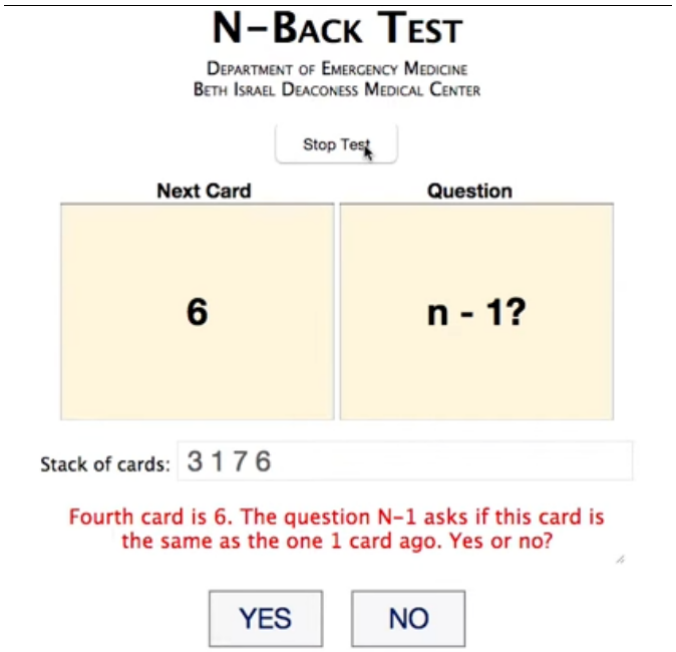

Figure 2.3: Example of N-Back Test Interface

2.5 Eyes

The eyes are one of the most essential communication and sensing tools of the human body. In general, the face, and in particular the eyes of people, are an important means of communicating their feelings and helping others understand what they are expressing [22]. The study of eye movements is, for example, used in areas such as marketing to know how a subject makes decisions and thus buys a product rather than another [23]. They are also studied in the medical field because changes in the timing of eye movements or the suppression of some of them can mean the presence of any pathological states [24].

The signals that can be extrapolated from the eyes are their movement and the size of the pupil. These signals can be acquired with different devices and methods: the oldest is the electrooculogram (EOG), which is deeply described in section 2.7. The most modern devices are eye trackers, which are used for measuring the position and movement of the eyes through different techniques.

2.5.1 Eye Movements

There are several types of eye movements that are distinguished mainly by their frequency, amplitude, and duration; they are explained below.

- **Fixations** are used to examine a visual scene, they are present when a person is staring at an object, and they last from 200 ms to several minutes. During fixation, the gaze does not remain perfectly fixed, but rather, it is in constant movement: these involuntary eye movements are called fixational eye movements and are divided into groups such as tremor, drift, microsaccades, and saccadic intrusions (SI) [25].

  Precisely, these small movements allow to capture the details of a visual scene. If a scene were completely static without a retinal image refresh, there would be a rapid adaptation to it [26]. In fact, if the retina always receives the same image, its receptors undergo uniform stimulation, and the visual scene dissolves from sight; this loss of visual information is contrasted by the fixational eye movements [27].

- **Saccades** are fast, jerky, and ballistic eye movements; they allow a shift of the gaze from one point to another. Their velocity reaches 50 angular degrees per second, and their duration ranges from 20 to 200 ms. These kinds of saccades are classified as "regular saccades" to distinguish them from other types of saccadic eye movements, as microsaccades and saccadic intrusions, explained below.

- **Microsaccades** and **SIs**, also being part of fixations, have the typical features of saccades; in fact, they are also rapid deviations of the gaze, but they do not change in substance its position; therefore the eye continues to stare at the same visual scene. They are present only on the horizontal axis of the eye and, they occur synchronously on both eyes in the same direction; it is therefore said that they are conjugated [25]. What differentiates microsaccades from SIs is mainly the amplitude of the eye movement. Microsaccades have amplitudes smaller than 0.4° [28]. Regarding SIs, their amplitude goes from 0.4° to 4.1°, and they have a dwell time, so the time where the gaze is away from the original position of the gaze, ranging from 60 to 870 ms [28]. They are composed of a first rapid shift of gaze position and a consequent coming back of the gaze to the original fixation point; the total path of the eye is rounded. Both microsaccades and SIs have a duration of the one-way path of 30 ms and occur 0.1 to 3 times a second [29].

  In Fig. 2.4, fixations, saccades, and SIs are graphically described.

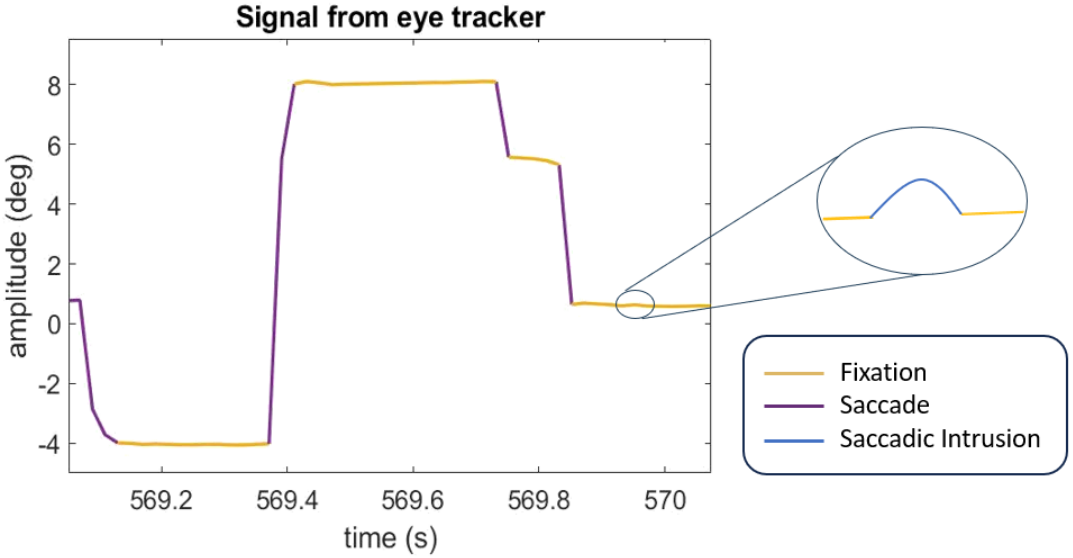

Figure 2.4: Graphical description of fixations, saccades and SIs

- **Tremor** is an aperiodic oscillatory eye movement with a bandwidth ranging from 30 to 80 Hz [28]. It has a small amplitude (15 minarc), and it is independent in the two eyes. Due to his amplitude and bandwidth, which are in the range of the recording system's noise, it isn't easy to detect it accurately [27].
- **Drift** is a slow and not-conjugated eye movement (1-3 minarc) [28] that occurs during the phases without microsaccades for maintaining accurate

visual fixation in their absence. Both tremor and drift might result from variability of neuronal firing to the ocular muscles and noise [27].

The characteristics of the eye movements just explained are resumed in Table 2.1.

Table 2.1: Summary table of characteristics of eye movements

Eye Movement	Amplitude	Lasting	Frequency	Velocity	Bandwidth
Fixation	<1°	200 ms - min	-	-	-
Saccade	>1°	20 - 200 ms	-	50 °/s	-
Tremor	15 minarc	-	-	-	30 - 80 Hz
Drift	1-3 minarc	-	-	-	0 - 0.5 Hz
Microsaccade	<0.4°	-	0.1 - 3 Hz	-	-
SI	0.4°- 4.1°	60 - 870 ms	0.1 - 3 Hz	-	-

2.6 State-of-the-art of eye movements assessment

A vast research in literature has been performed to understand which are the relationships between features extracted by ocular signal and variation of stress and MWL. From this analysis, it is possible to observe that the most studied features, in past and current studies, are related to blinking, pupil diameter adaptation, fixation and saccade eye movements, and to SIs. They are explained below:

Blinking

One of the main features related to blinking is the blink rate. In particular, the issue of blink rate variation with different levels of stress and MWL is very discussed:

- Brookings et al. [14] in 1996 dealt with the signal coming from eye movements acquired by EOG of eight air traffic controllers performing a Terminal Radar Approach Control. Their conclusions were that the blink rate decreases if MWL increases.
- Veltman et al. [13] in 1998 studied blink interval, defined as the time in between two successive eye blinks, and blink duration. The participants of this study were twenty pilots in a flight simulator; they had to carry out two kinds of trials. The first one is a pursuit task of a targeted jet at a large distance;

this task was expected to be easy because the considerable distance enabled a thorough anticipation of the maneuvers of the target jet. The second one was a tunnel task where they were required to fly in a tunnel, with varying levels of difficulty, with a continuous memory task with four levels of difficulty. EOG acquired the eye movements, and the results were that blink interval and blink duration decrease if MWL increases.

- Recarte et al. [19] in 2008 carried out two different trials: the first one was a cognitive task with no visual demand, while the second one was a cognitive task with visual need. The cognitive task could be listening, talking, or calculating. The device used is the eye-tracking system ASL 5000 with a video recorder. They found that the blink rate becomes bigger if MWL increases, but if a visual task is performed, the blink rate decreases.

Pupil diameter

First of all, this feature is very sensitive to variations of stress and MWL but, on the other hand, very sensitive to changes in the environment's illumination. The lighting of the domain is complex to keep constant; that is the reason why this problem is difficult to overcome and can fake the results. Nowadays, researchers are trying to find new sensitive features that are less dependent on external stimuli.

- Recarte et al. [19] also focused on the variation of pupil dimension with a variation of MWL: they claim that there is a direct proportionality between them.

- Tokuda et al. [29][25] in 2009 and later in 2011 carried out trials where participants were asked to perform N-Back Tests while staring at one point or while moving their gaze over a picture. The device used is the eye-tracker Tobii 1750 with a 50 Hz resolution. They found out that pupil diameter becomes bigger if MWL increases.

- He et al. [16] in 2012 studied, among other features, the relative dimension of the pupil and came to the same conclusions as the previous papers. The test they implemented was a Button Operation Experiment where subjects were required to press the keyboard key corresponding to the letter appearing on the screen; the time pressure varied to increase the difficulty of the task.

- Coyne et al. [17] in 2016 enrolled ten volunteers who participated in an experiment in which a Digit Span Task was employed to manipulate MWL. In this study, they performed the auditory version where to participants were presented series of numbers and then instructed to recall the digits in the same order which they were heard. Eye movements were recorded through the eye-tracker Eye Tribe. Researchers found out that pupil diameter increases if MWL increases.

- Hirt et al. [18] in 2020 used a virtual environment where there were different stressors and tasks to fulfill, to induce stress in participants. The setup consisted of an HTC Vive with an integrated eye tracker from Pupil Labs that works at 200 Hz. Their findings were that pupil diameter increases if MWL increases.

Fixation

Being one of the main groups of eye movements, fixations are deeply investigated. In particular, their frequency and lasting are studied.

- Recarte et al. [30] in 2003 studied the effects of MWL on visual search and decision-making in real traffic conditions with twelve participants who drove an instrumented car. An unobtrusive eye-tracking system (Applied Science Laboratories (ASL), Bedford, MA) with 50 Hz resolution was used. When more relevant traffic targets were in the visual field, eye fixations on roadside advertisements (irrelevant peripheral objects) were significantly reduced in favor of traffic information. Moreover, tasks with high spatial imagery content produced not only more pronounced effects but also a particular pattern of long fixations.

- He et al. [16] studied fixations frequency (pcs/min) and average fixation time; they found that both these features increase if MWL increases.

SI

SIs are a bit less investigated than the other eye movements:

- Tokuda et al. [25] found that features related to SIs are very sensitive to variations in MWL. In particular, they studied the SI value: it is the average amplitude in the evaluation period of a trial and represents the SI behavior. SI value has a unit of measurement of deg/sec, and it increases if MWL increases.

Saccade

Being one of the main groups of eye movements, saccades are deeply studied. In particular, their frequency and velocity are investigated.

- He et al. [16] were also interested in saccadic frequency (pcs/s) and average saccades velocity (rad/s); they concluded that they increase if MWL increases.

Table 2.2 shows the trend of the features already investigated in the literature.

Feature	Trend
Blinking frequency	↑ ↓
Blinking lasting	↓
Blinking interval	↓
Fixation frequency	↑
Fixation lasting	↑
Saccades frequency	↑
Saccades velocity	↑
SI value	↑
Pupil diameter	↑

Table 2.2: Features trend showed by the literature

2.7 Eye-Tracking: history, devices and applications

Eye tracking allows recording eye movements and analysing human processing of visual information for interactive and diagnostic applications. More precisely, this technology's application fields are neuroscience and experimental psychology, human-computer interaction, and still computer vision and marketing. The methods used to perform eye-tracking in the first studies concerning the movements of the eyes were based on invasive technologies that required, for example, the use of contact lenses and chin guards or bite bars to prevent the movement of the head and, therefore, any artifacts due to them. An example is the EOG [31].

The EOG is a technique that allows the recording of eye movements by measuring the electrical potential present between the cornea and the posterior pole of the eye. Since this biopotential is proportional to the angle of rotation of the eye, the fixation points can be found from the recorded signal. In Fig. 2.5 it is explained the EOG principle of operation. An advantage of EOG over eye trackers is that it identifies the fixation point even if the eyelashes are closed, and the surrounding environment is dark. It is also easy to configure as it consists of two electrodes placed on the right and left temples of the subject; the measured potential is amplified by DC or AC amplifiers. The disadvantages of this method of recording, which makes it not so popular, are the noise created by the movement of facial muscles and blinking and the drift present in long recordings when using DC amplifiers [32].

Due to all these limitations, the discomfort caused to the subject, and the fact that these technologies were applicable only to specific contexts, new eye-tracking

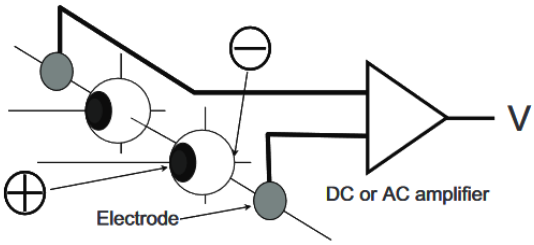

Figure 2.5: EOG principle of operation

methods were developed, moving towards video-based techniques. Around 1935, with the rise in popularity of film recordings, the first non-contact eye-tracker was built: it was based on the reflection of beams of light from the cornea and their recording on moving films. Later, in the post-second World War era, the issue of immobilizing the head during eye-tracking experiments was no longer a problem thanks to the development of the first head-mounted eye-tracker. However, there was still the issue of discomfort due to the use of a mouth plate [33]. Nowadays, the newer video-based techniques used for eye-tracking are less invasive, more comfortable, and more accurate; in addition, they have different possible configurations: there are head-mounted, desk-mounted, and eyeglass-mounted high-speed cameras [31]. Moreover, today, there is a move towards greater accessibility of eye-tracking by the public; in fact, software that, without the help of additional devices and sensors, works on hardware such as phones or tablets, has been developing. However, the accuracy of the tracking of eye movements differs from that obtained with video-based techniques [34].

2.8 Data set acquisition

This thesis is part of the BiLoad research project of the Politecnico di Torino. The protocol on which the acquisitions carried out in the study are based has been approved by the Politecnico itself with a Protocol Number 1606. The Stroop Test and the N-Back Test interfaces have been implemented during a previous step of the project to generate stress and MWL states. In this section, the acquisition protocol will be described; in particular, there will be a focus on the different biosignals acquired, on Tobii Glasses 3 and Tobii Pro Lab that are, respectively, the devices used for acquiring the signal and for processing it, and the structure of the trial. In Fig. 2.6 and 2.7 the trial setup is shown.

Figure 2.6: Trial setup: side shot

Figure 2.7: Trial setup: frontal shot

2.8.1 Biosignals acquired

The participants were asked to perform two different kinds of tests, the Stroop Test and the N-back Test, while wearing sensors costumed for acquiring various biosignals. All the signals acquired, except to the ocular signal, were recorded through Biosignalplux Professional KIT [35].

- **fNIRS** is the signal that measures the oxygenation of the brain and, consequently, the most used areas in a particular moment. The signal is collected with a single sensor placed on the forehead and held down through a belt; it uses two different lights, one red and one infrared.
- **Respiration** is measured by a belt placed around the chest and closed with a sensor consisting of an estensimeter that detects the variation of the dimension of the chest and relates it with a variation of impedance. In this way, it allows the estimation of the inhaled air volume.
- **Body temperature** is acquired by a thermocouple placed on a hand finger; the principle of its functioning is that if the finger's temperature changes, the voltage measured by the sensor changes.
- **Electrodermal Activity (EDA)** represents the electrical activity of the skin surface. This signal is measured by two electrodes placed on the forefinger and the middle finger of one hand.
- **Electrocardiogram (ECG)** indicates the electrical activity of the heart, measured on the body surface. Three electrodes placed on the upper left side of the chest are used.
- **Eyes Movements** are recorded by Tobii Glasses 3 eye-tracker worn by the participant.

2.8.2 Tobii Glasses 3

Eye movement data are taken using the Tobii Glasses 3 eye-tracker [36], shown in Fig. 2.8, and considered the gold standard for acquiring data on fixations, saccades, and the subject's pupil size. Tobii Glasses 3 is a wearable eye-tracker for academic, commercial, and industrial use to capture truly objective and deep insights into human behavior in a real-world environment. The glasses are designed to capture what the wearer is viewing while providing robust and accurate eye-tracking data. The complete Tobii Glasses 3 system consists of a head unit, a recording unit, and the Tobii Glasses 3 controller application installed on an external device. Tobii Glasses 3 collects the visual data of the wearer and records a video with audio of the surrounding environment. The time resolution is 50 Hz while the spatial

Figure 2.8: Tobii Glasses 3 setup and his additional devices

resolution is 0.6°, the video resolution is 1920x1080 at 25 fps and the Field of View (FOV) is 63° vertically and 95° horizontally. In Table 2.3, the characteristics just named of resolution and Field of View of the device Tobii Glasses 3 are listed. It is

Table 2.3: Tobii Glasses 3 characteristics of resolution and Field of View

Resolution			**Field of View**	
Time	Spatial	Video	Vertical	Horizontal
50 Hz	0.6°	1920x1080	63°	95°

based on the most widely used technique PCCR [37], the corneal reflection of the center of the pupil. This method uses a light source close to infrared to illuminate the eye, thus causing evident reflections, which are highlighted in the pupil and cornea and photographed by an infrared camera. The image obtained allows to identify the reflections of the light source. Then it is performed the calculation of

the vector, composed of the angle between the two reflections on the different parts of the eye: the direction of the vector is then used to determine the direction of the gaze [33].

The coordinate systems used by Tobii Glasses 3 are two: the first one is called Media Coordinate System (MSC), it is a 2D system with its origin in the upper left corner of the FOV; the x-axis is horizontal and directed towards the right while the y-axis is vertical and downward, as shown in the Fig. 2.9.

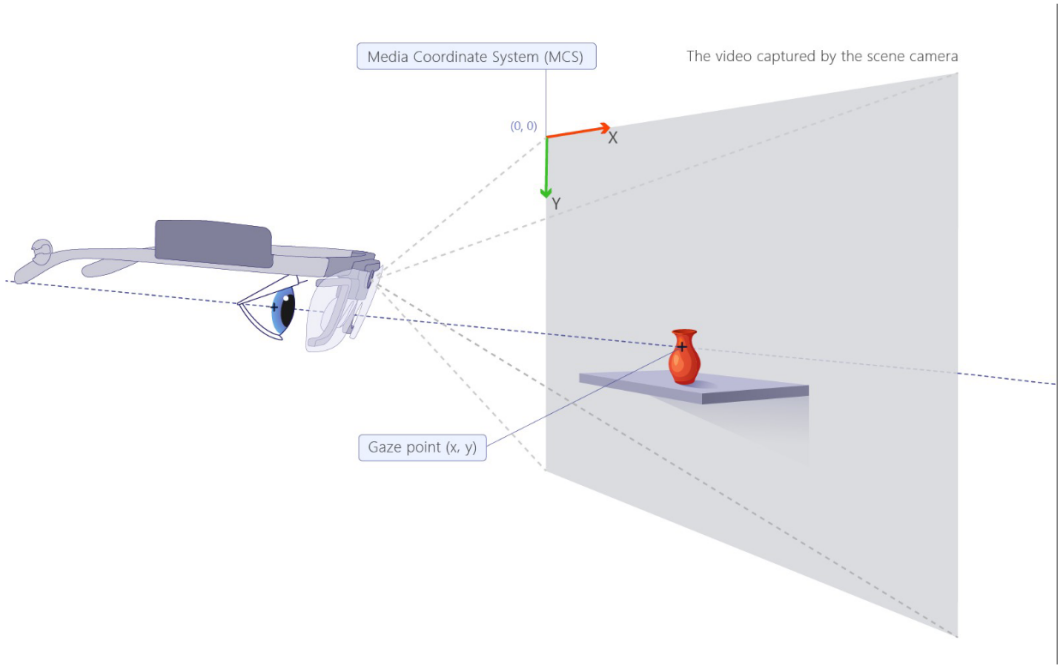

Figure 2.9: Tobii Glasses 3 coordinate system MCS

The second is the Head Unit Coordinate System HUCS which is a 3D coordinate system with its origin in the center of the scene camera; the axis are directed as shown in the Fig. 2.10. In both coordinate systems the gaze location uses pixels as measurement unity. The data exported from Tobii Glasses 3, called Raw Data, has different fields, as explained below:

- Timestamp: it is the instant related to an eye movement, time axis should have samples at 20 ms intervals, but some samples are missing due to the large amount of data recorded;
- Position of the gaze according to the MSC coordinate system in pixels (gaze2D);
- Position of the gaze according to the HUCS coordinate system in pixels (gaze3D);

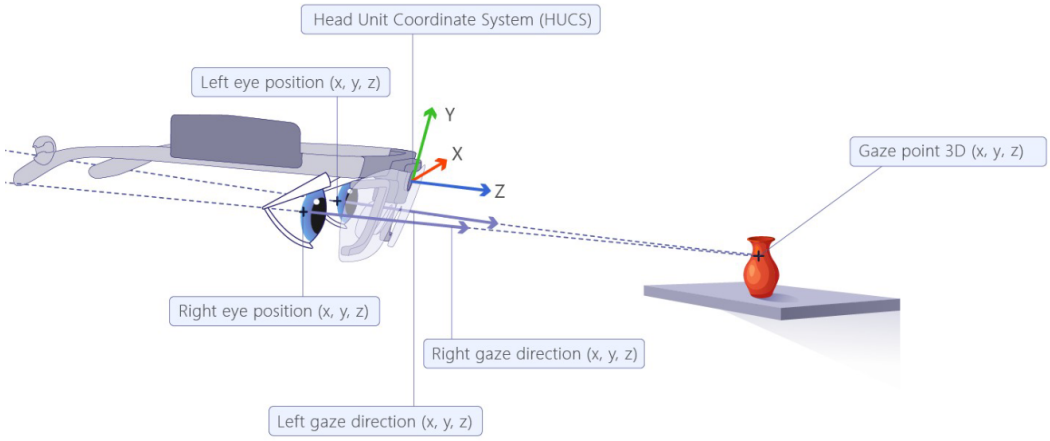

Figure 2.10: Tobii Glasses 3 coordinate system HUCS

- For both, the left and right eye, the gaze origin, the gaze direction, and the pupil diameter.


```
{"type":"gaze","timestamp":0.000828,"data":{}}
{"type":"gaze","timestamp":0.060991,"data":{}}
{"type":"gaze","timestamp":0.161118,"data":{"gaze2d":[0.59229453025255785,1.0555090697581364]
{"type":"gaze","timestamp":0.201190,"data":{"gaze2d":[0.60147452033237259,1.0660505181707234]
{"type":"gaze","timestamp":0.221281,"data":{"gaze2d":[0.6032623581830614,1.0766060906391814],
{"type":"gaze","timestamp":0.241263,"data":{"gaze2d":[0.60594676478856191,1.0836833790895679]
{"type":"gaze","timestamp":0.261353,"data":{"gaze2d":[0.60621819516042919,1.0852039983193138]
```


Figure 2.11: Extract of raw data

In Fig. 3.9 the first fields of raw data are shown: the timestamp and the gaze position in the MCS coordinate system. In particular, the first two rows don't show the gaze coordinates because the eye tracker couldn't detect it during the acquisition phase, this is due to eyelash interference or blinking.

2.8.3 Tobii Pro Lab

Tobii Pro Lab software [38] is adopted for eye-tracking data processing. In general, it receives input recordings done by Tobii Glasses 3 and provides a visual user interface and dedicated features extraction software that efficiently guides and supports the user through all phases of eye-tracking data management and processing.

It is an efficient tool specifically designed for this kind of application. The interface of Tobii Pro Lab is shown in Fig. 2.12 with all the indications on its components. Two predefined filters can be applied to raw data: both are based on velocity threshold algorithms, later explained in Section 3.2, and they are different

Figure 2.12: Interface Tobii Pro Lab: 1) Data Export; 2) Metrics; 3) Recording information; 4) Snapshot; 5) Events list; 6) TOIs; 7) TOIs shown graphically; 8) Events shown graphically

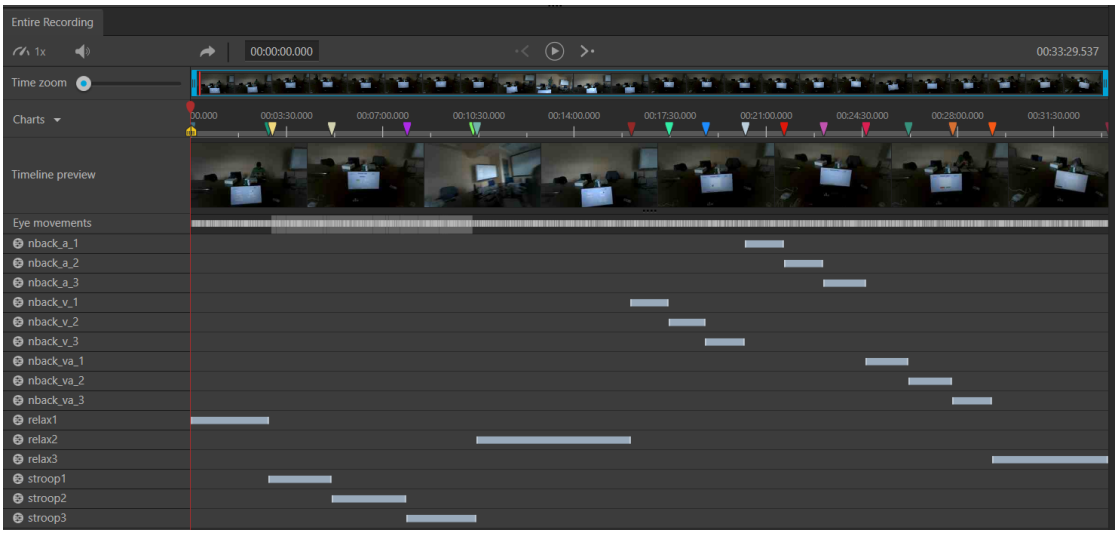

Figure 2.13: TOIs and events shown graphically

just for the threshold they apply in order to distinguish fixations and saccades. "Fixation" filter has a threshold of 30 degrees/second while "Attention" filter of 100 degrees/second. The first one is used in situations where the subject wearing the glasses is not moving a lot, while the second is used if, for example, the participant

is walking or doing other physical activities. Tobii Pro Lab offers the possibility to

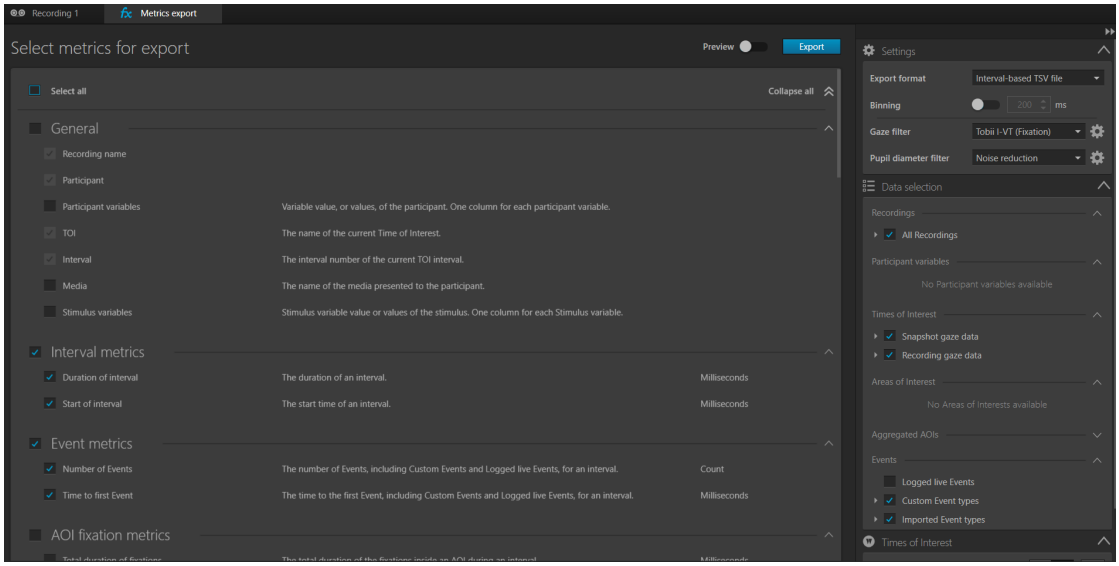

Figure 2.14: Metrics file creation

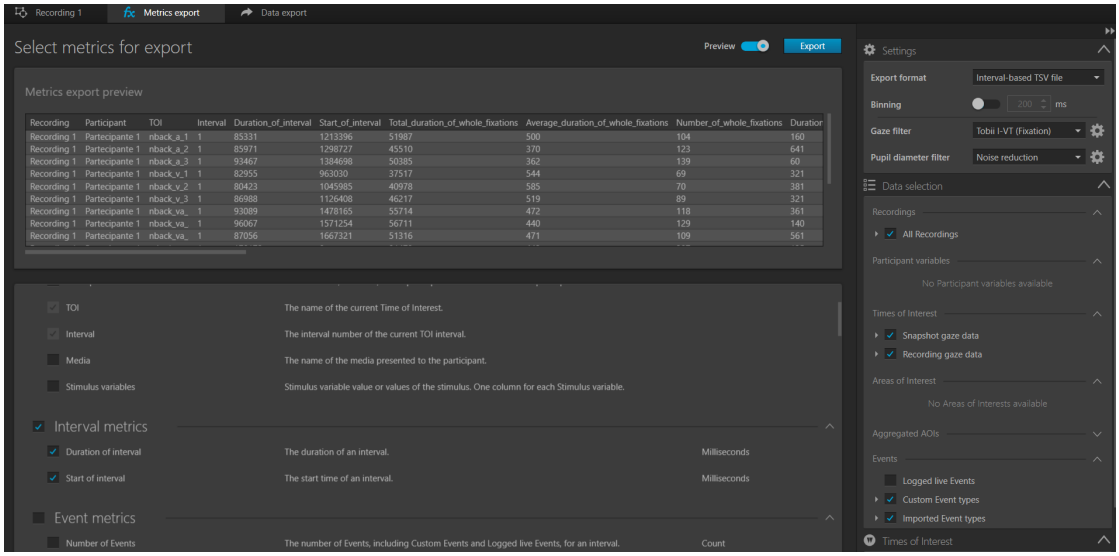

Figure 2.15: Data Export file creation

the user to design other filters, different from the predefined ones, according to the application and the data quality. These filters are used to classify eye movements; in fact, Tobii Pro Lab processes data from the eye-tracker and performs a classification of eye movements. It distinguishes among "Fixations", "Saccades", "Unclassified",

Recording timestamp	Computer timestamp	Sensor	Gaze point X	Gaze point Y	Pupil diameter left	Pupil diameter right	Validity left	Validity right	Eye movement type
828	828	Eye Tracker					Invalid	Invalid	EyesNotFound
60991	60991	Eye Tracker					Invalid	Invalid	EyesNotFound
161118	161118	Eye Tracker	1137	1140	4,303	4,213	Valid	Valid	Fixation
201190	201190	Eye Tracker	1155	1151	4,246	4,27	Valid	Valid	Fixation
221281	221281	Eye Tracker	1158	1163	4,278	4,247	Valid	Valid	Fixation
241263	241263	Eye Tracker	1163	1170	4,265	4,257	Valid	Valid	Fixation
261353	261353	Eye Tracker	1164	1172	4,265	4,244	Valid	Valid	Fixation
281335	281335	Eye Tracker	1165	1174	4,277	4,265	Valid	Valid	Fixation
401553	401553	Eye Tracker	1172	1180	4,298	4,275	Valid	Valid	Fixation
421644	421644	Eye Tracker	1177	1178	4,297	4,306	Valid	Valid	Fixation

Figure 2.16: Data Export from Tobii Pro Lab

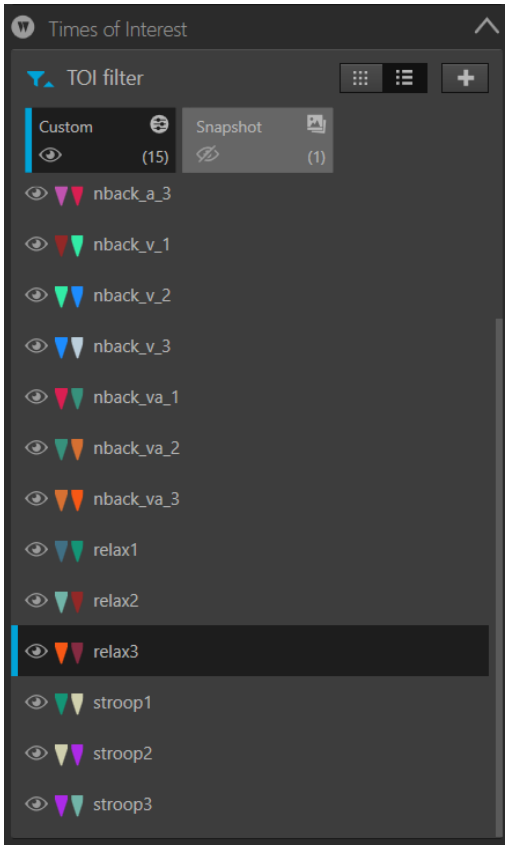

Figure 2.17: TOIs creation

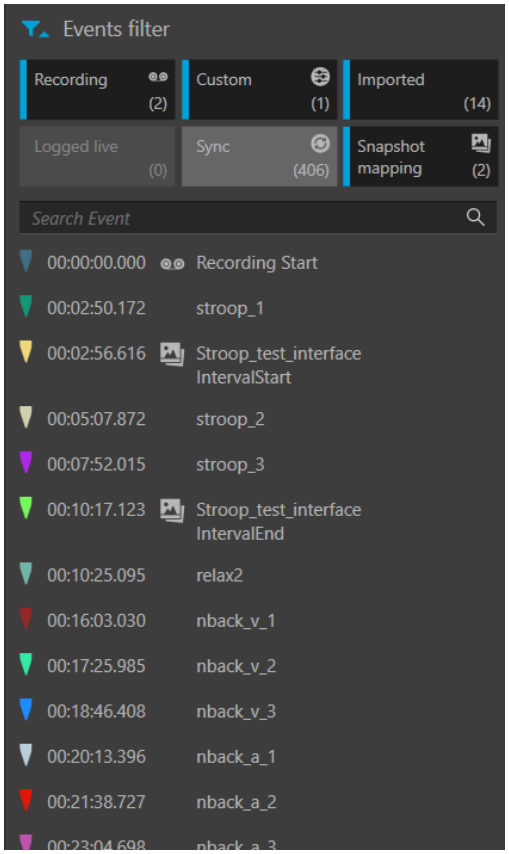

Figure 2.18: Events creation

and "Eyes Not Found". The samples that result as "Unclassified" do not match the requirements for being classified as saccades and neither for fixations. "Eyes Not Found" (ENF) is the label to indicate a non-detection of the gaze by the eye-tracker; this could be due to blinking, eyelash interference, or other reasons. Another tool is the possibility of performing an Assisted or Manual Mapping of the fixations in a snapshot. Snapshots are still images of environments and objects of interest; they can be imported in box 4 if Fig. 2.12. It is also possible to create Events, Time of

Interest TOIs and Areas of Interest AOIs by mapping eye movements. Moreover, data chosen by the user, if needed, even related to TOIs and AOIs, can be exported in TSV files. The user can also decide if to export them in the MSC or the HUCS coordinate systems; in both coordinate systems, the gaze location uses pixels as measurement unity. TOIs creation is possible through the boxes 6 and 7 in Fig. 2.12, that are better shown in Fig. 2.17 and 2.13. Events are created through boxes 5 and 8 in Fig 2.12, which are better shown in Fig. 2.18 and 2.13. Finally, it extracts a set of features about saccades, fixations, glaces, and visits from data. An example of Data Export is shown in Fig. 2.16 where among the fields exported there are the timestamp, the gaze position in MCS coordinate system, the pupil diameter right and left, and the eye movement type. The interfaces for exporting metrics and data are in Fig. 2.14 and 2.15 and they can be obtained from boxes 1 and 2 in Fig 2.12.

2.8.4 Structure of the trial

The first very important part of the test session is the explanation of the tests the participants have to perform. This is followed by the filling of the documents regarding the privacy policy by the participants themselves, as wanted by the approved protocol. All the sensors of Biosignalplux Professional KIT and the Tobii Glasses 3 are worn by the subjects. The trial is structured in different phases:

1. **Relax phase**: for recording the baseline of every biosignal, it lasts 2 minutes for the first six participants and 3 minutes for the others.

2. **Stroop Test phase**: the Stroop Test principle as been already explained in section 2.4.1, in particular, the test designed for this study consists of a word shown in the central upper part of the screen, whose meaning is a color, but its color could be different from the meaning. The participant has to decide among five different possibilities of colors that are shown in the part of the screen below. The test aims to click on the color of the region below corresponding to the meaning of the upper word. The Stroop Test interface adopted in the research is shown in Fig. 2.19.

   At the beginning of the test there is a rest of 5 seconds. Then, it is divided into three levels designed to be the first one the less stressful and the last one the most stressful. In particular, the first one is based on congruent words and noise, the second one on incongruent words and noise while the third one on incongruent words, noise and color names playing. Each level is divided from the others by a short phase of relaxation lasting 15 seconds.

3. **Relax phase**: for coming back to a state of relaxation, it lasts 3 minutes.

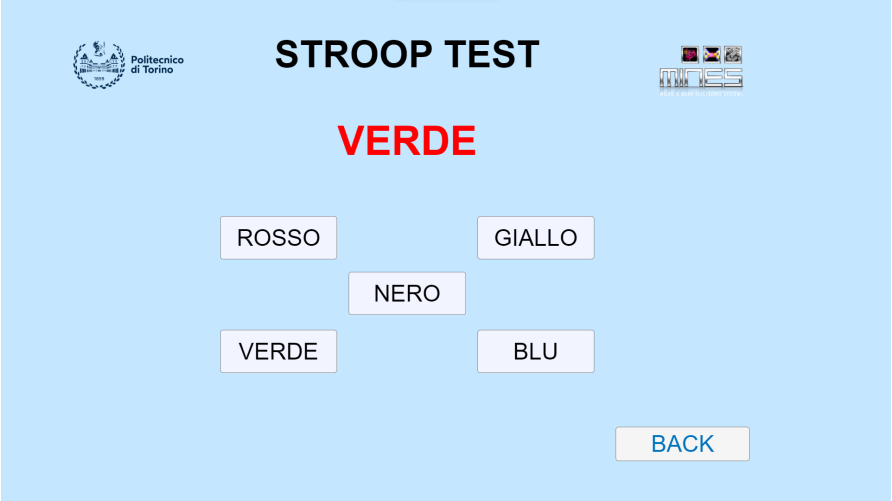

Figure 2.19: Stroop Test interface

4. **N-Back Test phase**: based on the examples of N-Back Tests in literature explained in section 2.4.2, in this study the test was developed in three different versions: visual, auditory, or both visual and auditory tasks. Moreover, it was decided to structure the tests as follows:

   - **Visual N-Back Test**: it consists of a rectangle that, every two seconds, moves in a grid-changing position.
   - **Auditory N-Back Test**: it consists of listening to a set of letters through earphones.
   - **Dual N-Back Test**: it consists of a combination of visual and auditory tasks, where at the same time, both previous tasks must be managed.

   The test aims to remember the order of apparition of the rectangle or the order in which letters are heard, for remembering the N-th position of the rectangle before the current one or the N-th letter heard before the current one. The Dual N-Back Test interface is shown in Fig. 2.20.

   The first group of the N-Back Test is the visual one, the second group is the auditory one, and the last group is a combination of visual and auditory tasks. Each group has three different levels of difficulty, from the easiest one to the most difficult one (N=1,2,3).

   At first, there are 5 seconds of rest; then, there are the three levels of the Visual N-Back Test divided each one with 5 seconds of rest. Before the Auditory N-Back and the Dual N-Back there are 10 seconds of rest, while each level is divided from the others with 5 seconds of relaxation.

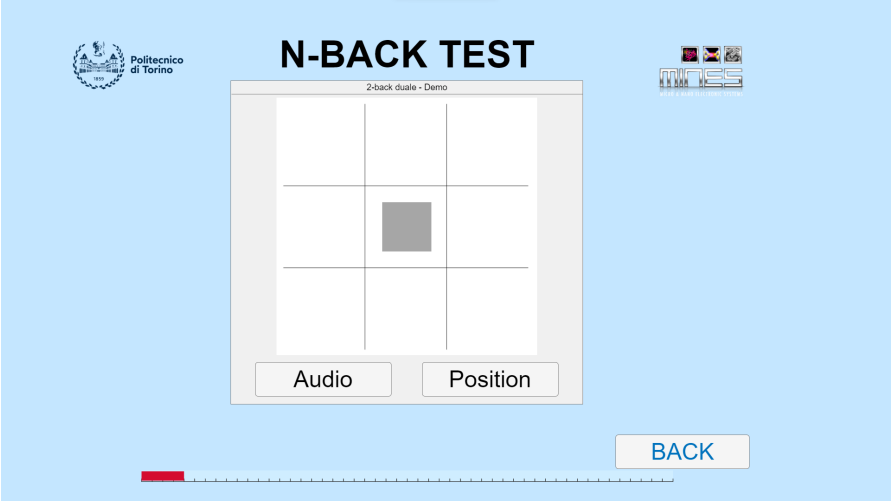

Figure 2.20: Dual N-Back Test interface

5. **Relax phase**: for coming back to a state of relaxation, it lasts 20 seconds. During it, the participant can look at his results.

6. **Questionnaire**: At the end of the trial, the participant has to fill in a questionnaire where he is asked to match each level of the tests with one of the three possible levels of "difficulty". This "difficulty" is not just a simple concept of effort needed to do the tests but is differentiated by stress and MWL. In Fig. 2.21 and 2.22 two parts of the questionnaire are shown: the request of personal data to the participant and the request of the level of "difficulty" faced in each test.

In Fig. 2.23 the flowchart of the acquisition protocol is explained.

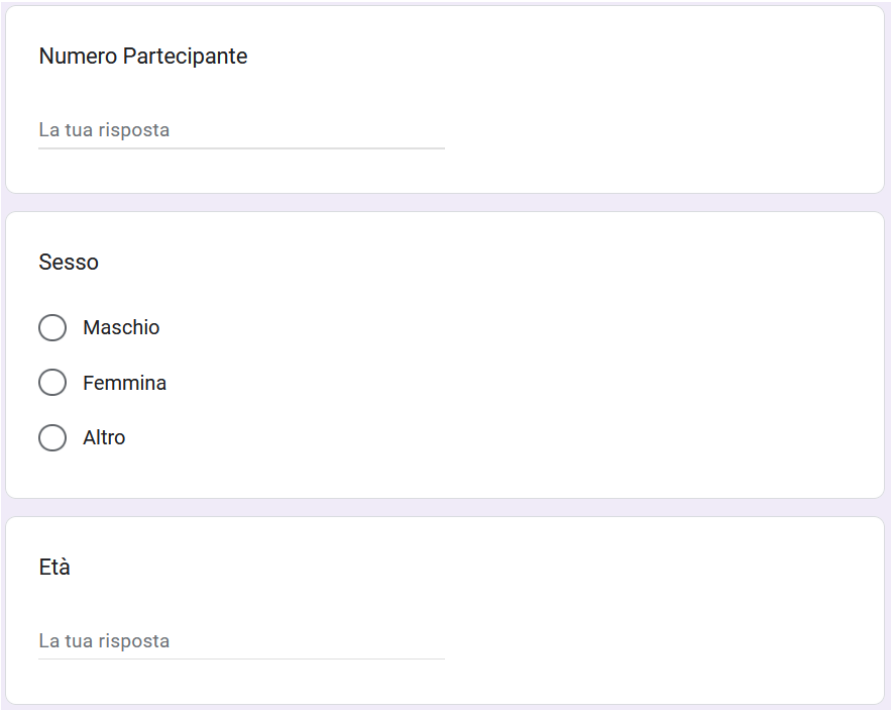

Figure 2.21: Questionnaire: personal data

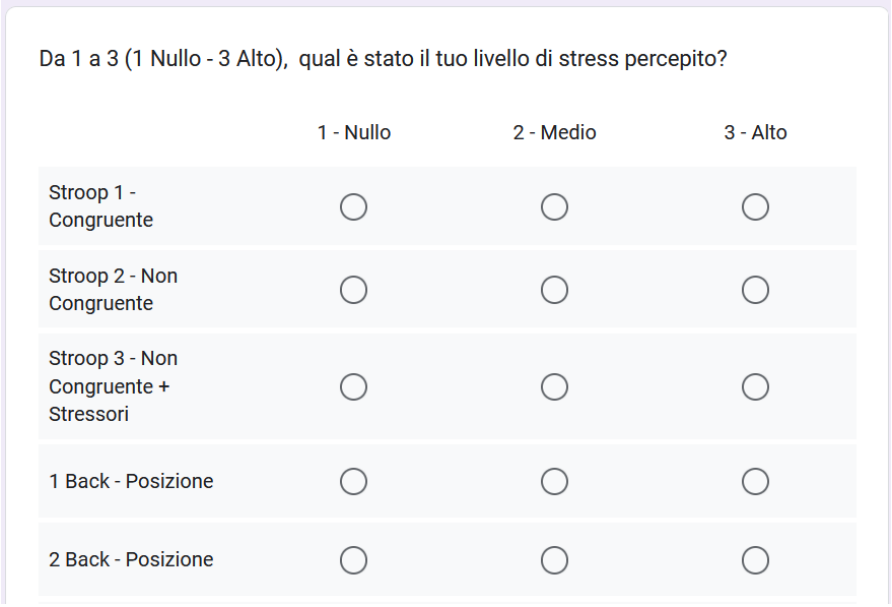

Figure 2.22: Questionnaire: level of "difficulty"

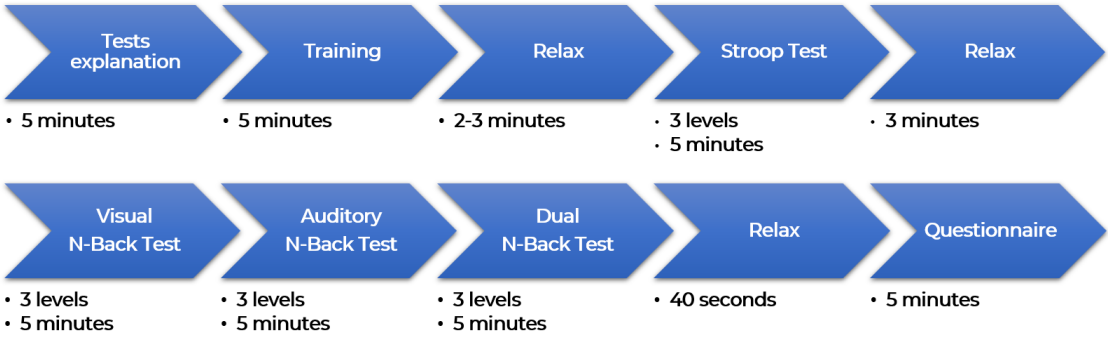

Figure 2.23: Flowchart acquisition protocol

2.9 Different kinds of normalization

To find the most meaningful representation of the data a study about four different kinds of normalization is carried out during dataset investigation. The types of normalization are:

- *Z-score normalization or standardization*: the formula for passing from the original data vector X to the standardized one Z is Equation2.1.

$$z_i = \frac{x_i - \mu}{\sigma} \tag{2.1}$$

  where z_i is the value standardized, x_i is the original value of the dataset X, μ is the mean of all the elements in X and σ is the standard deviation of the elements in X.

  Z is the distance of a data point from the mean in terms of the standard deviation. The standardized data vector has a mean equal to zero, a standard deviation equal to one, and retains the shape properties of the original dataset, in particular skewness and kurtosis. Moreover doesn't have a fixed range, isn't much affected by outliers and it is the best kind of normalization if the distribution of the data is gaussian.

- *Euclidean norm standardization*: the formula for passing from the original data vector X to the standardized one S is Equation 2.2.

$$s_i = \frac{x_i}{||x||_2} \tag{2.2}$$

  where s_i is the value standardized, x_i is the original value of the dataset X and $||x||_2$ is the Euclidean norm of the elements in X.

The Euclidean norm is calculated in Equation 2.3.

$$||x||_2 = \sqrt{\sum |x_i|^2} \tag{2.3}$$

The standardized dataset has no fixed range and has an Euclidean norm equal to zero.

- *Mean normalization*: the formula for passing from the original data vector X to the standardized one S is Equation 2.4.

$$s_i = \frac{x_i - \mu}{max(X) - min(X)} \tag{2.4}$$

where s_i is the value standardized, x_i is the original value of the dataset X and μ is the mean of all the elements in X.

The standardized dataset has no fixed range.

- *Min-Max Scaling*: the formula for passing from the original data vector X to the standardized one S is Equation 2.5.

$$s_i = \frac{x_i - min(X)}{max(X) - min(X)} \tag{2.5}$$

where s_i is the value standardized and x_i is the original value of the dataset X.

The new standardized data are scaled in a range [0,1] if there are just positive values. It is easily affected by outliers but is the best normalization if the distribution is unknown.

2.10 Comparison Tobii Pro Lab - Algorithm

In this thesis project, an algorithm that extracts the features from the ocular signal starting from the classification of eye movements is developed. Saccades and fixation points found with the developed algorithm are compared with the ones found by the gold standard Tobii Pro Lab. The parameters calculated are accuracy, sensitivity, specificity, PPV and NPV, and then the F1 score.

The generic confusion matrix is shown in Fig. 2.24 and the equations for the calculation of each parameter are 2.6, 2.7, 2.8, 2.9, 2.10, 2.11.

$$accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{2.6}$$

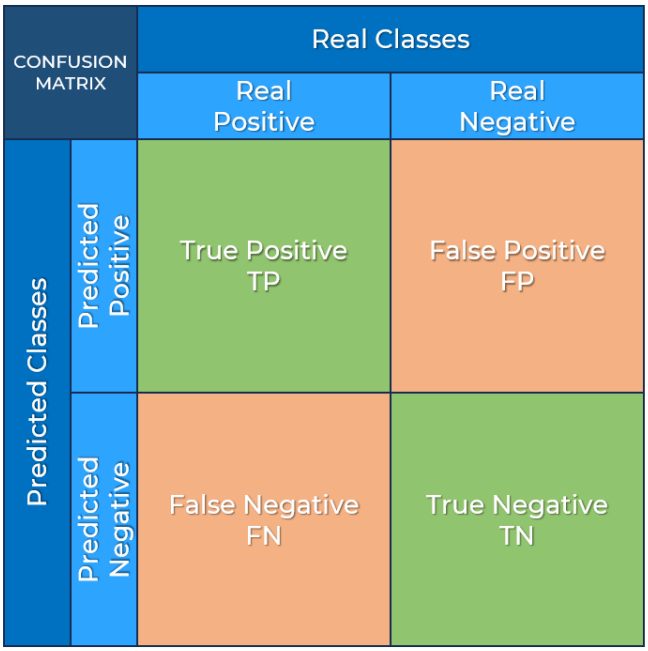

Figure 2.24: Confusion matrix

$$sensitivity = \frac{TP}{TP + FN} \tag{2.7}$$

$$specificity = \frac{TN}{TN + FP} \tag{2.8}$$

$$PPV = \frac{TP}{TP + FP} \tag{2.9}$$

$$NPV = \frac{TN}{TN + FN} \tag{2.10}$$

$$F1score = \frac{2 * PPV * sensitivity}{PPV + sensitivity} \tag{2.11}$$

where TP are the eye movements correctly identified as the eye movements of interest by the algorithm with respect to the gold standard; TN are eye movements correctly identified as not being the eye movements of interest by the algorithm with respect to the gold standard; FP are the eye movements incorrectly identified as the eye movements of interest by the algorithm respect to the gold standard; FN are eye movements incorrectly identified as not being the eye movements of interest by the algorithm respect to the gold standard.

2.11 Statistical Analysis

Statistical analysis is an important step for the analysis of results. In fact, it allows to generalize the results found for a population sample to the whole population. It has been carried out with statistical methods widely used in research. At first, a Kruskal-Wallis H Test [39] is chosen because the distribution of the classes for each feature is not Gaussian, so an ANOVA test is not suggested. Then, a Python code is selected because the classes are not equally populated and, in this language, the function of this statistical test supports this condition. The calculation of statistical parameters, *p-value* and *H* statistic, is done for each feature, for each data set, and differentiating between stress and MWL.

2.11.1 Kruskal-Wallis H Test and Mann-Whitney U Test

The Kruskal-Wallis H Test returns two outputs; one is the *p-value* and the second is the *H* statistic. The null hypothesis says that the data in each class of the dataset comes from the same distribution; the alternative hypothesis is that not all samples come from the same distribution. The null hypothesis is rejected if the *p-value* is smaller than the level of significance (α). The level of significance is chosen to be 0.05. The Kruskal-Wallis H Test is performed on four classes, this means that if the *p-value* is smaller than the significance level, at least one couple of classes doesn't come from the same distribution. *H* statistic is calculated with the equation 2.12.

$$H = \frac{12}{N * (N+1)} * \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3 * (N+1) \tag{2.12}$$

where H is the statistic of Kruskal-Wallis H Test, k is the number of classes, n_i the number of observations of class$i$, N is the sum of all the n_i with i going from 1 to k, R_i the rank assigned to class i. H statistic follows the χ^2 distribution depending from the level of significance and the degrees of freedom, equal to k-1. H statistic increases if the statistical difference between classes increases either and, in particular, if H statistic is higher than χ^2 distribution corresponding to a level of significance of 0.05 and a number of degrees of freedom equal to 4-1=3, then the null hypothesis is rejected. In particular, with a level of significance of 0.05 and a number of degrees of freedom equal to 3, then the H statistic limit is 7.815. If the H statistic is higher than this number, then the feature results to be significant.

To better understand which and how many couple of classes are statistically different in each data set, further statistical analysis post hoc is carried out. A Mann-Whitney U Test [40] is performed on each couple of classes in each data set and, even for this test, if the *p-value* is smaller than the level of significance the null hypothesis is rejected in favor of the alternative hypothesis, where the null and the alternative hypotheses are the same as the previous ones.

Chapter 3

Materials and Methods

This chapter describes how the validation of pixels-degrees conversion is carried out. Then the signal pre-processing and processing to extract features from it, are explained. Finally, the dataset is described.

3.1 Validation of Pixels - Degrees Conversion

Tobii Pro Lab Data Export gives the coordinates of each gaze point according to the MCS reference system originating in the upper left corner of the Field of View (FOV). The x-axis is directed horizontally, and it is positive on the right side of the origin, while the y-axis is directed downwards. The developed algorithm, explained in section 3.6, works with degrees coordinates with a reference system centered in the center of the FOV, with the x-axis directed horizontally and positive on the right side of the origin and the y-axis directed upwards. Considering the two different reference systems, a conversion from the first to the second one, and between pixels and degrees should be carried out. The different coordinate systems and the passing from the first described to the second one are shown in Fig. 3.1.

Diaz et al.[41] deal with this kind of double conversion and, referring to Fig. 3.2 they reached the formulas 3.1 and 3.2:

$$\phi_S = -tan^{-1}\frac{2y - pix_Y}{2d} \tag{3.1}$$

$$\theta_S = tan^{-1}\frac{2x - pix_X}{2d} \tag{3.2}$$

Where ϕ_S is the angle on the vertical plane that links the participant to the screen and θ_S is the angle on the horizontal plane that connects the participant to the screen. x and y are the gaze coordinates in the MCS coordinate system, while

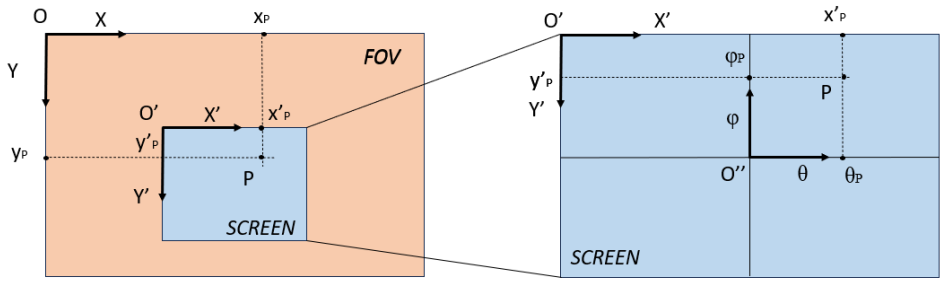

Figure 3.1: Reference systems of the Field of View of Tobii Glasses 3 and of the snapshot representing the screen of the PC

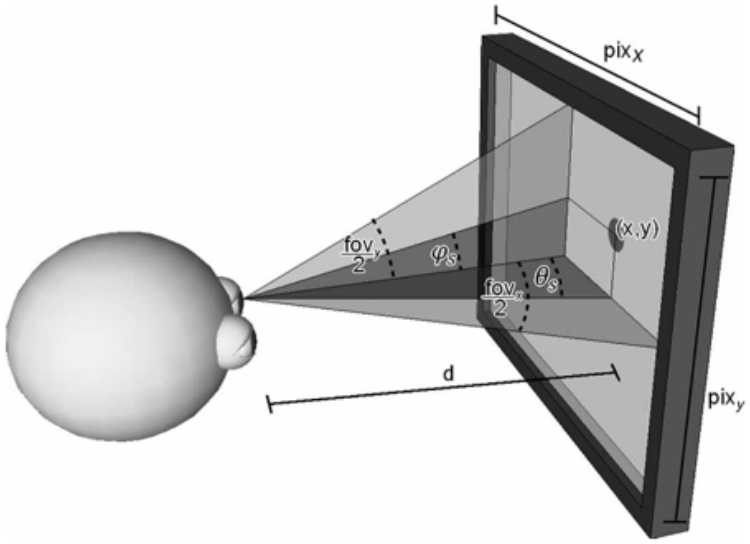

Figure 3.2: Reference systems explained of Diaz et Al. paper

pix_x and pix_Y are the width and the height of the FOV with the MCS reference system with origin in the upper left corner.

In order to validate these formulas, a test has been carried out. Two Matlab interfaces have been created; a rounded marker moves horizontally and vertically, and his location, which changes every 5 seconds, is known step by step. These Matlab interfaces are full-screen, this means that their resolution is 1920x1080 pixels as specified by the PC specificities. The interfaces are shown in Fig. 3.3 and 3.4.

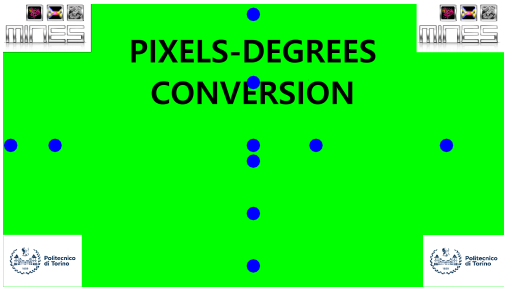

Figure 3.3: Interface n.1 with all marker positions

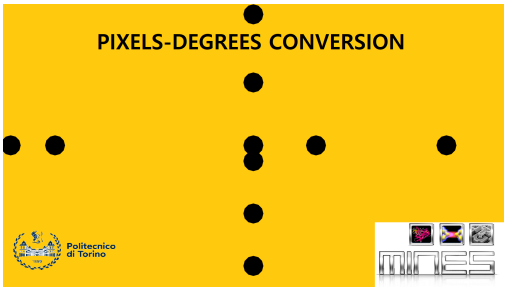

Figure 3.4: Interface n.2 with all marker positions

The participant sits in front of the screen with the interface at a known distance and, wearing Tobii Glasses 3 and without shifting the head, moves his gaze in order to stare at the marker. In Fig. 3.5 the real setup of the experiment, with all the parameters used, is shown. In the picture, the FOV indicated is the one of Tobii Glasses 3, thanks to Tobii Pro Lab the Data Export has a reference system with the origin in the upper left corner of the screen of the PC. This can be done thanks to the snapshot corresponding to the screen of the PC imported in Tobii Pro Lab. Width corresponds to pix_x while High corresponds to pix_Y of the Equations 3.1 and 3.2. The trial record is loaded on Tobii Pro Lab, and Assisted and Manual Mappings are performed on the snapshot in Fig. 3.3 and 3.4, representing the interface. Tobii Pro Lab carries out a mapping of gaze fixation points on the snapshot giving them in a MCS reference system, with the origin in the upper left corner of the snapshot itself. Both types of Mapping were performed in order to understand if eventual errors in the conversion are due to the formulas themselves or are just due to an offset caused by the calibration of Tobii Glasses 3. Another possible problem could be that there are involuntary saccades of the participant while staring at the marker that leads the gaze away from it. During the elaboration of the data coming from Tobii Pro Lab the position of the gaze is found by averaging the gaze points found by the software; if there are several saccades and so several gaze points far from the position of the marker, the average location of the gaze results shifted respect to the marker position.

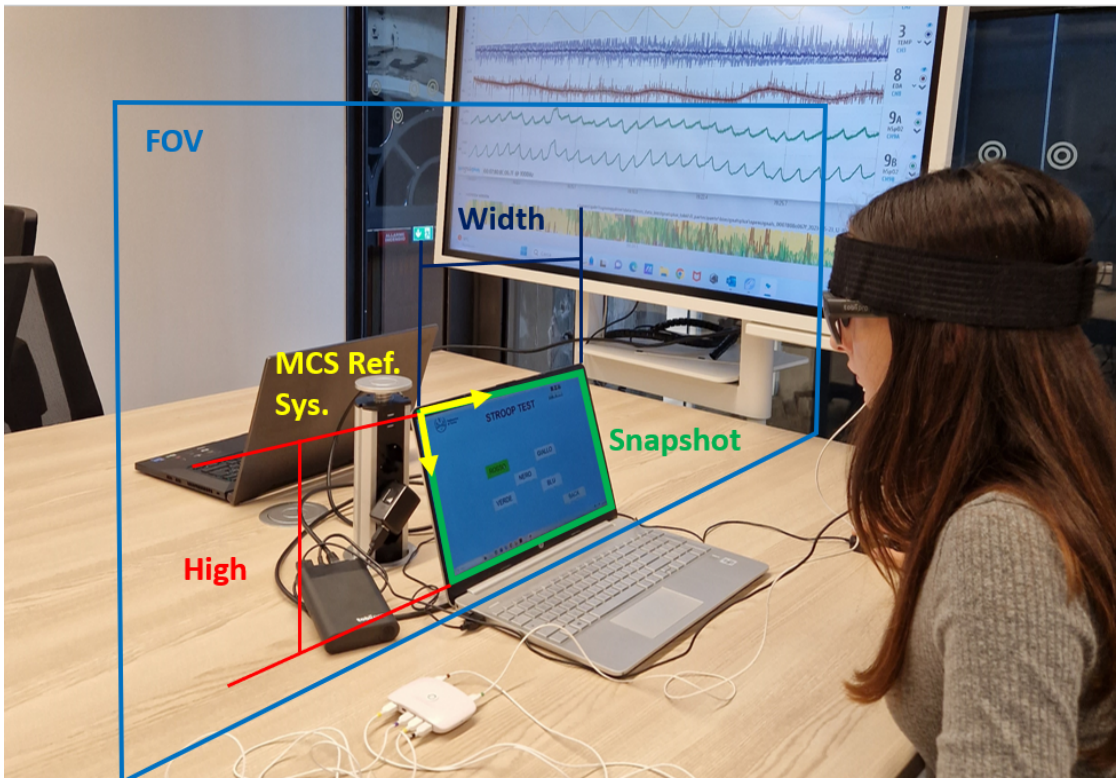

Figure 3.5: Real setup of the experiment

It is also added an event *Event_start* that denotes the marker's appearance in the first location. Then, the mapped gaze points, the events, the time axis, the Width and the Height of the snapshot are exported and loaded in a Matlab script. The flowchart of the validation of the pixels-degrees conversion is shown in Fig. 3.6.

The steps followed are:

1. A Conversion Factor is calculated: the length of the PC screen in mm is divided by the horizontal resolution of the screen, that is 1920 pixels, and it results to be 0.1765 mm/pixel;

2. The measured distance between eyes and screen in mm is converted in pixels through the Conversion Factor;

3. TOBII PRO LAB: The instant where *Event_start* was added is detected; from this moment it is known that every 5 seconds the marker changes position and so the participant's gaze;

4. TOBII PRO LAB: From *Event_start* the Mapped Gaze Points detected every 5

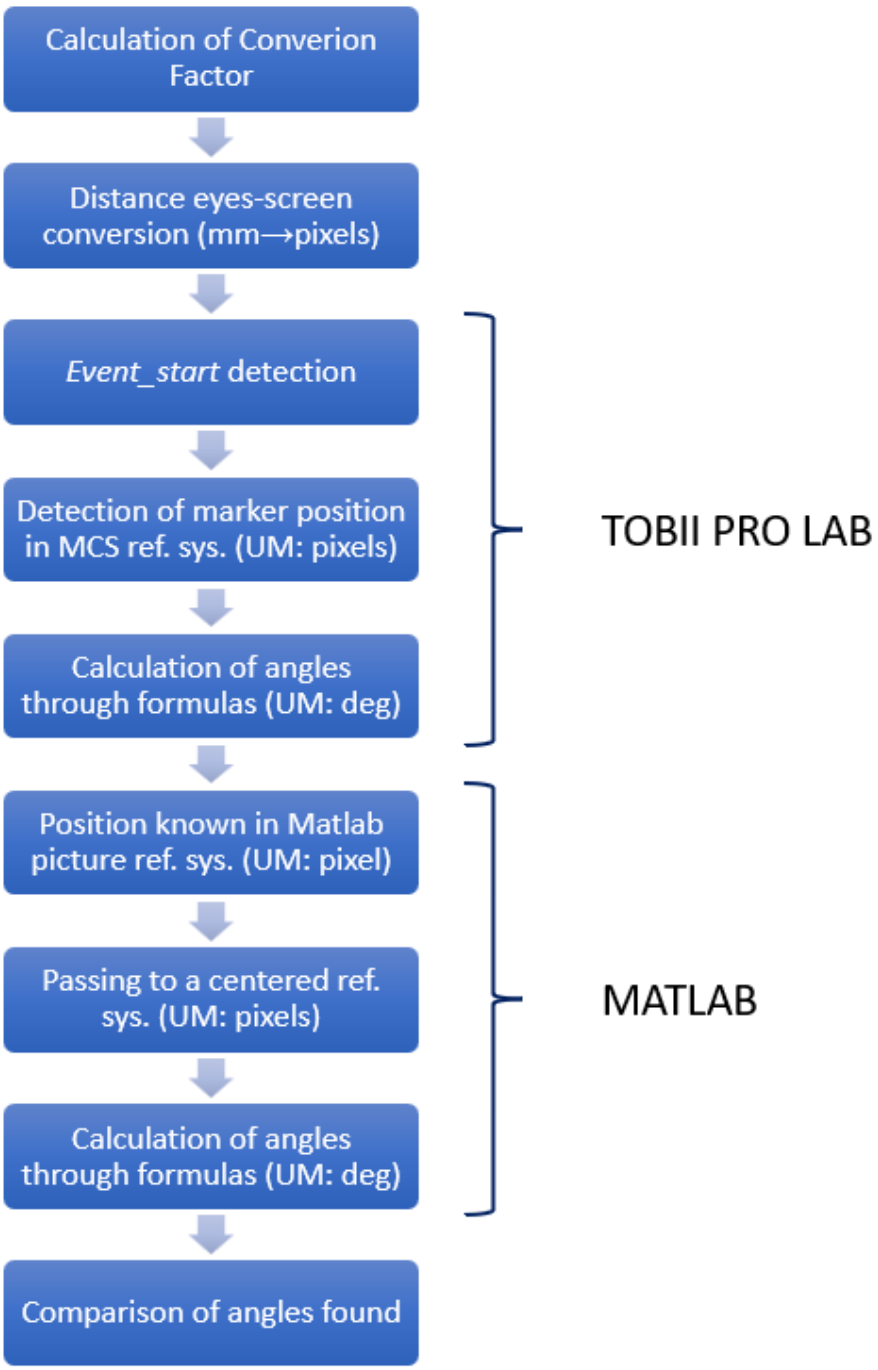

Figure 3.6: Flowchart pixels-degrees conversion validation

seconds are averaged; they represent the position of marker according to Tobii Pro Lab with a pixel measurement unit and referred to the MCS reference system (green snapshot in Fig. 3.5);

5. TOBII PRO LAB: The angles on the vertical ϕ_S and on the horizontal θ_S planes that link the participant to the screen, which refers to the average position of the gaze of the participant while staring at the marker are calculated through the Equations3.1 and 3.2. The picture of reference is Fig. 3.2;

6. MATLAB: While creating the interface, the positions of the marker are set up and are referred to the reference system of the Matlab picture that has the origin in the upper left corner of the image, moreover they are in pixel measurement unit;

7. MATLAB: the known coordinates are expressed in a new reference system centered in the center of the picture as shown in Fig. 3.1;

8. MATLAB: Equations 3.1 and 3.2 are applied to the known positions of the marker, and ϕ_M and θ_M are obtained;

9. The two couples of angles are compared.

3.1.1 Dataset description for validation

The participants are 4, for a total number of registrations of 15. Some of these recordings are done with a distance between the person and the screen of about 50 cm while others with a distance of about 1 meter and a half.

3.2 Algorithm for saccades identification

There are two main categories of algorithms used to classify fixations and saccades.

- Fixation picker method: the fixation candidates and, therefore, the inter-fixation intervals are identified; if the inter-fixation intervals exceed a minimal duration and amplitude, they aren't removed from the list of inter-fixation intervals; if they don't exceed the two fixations at the ends are merged.

- Saccade picker method: the saccade candidates are identified, and the ones that don't satisfy the selection rules based on minimal saccade amplitude and minimal saccade duration are removed. In this way, the inter-saccade intervals are fixation candidates. Finally, with the minimal fixation duration rule, the fixations are identified.

In these two categories, there are different kinds of algorithms, based on different principles, for identifying the two types of eye movements [42].

- *Dispersion-based algorithms*: they belong to the fixation picker category; they are based on the fixations identification through rules about the location and the duration of the fixations itselves. The more used is the Disperion-Threshold Identification algorithm; although it is well established and an integrated part of many commercial software, it is very sensitive to noise and drifts in the data and is poor at providing accurate temporal estimates of event onsets and offsets [43]. Robustness and accuracy are the best between the different groups of algorithms, and speed and implementation ease are high [44].
- *Velocity-based algorithms*: they are based on a velocity threshold that decides if a sample belongs to a fixation or a saccade. They are easy to implement and more transparent than dispersion-based algorithms, but they are more sensitive to noise around the threshold, and even choosing the right threshold isn't easy. They are sometimes combined with acceleration criteria to find saccade onset and offset; the issue rising is that additional filtering is needed since the level of noise increases with the numerical differentiation [43]. The accuracy is good, speed and implementation ease are the best between the different groups of algorithms, while robustness is not as good [44].
- *Area-of-Interest Identification algorithms*: they identify fixations that occur in certain AOIs and use a threshold on the duration of the fixation. The methods' accuracy is not as good while speed and robustness are high; moreover, it is easy to implement [44].

The characteristics of the different methods are summarized in Table 3.1.

Method	**Accuracy**	**Speed**	**Robustness**	**Impl. Ease**
Disperion-Based Alg.	✓✓	✓	✓✓	✓
Velocity-Based Alg.	✓	✓✓	X	✓✓
Area-of-Interest Alg.	X	✓	✓	✓

Table 3.1: Summary of methods for saccades and fixations identification

3.2.1 Implemented algorithm description

The algorithm developed in this study follows Tokuda's one, implemented in 2009 and improved in 2011. It belongs to the saccade-picker group and, in particular, to the velocity-based algorithms.

The algorithm wants to detect and distinguish saccades and fixations. At first, the algorithm wants the division of saccades and fixational eye movements through two thresholds: the first is an amplitude threshold of 1 degree of visual angle in the two-dimensional plane (signal on horizontal and vertical direction). The second one is a time threshold that wants the gaze to stay in the new gaze location and not return to the previous one for at least 1000 ms. Saccades are identified with the double threshold in both signals, the gaze displacement on the horizontal plane and the one on the vertical plane; a logical OR between them is performed. Since saccades are identified, inter-saccade intervals are fixations [25]. In Fig. 3.7 the saccades identification made by the paper is shown. In particular, the signal of the horizontal eye position of a subject, from 20 to 32 seconds, is shown and fixations and saccades are pointed.

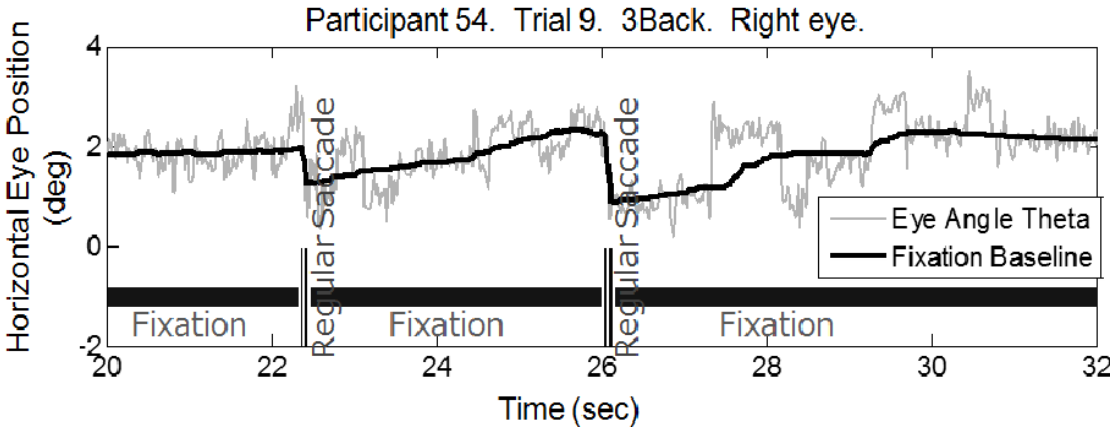

Figure 3.7: Tokuda algorithm

Drift and tremor removal are performed through a double moving median procedure: the first moving median has a window length of 5 samples, so 100 ms; it is used to remove tremor from the signal. Then, a second moving median with a window length of 100 samples, so 2 seconds, is carried out to calculate the drift. The calculated drifts are removed from the tremor-free line [29].

3.3 Tobii Pro Lab Use

Each participant has been imported in Tobii Pro Lab, the events related to the start and the end of each phase are placed in order to create TOIs for every phase of the trial. TOIs are shown in Fig. 2.17 and 2.13. Then the data are exported in TSV files in the form of Data Export with, among the fields, time axis, gaze position on x and y direction in an MCS coordinate system and eye movements assigned to each sample by Tobii Pro Lab (Fixation, Saccade, ENF, and Unclassified). Finally,

metrics calculated by Tobii Pro Lab and related to each phase are exported in TSV files. Metrics TSV files and Data Export TSV files are created from the interfaces in Fig. 2.14 and 2.15.

3.4 Eye movements and gaze points

In this section it is explained the distinction between gaze point and eye movement. The gaze point is a single sample acquired every 20 ms by Tobii Glasses 3, it is classified as saccade, fixation or SI point. The eye movement is a set of consequent gaze points classified in the same way. In Fig. 3.8 this difference is shown: the single points are gaze points while the lines represent the eye movements.

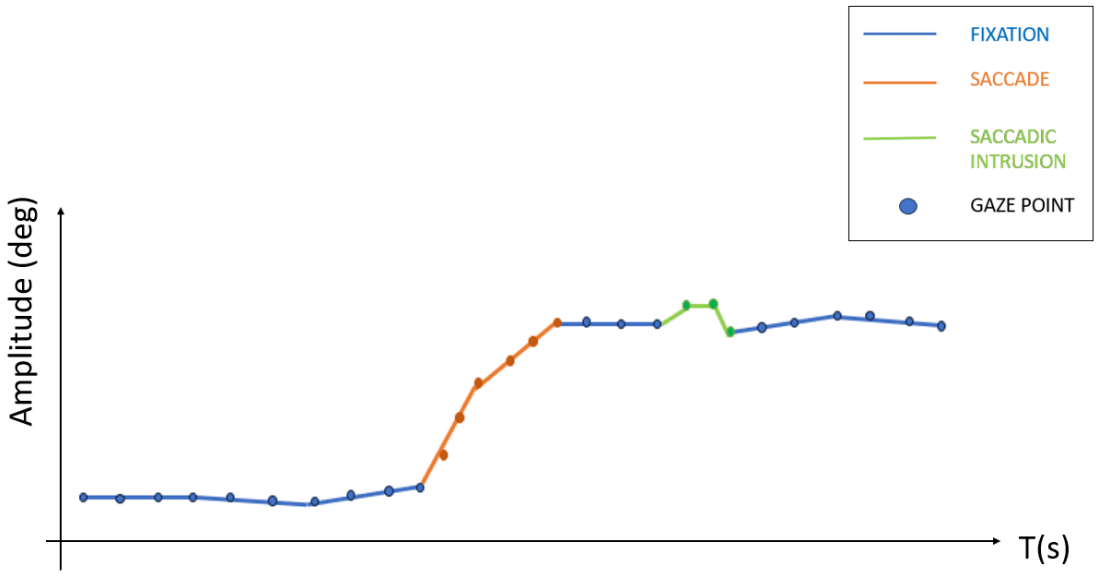

Figure 3.8: Eye movements and gaze points

3.5 Preprocessing

In this project, the "main" code follows the sequent steps for each participant:

1. A table with data coming from Raw Data extracted by Tobii Glasses 3 is imported, and the fields with time axis and gaze position on x and y directions in an MCS coordinate system are saved in vectors. In the table, where no signal was detected by Tobii Glasses 3, the gaze location is assigned at zero pixels. In Fig 3.9 a set of raw data imported from Tobii Glasses 3 are shown; it could be seen that there are horizontal lines due to a time axis with, sometimes,

a difference between samples smaller than 20 ms. Moreover, in the initial part, a more regular signal could be noticed: that's why the participant was performing the Stroop Test so the moving of the eyes is mainly limited to the screen. Later, the signal is more irregular: that's why the participant is doing a phase of rest, so she or he can move the gaze all around.

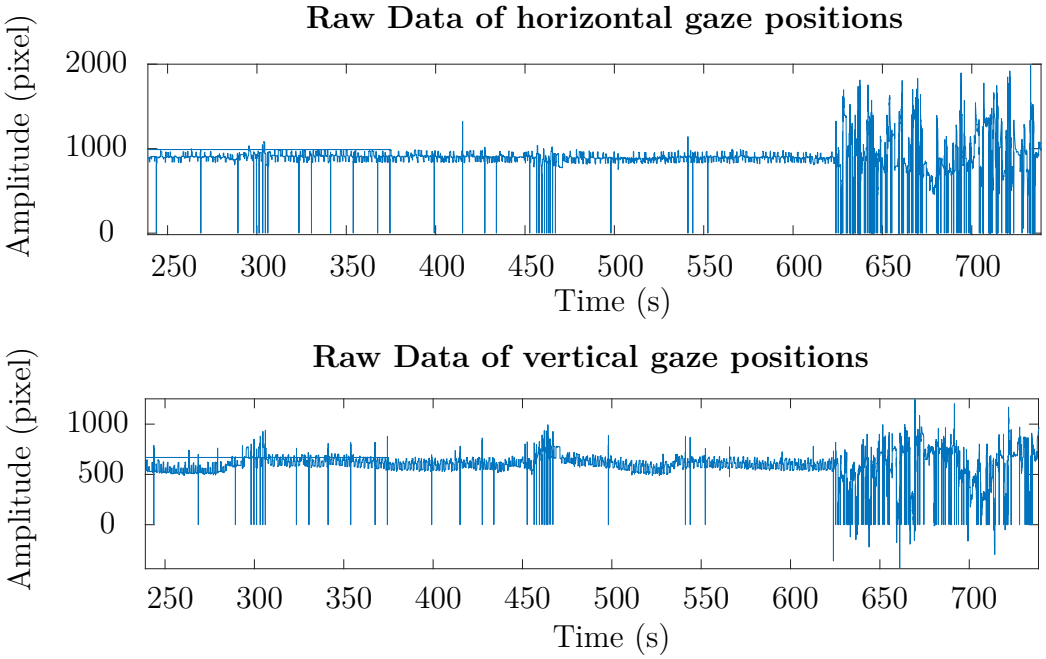

Figure 3.9: Raw data imported from Tobii Glasses 3

2. Data Export, whose obtaining is explained in section 3.3, coming from the elaboration of Raw Data by Tobii Pro Lab are imported, and the fields with the time axis, gaze position on x and y direction in an MCS coordinate system, eye movements assigned to each sample by Tobii Pro Lab (Fixation, Saccade, ENF, and Unclassified) and the events created in Tobii Pro Lab are saved in vectors. The time axis extracted by Raw Data and the one from Data Export are not the same because some samples are missing in the Data Export time axis.

3. Time instants and samples related to the events created by the user in Tobii Pro Lab for identifying tests beginning and ending are identified. This step is carried out by looking in the vector containing the events for the ones of interest and, therefore, for their corresponding indexes, that are the same on the time axis. Knowing the indexes, the timestamp related to each event can be saved.

4. The time axis extracted by raw data, as already introduced, has missing samples, so the difference between the timestamp of two consecutive samples is not constant to 20 ms. A time axis "filling" is carried out: the difference between each consecutive timestamp is calculated; if it is lower than 20 ms, the timestamp is corrected with the average of the two extreme timestamps. If it is bigger, new samples are created corresponding to the instants missing in the original one, each one separated from the previous one by 20 ms. The location of the gaze in the x and y directions is calculated by looking at their values at the extremes of the missing data and assigning them a value with a linear interpolation.

5. Since that the time axis has been "filled", the indexes of the events found before aren't the right ones anymore, so they are calculated again on the new time axis.

6. Where the gaze signal is zero, there should be missing samples, as explained in point 1, because Tobii Glasses 3 couldn't detect the gaze position. This means these samples correspond to the ENFs classified by Tobii Pro Lab. A vector with the indexes of ENFs is created, looking to the zeros in the signal, and the vectors with the position of the gaze on x and y directions, in these indexes, are changed from zeros to a missing value (NaN).

7. Cause the data of gaze signal are in a pixel unity measure in an MCS coordinate system, but the algorithm of Tokuda et al. [25] works on data in a degree unity measure in a centered coordinate system; a conversion from pixels to degrees, and between the two coordinate system, is carried out. It implements the formulas explained in the section 3.1. In Fig 3.10 a piece of signal is shown: in particular, ENFs are shown, in the picture above they are zeros, while below they are converted to NaNs.

   In Fig. 3.11 horizontal and vertical signals after "time axis filling", changing of ENFs as NaNs and pixels-degrees conversion are shown.

8. Saccade detection is carried out following the Tokuda algorithm already explained in Section 3.7. Samples classified as saccade points on both signals coming from the location of the gaze in the x and y direction are detected. Then a logic OR between them is performed. A vector with all the indexes of saccade samples is created. Close saccade points represent together, a unique saccadic eye movement, as shown in Fig. 3.8. In Fig. 3.12 Tokuda algorithm applied on the horizontal signal is shown. In particular, it could be noticed that sample i-1 and i have a difference in amplitude bigger than 1° and the signal doesn't come back to the i-1 location before 1000 ms.

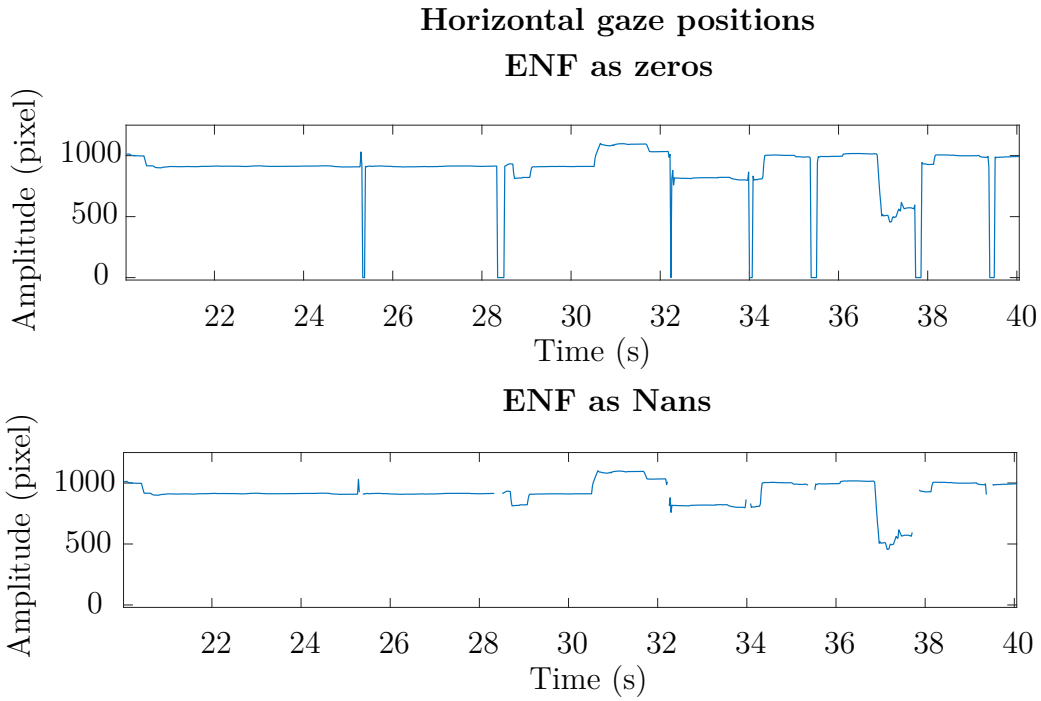

Figure 3.10: Eyes Not Found

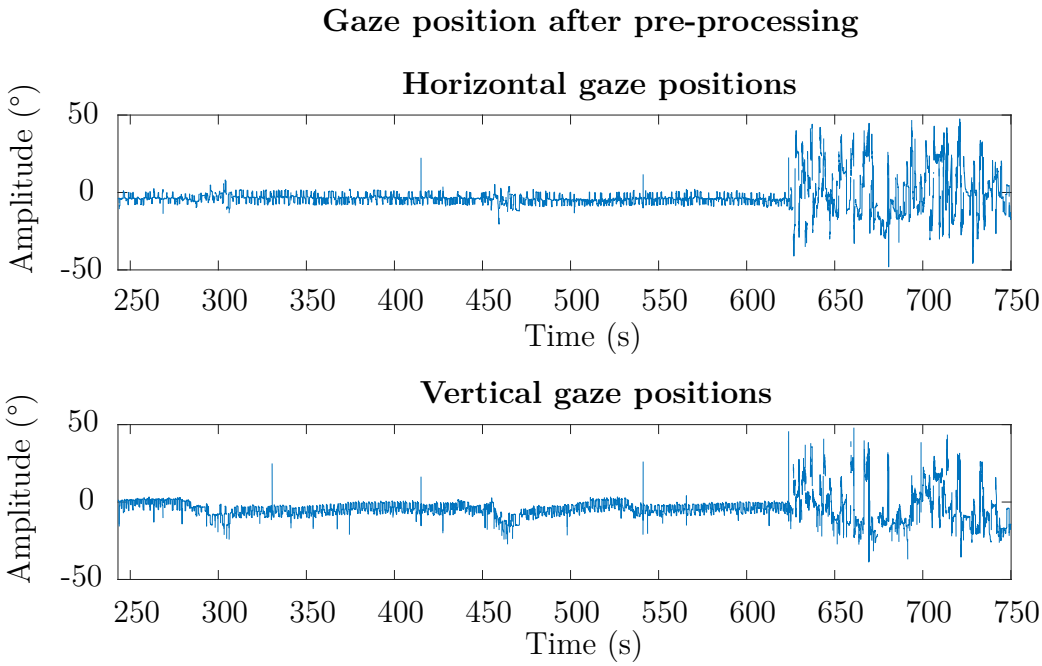

Figure 3.11: Horizontal and vertical signals after "time axis filling" and pixels-degrees conversion

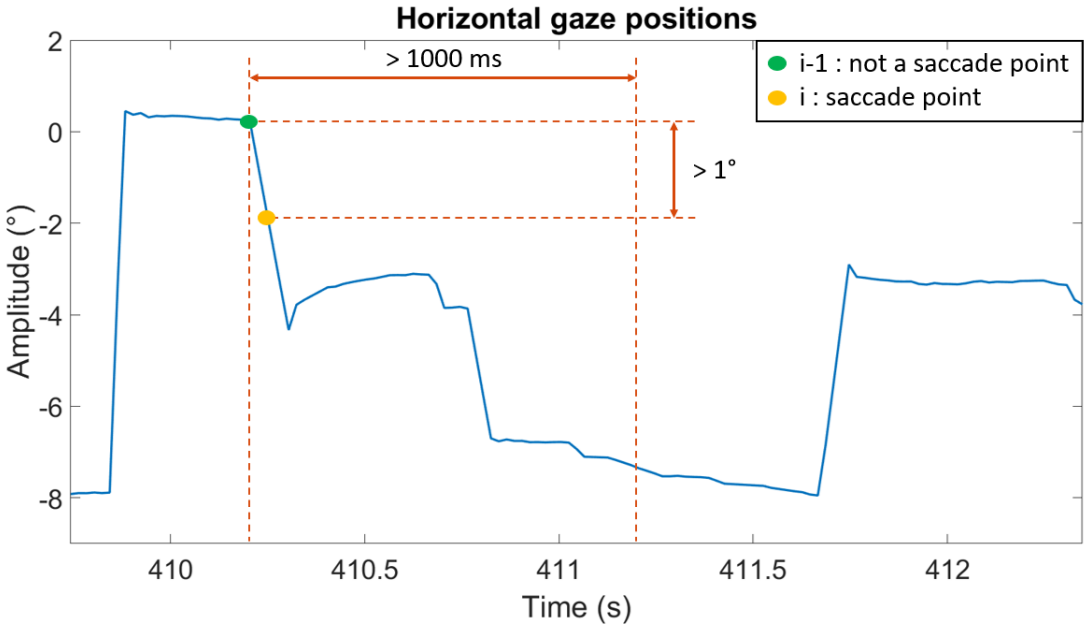

Figure 3.12: Tokuda algorithm applied to the horizontal signal

9. Fixation candidates points are identified as the samples that are not saccades and are not ENFs. For finding the real fixations, defined as sets of fixation points, some selection rules are applied. As suggested by the Tobii Pro Lab manual, a set of ENFs lasting less than 50 ms are not blinking but are due to eyelash interference or other kind of movements. So, another vector is created with just the indexes of the ENFs that probably are blinking. If before and after a short set of ENFs, there are fixations with a difference in amplitude between them smaller than a threshold of 1 degree, this set of ENFs becomes part of a unique bigger fixation which includes the previous and the following fixations. The last selection rule is related to short fixations that are excluded: the threshold is 60 ms. This threshold and the saccade threshold amplitude are the selection rules identified by Hooge et al. [42] in 2022 for finding a similar fixation duration distribution, whichever is the algorithm applied. A vector with all the indexes of the fixation points is created. Close fixation points represent together, a unique fixation eye movement, as shown in Fig. 3.8.

10. Saccadic intrusions are identified following the steps described by Tokuda et al. [25]: once saccades and fixations are identified, just the indexes related to fixation points are considered since saccadic intrusions are fixational eye movements. Moreover, just the signal of gaze movements in a horizontal direction is considered because SIs are just on this axis. At first, the signal is

interpolated in order to avoid the loss of samples once the filters are applied. The baseline is calculated through the application of two median moving filters: the first moving median has a window length of 5 samples, so 100 ms, is used to remove tremor from the signal. Then, a second moving median with a window length of 100 samples, so 2 seconds, is carried out to calculate the drift. The calculated drifts are removed from the tremor-free line. Then, just the samples where the absolute value of the gaze deviation is bigger than 0.4 degrees are saved as possible SIs. The second criteria is related to the dwell time, which should be between 60 and 870 ms, so the gaze shifts from one location to another and comes back within that dwell time [29]. A vector with all the indexes of the SIs points is created. Close SI points represent together, a unique SI eye movement, as shown in Fig. 3.8.

11. Duration phases are calculated and the closest indexes to event indexes, in gaze signal, in each vector containing the indexes of ENFs, fixations, saccades, and SIs, are found. In Fig. 3.13 horizontal and vertical signals with division in tests phases are shown.

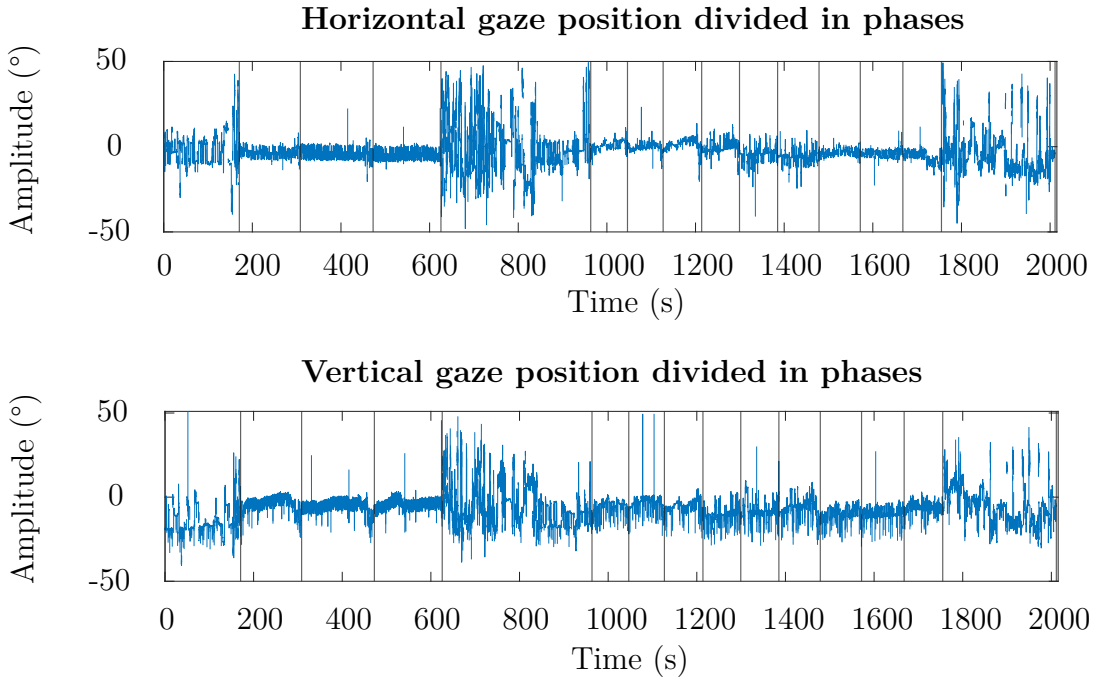

Figure 3.13: Horizontal and vertical signals with division in tests phases

In Fig.3.14 the flowchart of the pre-processing of the signal is shown. The next step will be the feature extraction through the processing of the signal.

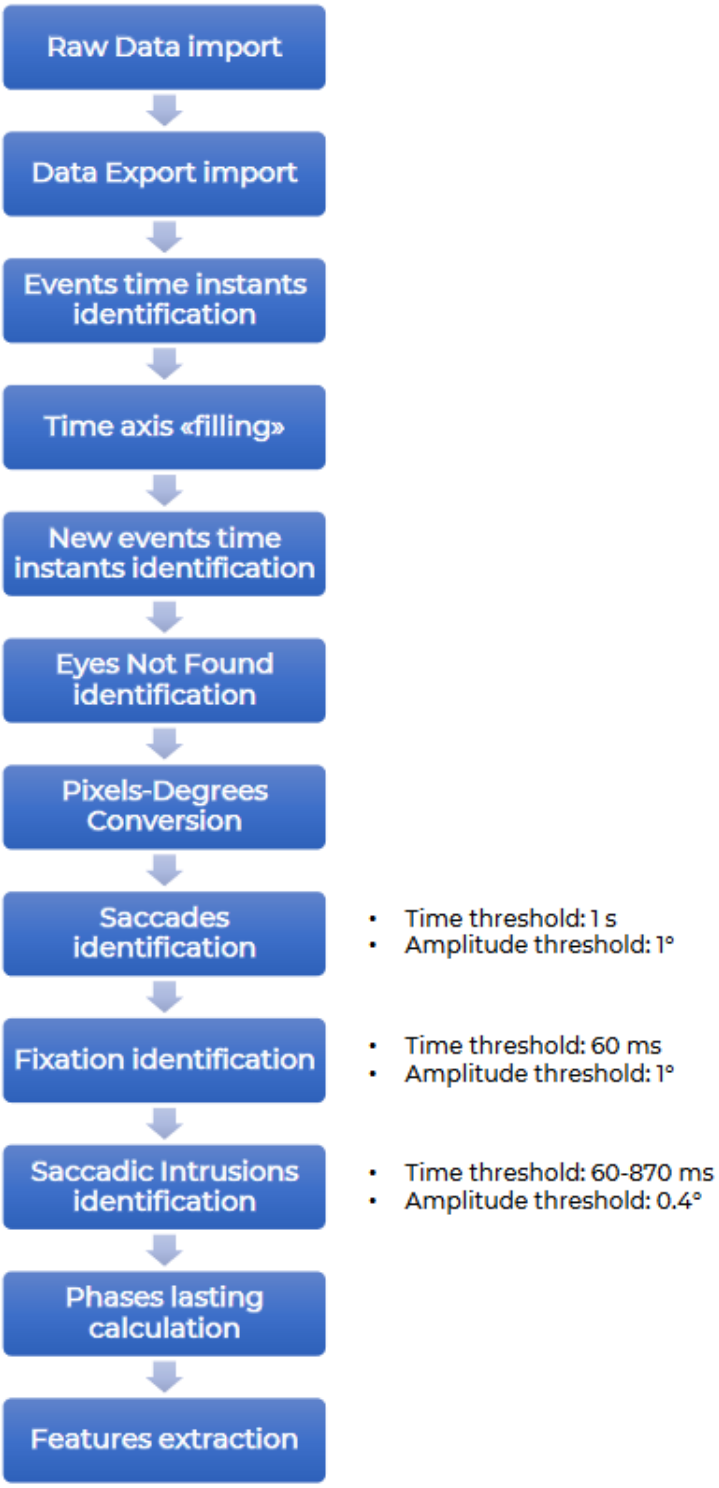

Figure 3.14: Preprocessing flowchart

3.6 Processing

In this section, it is described the processing of the signal for the feature extraction. In Fig.3.15 the flowchart used for describing the features extracted is shown.

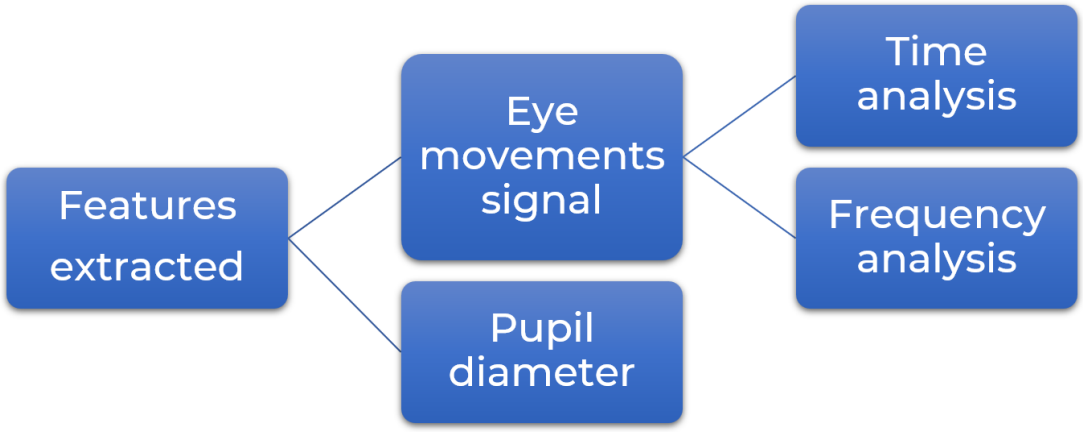

Figure 3.15: Processing flowchart

The calculated features can be divided mainly into two groups: those based on the pupil diameter and those calculated based on the signal from the eye movements. The first are extracted from the signal concerning the pupillary diameter provided directly by the device, while the latter are based on eye movements previously identified in the preprocessing of the signal provided by Tobii Glasses 3, and therefore are related to fixations, saccades, and SI.

Speaking about the first category of features, they are related to the relative pupil diameter of the left and right eyes. The absolute pupil diameter signal can't be used because it is very sensitive to changes in the environment's illumination. The lighting of the environment is complex to keep constant; that is the reason why a relative measurement with respect to a natural condition is preferred. The diameter of the eyes in the natural condition is found as the average diameter of the pupil in the first relaxation phase. The relative diameter is calculated with the formula 3.3 for right and left eyes:

$$RelativeDiameter = \frac{P_i - P_0}{P_0} \tag{3.3}$$

where P_i is the pupil diameter measured during the experiment, and P_0 is the pupil diameter in natural conditions, so during relaxation. P_0 is obtained through the mean of the pupil diameter during the first relaxation phase.

Then, the second group of features extracted can be divided into metrics obtained through a time analysis (Time-Based TB) of the signal and those obtained through

a frequency analysis (Frequency-Based FB) of the signal. Considering the TB group, for each type of eye movement, some features extracted have already been analyzed in the literature, while others are investigated in this study for the first time. In the FB group, features extracted from the power spectral density (PSD) and related to stress and MWL could be found; moreover, they have not yet been analyzed in the literature.

The description of the TB group follows:

- BLINKING: Duration, frequency, and interval blinking are found.
- FIXATIONS: Duration and frequency saccades are calculated. Fixation lasting is shown in Fig. 3.16.
- SACCADES: Duration, frequency and velocity of saccades in x and y direction are calculated. Saccade lasting is shown in Fig. 3.16.
- SI: SI value is calculated as the absolute values of the SI measures accumulated over the evaluation phase divided by the lasting of the phase. Duration and frequency of SIs are calculated and the mean of single SI velocities and the mean of single SI velocities throughout the phase are obtained. For these two features the velocity is calculated as the maximum displacement with respect to the baseline of a single set of SI gaze points, in absolute value, divided by the lasting of the single SI. SI velocity is calculated as the mean of the difference in amplitude of the gaze position of the first index of a set of SI gaze samples and the index before, if it is not an ENF, over the lasting of that set of SI gaze points.

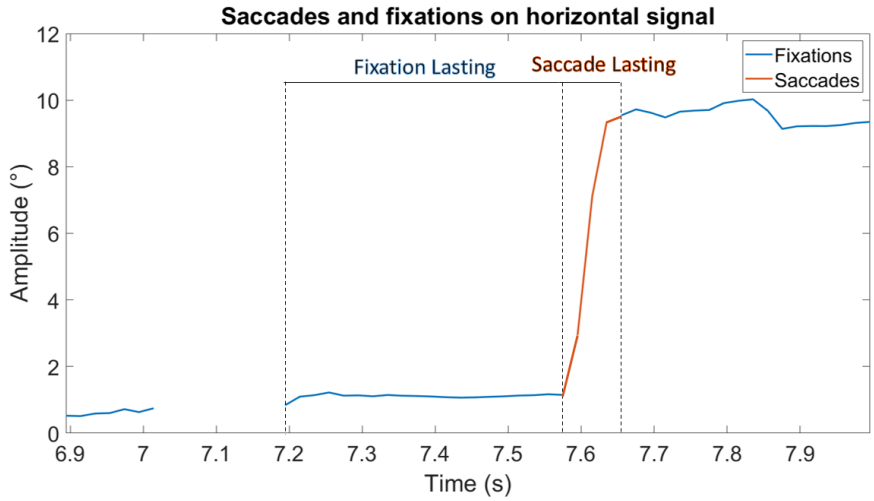

Figure 3.16: Fixation and saccade lasting

The FB group treats features coming from analysis in frequency observation. In particular, for both signals in the x and y direction, are found the power provided by the low-frequency (LF) bandwidth (0-1 Hz) and by the high-frequency (HF) bandwidth (1-3 Hz), as the area underlying the normalized PSD in the right range of frequency. Moreover, the ratio between these two powers is calculated.

3.6.1 Frequency Analysis

In order to achieve features coming from frequency analysis, several steps have been performed:

1. Steps from 1 to 7 of preprocessing are carried out.

2. The signal is interpolated in order not to have the contribution of ENFs and to preserve the original time axis. In Fig. 3.17, the PSD of the interpolated raw signal of one participant is shown. The PSDs of other participants are similar to the one shown, and the peak of the PSD is close to zero, and it ranges from 0.02 to 0.05 Hz. The peak for this participant is better shown in the graph below in Fig. 3.17 thanks to a zoom of it, in this particular case the peak is at 0.025 Hz.

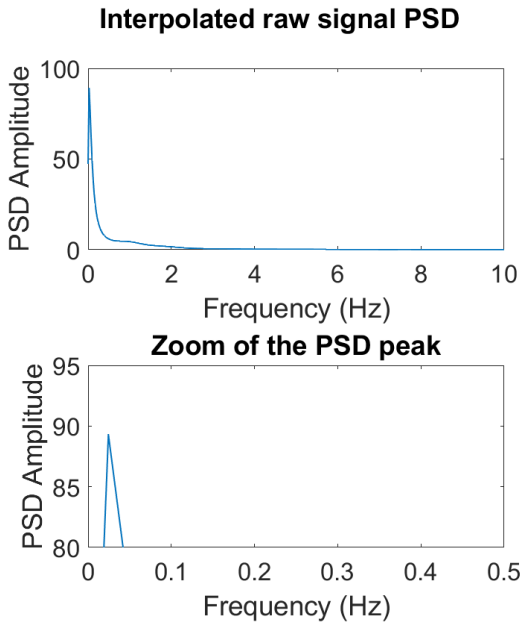

Figure 3.17: PSD of interpolated raw signal

3. Filtering: according to literature, there are two main kinds of noise in ocular signal [27]: the drift has a bandwidth from 0 to 0.5 Hz, and, as shown in Fig. 3.17 it is the noise that provides more power to the signal. The second kind of noise is the tremor; it has a bandwidth of 30-100 Hz, and Fig. 3.17 shows that his contribution is really low because the signal in that range of frequencies provides no power.

    - Tremor filtering: The first operation carried out is the tremor removal; it is tried to remove it with two methods. The first one is a moving median filtering with a moving window of five samples, so 100 ms [29]. While, the second one is a Butterworth filter of sixth order with the mask shown in Fig. 3.18 and with a cutting frequency of 24 Hz, the maximum possible with this kind of filter (sampling frequency/2-1). In Fig. 3.19 they are shown: in the picture above the PSDs of the raw interpolated signal, the filtered signal with moving median method and the filtered signal with Butterworth filter method. In the same figure, but in the picture below, the three signals are displayed. As shown the three PSDs are very similar, in particular the PSD of raw interpolated signal and of Butterworth signal are practically overlapped, this is confirmed by the signals displayed in the second picture. The moving median filter cuts more the high frequency respect to the Butterworth filter.

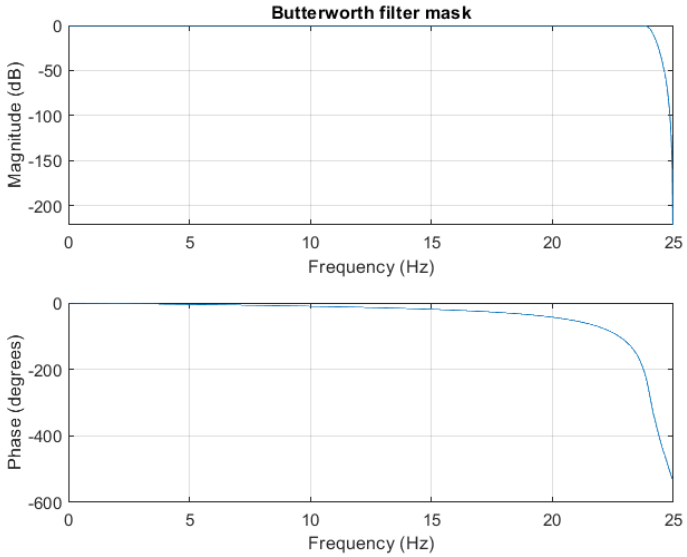

Figure 3.18: Mask of Butterworth filter for tremor removal

It is chosen to go on with the signal filtered with the moving median because it smooths more the raw signal and, moreover, it is the strategy

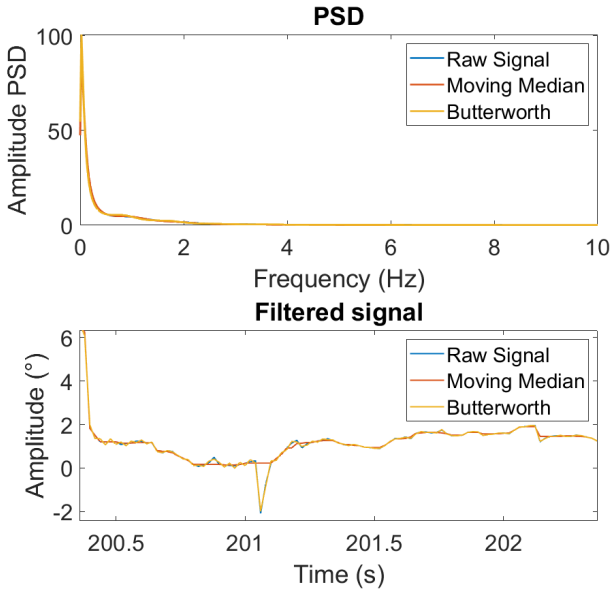

Figure 3.19: Raw signal, filtered signal with moving median and filtered signal with Butterworth filter: PSD and signal

used by Tokuda et al. [25].

- Drift filtering: This filtering of the signal is performed with two methods.

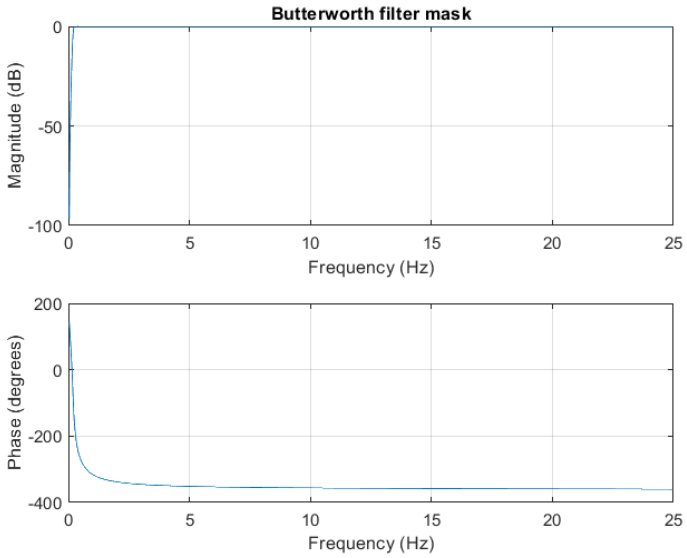

Figure 3.20: Mask of Butterworth filter for drift removal

The first one wants to calculate the drift with a moving median filter [25] with a window length of 100 samples and so of 2 seconds, and later to subtract the drift found to the signal. The second one is a highpass Butterworth filter with a cutting frequency of 0.5 Hz and sixth order, that wants to remove directly the continuous component of the signal. In Fig. 3.20 the mask of the Butterworth filter is shown. In Fig. 3.21 they are shown: in the picture above the PSDs of the signal at which the tremor has been already removed, the filtered signal with moving median method and the filtered signal with the Butterworth filter method. In the same figure, but in the picture below, the three signals are displayed. It could be noticed that with both methods the main original peak in 0.02-0.05 Hz has been removed; in the second picture it could be seen that the drift is effectively removed and the signal is shifted to a baseline of 0°. It is decided to use the signal filtered with the Butterworth filter.

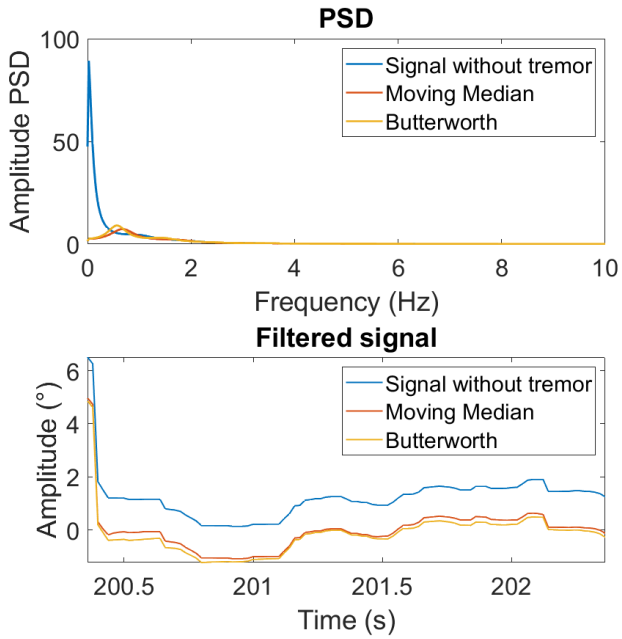

Figure 3.21: Raw signal, filtered signal with moving median and filtered signal with Butterworth filter: PSD and signal

4. For data filtered, for a better understanding of the location of the peaks, PSDs in the first phase of relax, in the Stroop phase, in the second phase of relax, and in the N-Back phases are calculated. It can be noticed in Fig.3.22 that the main peak in all phases is around 0.5 Hz, while the second main peak is

approximately 1.2-1.3 Hz. The figure is related again to the same participant whose pictures are shown before, but this trend is similar in all participants.

This is the reason why it is decided to distinguish the LF bandwidth from the HF bandwidth with the delimiter frequency at 1 Hz. In particular, LF bandwidth ranges from 0 to 1 Hz while HF bandwidth ranges from 1 to 3 Hz; in fact, after 3 Hz the contribution of PSD is really low. The two frequency limits are shown in the figure with two black vertical lines. Moreover, it could be noticed that in both phases of relaxation, there is a higher contribution of LF bandwidth than in the Stroop and N-Back phases.

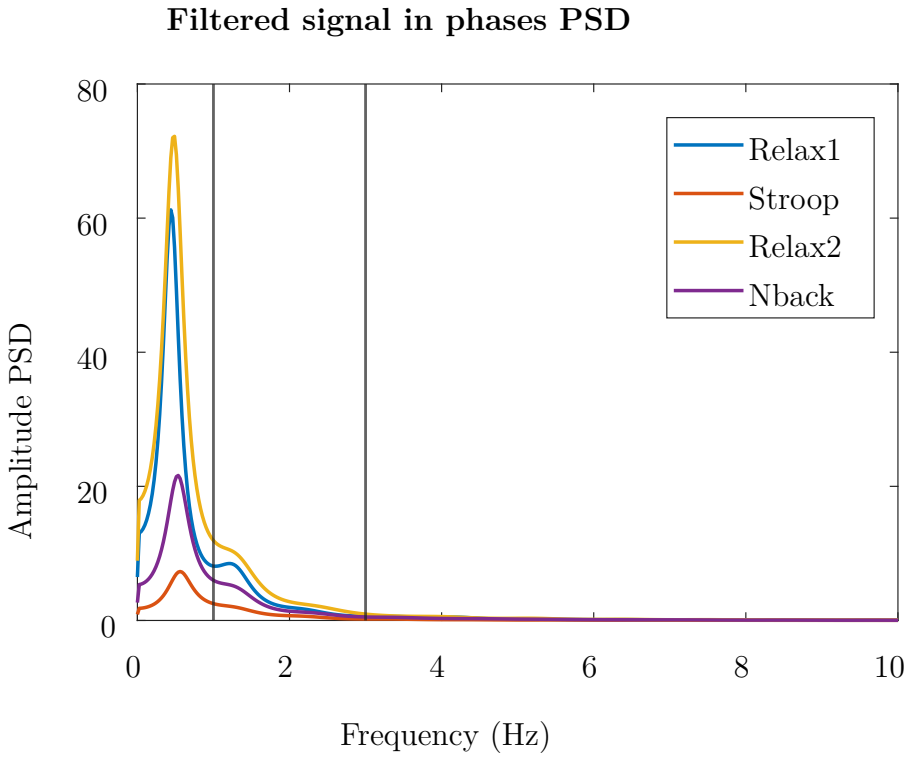

Figure 3.22: PSD of filtered signal in each macro-phase

5. In order to calculate the feature of power from LF and HF bandwidth and their ratio, the PSD of the filtered signal in all phases has been calculated and normalized with respect to each maximum value of PSD. Later, the power of the signal in LF and HF bandwidth was calculated as the area below the signal in the right range of frequencies. Finally, the ratio between the power in the LF bandwidth and the power provided by the HF bandwidth is obtained.

Each feature is resumed in a matrix containing the values related to it for all

the participants. Each participant is represented on each row, while the columns represent the phases of all the tests. The resume of all the features extracted is shown in table 3.2.

Table 3.2: Eye movement features and trends, if known in literature, are shown.

Eye Movement	**Feature**	**Expected Trend**
Fixations	Lasting	↑
	Frequency	↑
Saccades	Lasting	-
	Frequency	↑
	X velocity	↑
	Y velocity	↑
Blinkings	Lasting	↓
	Frequency	↑↓
	Interval	↓
SIs	Lasting	-
	Frequency	-
	SI value	↑
	SI velocity	-
	Single SI velocity	-
	Single SI velocity over phase	-
Pupil	Relative diameter Left	↑
	Relative diameter Right	↑
PSD	Low frequency power X	-
	High frequency power X	-
	Ratio X	-
	Low frequency power Y	-
	High frequency power Y	-
	Ratio Y	-

3.7 Dataset description

The data set gathered during the trial comes from 64 participants; one participant was removed because of missing data from Tobii Glasses 3. Without the participant removed the percentage of females is about 49 while the percentage of males is about 51; this means that the investigated population is well distributed regarding gender. Looking at the age of the participants goes from 19 to 40 years old, with an average of 23.5 years old and a standard deviation of 3 years. This leads to the conclusion that the investigated population isn't representing all the age groups.

Chapter 4

Results

This chapter is structured in the following way: at first, the results of pixels-degrees conversion validation are shown. Then, it is explained the assessment of stress and MWL in results and the steps for choosing the best normalization possible. Later, the dataset investigation is performed and the comparison between the gold standard and the developed algorithm is shown. Finally, the statistical analysis of the results is carried out.

4.1 Validation of Pixels-Degrees Conversion

For the pixels-degrees conversion validation, already treated in section 3.1 the snapshots imported in Tobii Pro Lab and used for the Manual Mapping are in Fig. 3.3 and 3.4, while for the Assisted Mapping, the snapshots are the same as before but without drawn all the positions of the marker. The snapshots must be easily recognizable by Tobii Pro Lab Assisted Mapping, which is why some logos and writings are present. At first, for both the Assisted and the Manual Mapping, a plot is performed with the ideal and calculated angles ϕ and θ. They are, respectively, the angles on the vertical and on the horizontal planes that link the participant to the screen. In Fig. 4.1, it is shown the comparison of ideal and calculated angles with both, Assisted and Manual Mapping of one recording. For further graphs about the other recordings go to Section A. In order to resume the results for all the 15 recordings the mean error and the standard deviation of the gaze location corresponding to each marker position with Assisted and Manual Mapping is calculated and shown in Fig. 4.2. The results will be later commented in section 4.1.1.

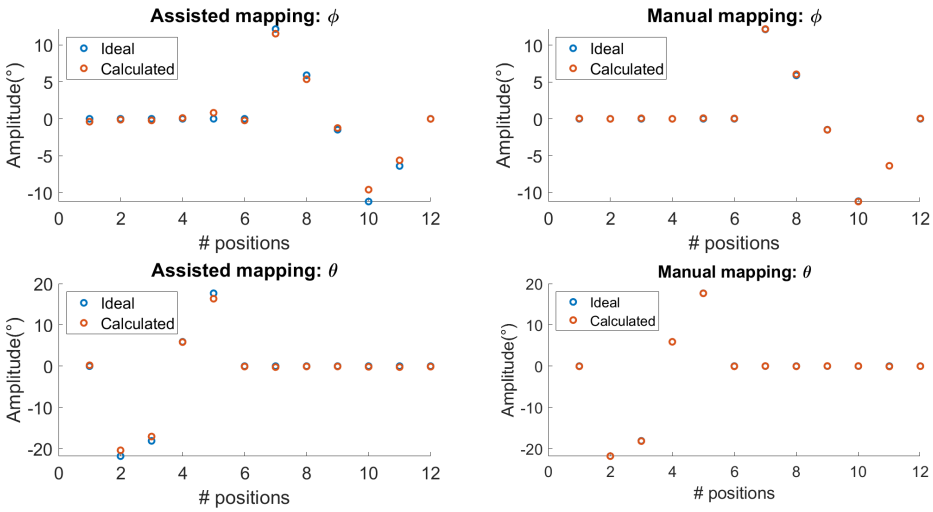

Figure 4.1: Comparison of ideal and calculated angles of one recording of example

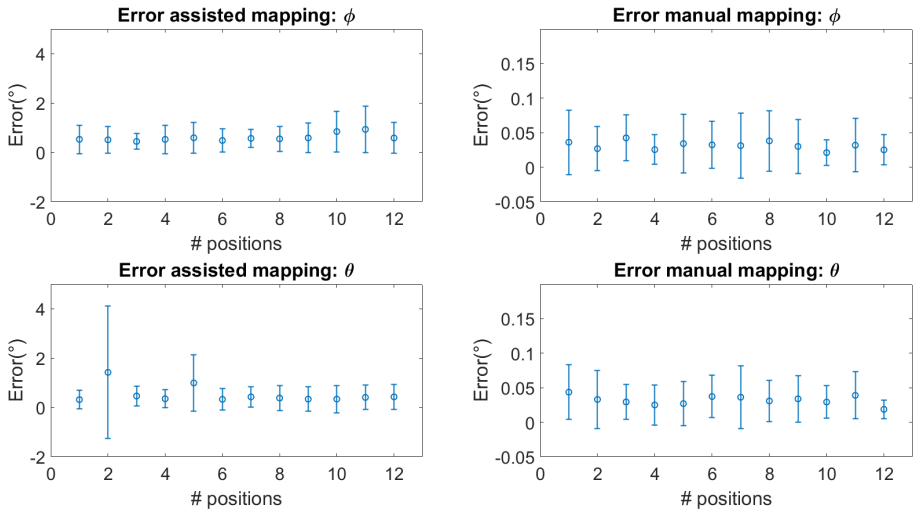

Figure 4.2: Mean error and standard deviation of the gaze location corresponding to each position of the marker with Assisted and Manual Mapping

4.1.1 Discussion of validation of pixels-degrees conversion

As shown in Fig. 4.1, both types of mapping give angles that are close to the ideal ones; this leads to the conclusion that the **formulas for the conversion** of gaze location from pixel to degrees are **consistent** in this use case. Another observation is that for most of all plots, the Assisted Mapping gives an error visibly higher than the Manual Mapping and usually with a constant bias. This observation is strengthened by Fig. 4.2, which shows a mean error an order of magnitude lower for the Manual Mapping with respect to the Assisted Mapping. The higher error made by Assisted Mapping could be due to the initial not perfect calibration of Tobii Glasses 3 that causes the shifting of the gaze in the whole recording with respect to the real position of the gaze itself. In the code created for this contest, every 5 seconds the position of the marker changes; so later, while elaborating the data, the positions of the gaze recorded by Tobii Glasses 3 in these 5 seconds are averaged. If the participant involuntary moves his gaze while staring at the marker, even this location will be averaged, causing a shifting of the calculated average position. This could be another issue that leads to the worst performance of the Assisted Mapping.

4.2 Stress and MWL assessment in results

The questionnaire is structured in a way that for each phase of the tests, the participant can give a score from 1 to 3 for both, the stress and MWL levels felt. Anyway, it is chosen to consider the answers of the stress just for the Stroop Test and the ones of the MWL just for the N-Back Tests. This choice is made because the participants could find it difficult to differentiate between the two concepts of stress and MWL, and they could be confused if it is asked to give a score about the stress felt on a test designed just for creating different levels of MWL and vice versa.

4.3 Steps for choosing the best normalization method

In section 2.9 they are described the four kinds of normalization applied to the datasets. The steps followed for assessing the best kind of normalization for reducing error bars and having the most meaningful representation of the data, are the following ones. For each feature:

1. The matrix containing all the values is imported;

2. The matrix is divided in the four datasets, one for each test (Stroop, Visual N-Back, Auditory N-Back, and Dual N-Back). Each row corresponds to a participant and in the first column, of each dataset there are the values of the feature in the initial relax, while in the last three columns, there are the values of the feature in the three levels of each test. For each dataset:

   - The four kinds of normalizations are applied on each row, so for each participant.
   - Four different matrices are created, one for each normalization. For each matrix:
     - The matrix is divided into classes.
     - The standard deviation is calculated on each class.
     - In a vector, created for a specific kind of normalization, the four values are added.

This process is carried out for all the 23 features. For each dataset, the resulting four vectors, one for each kind of normalization, are 23*4 long. Four histograms for each test are represented, one for each kind of normalization.

The normalization that permits the smallest error bar is the one that has a distribution of the standard deviation closer to zero. This means that the majority of standard deviations are small, and so the error bars are shorter.

4.3.1 Evaluation of different kinds of normalization

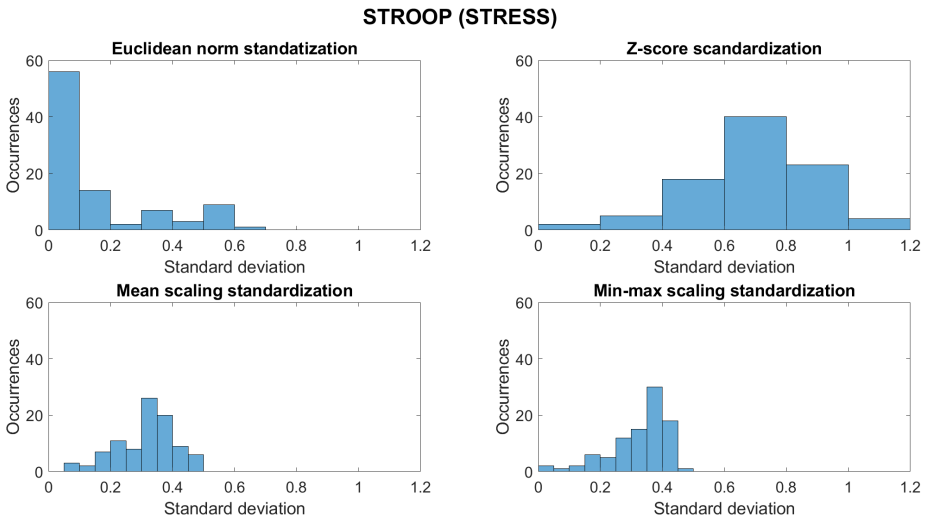

Figure 4.3: Stroop Test: histogram of standard deviations

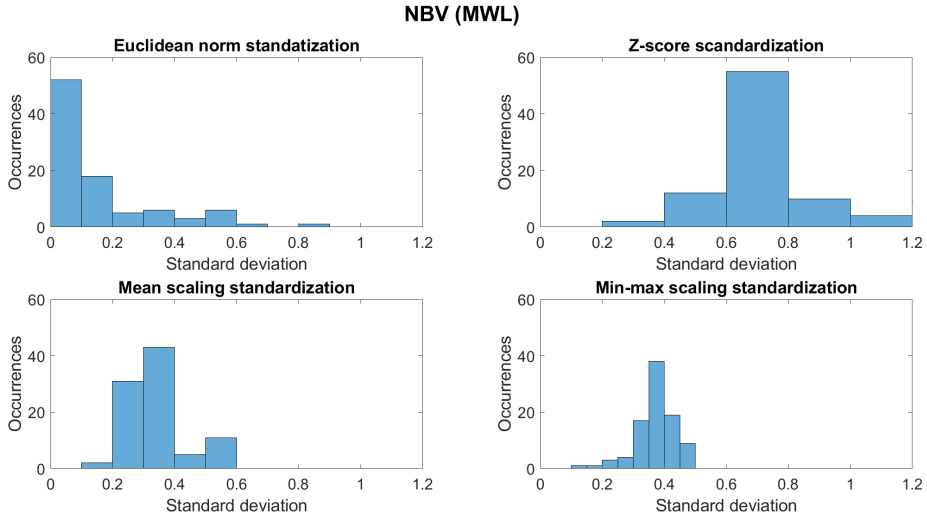

Figure 4.4: Visual N-Back Test: histogram of standard deviations

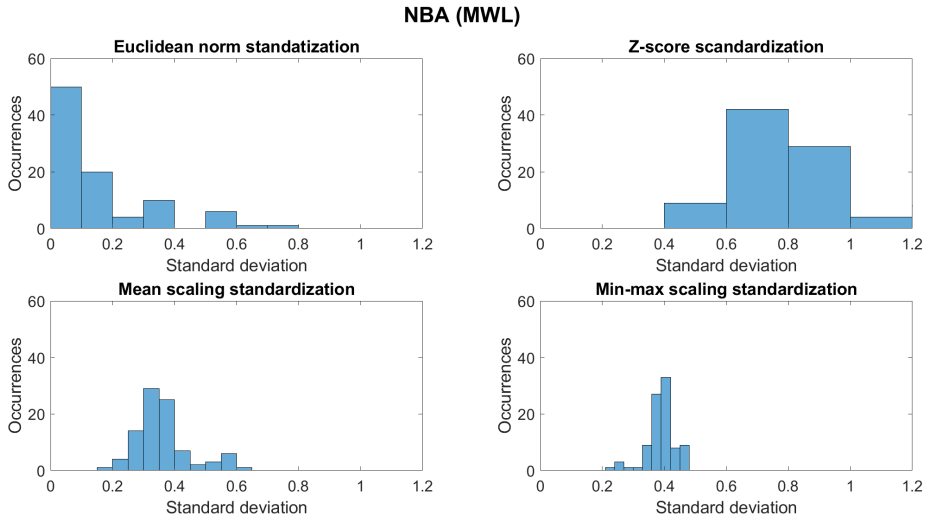

Figure 4.5: Auditory N-Back Test: histogram of standard deviations

Looking at the histograms shown in the Fig. 4.3, 4.4, 4.5, 4.6, for each test, could be concluded that:

- The normalization through *Euclidean norm* gives values of standard deviations close to 0, the majority of them are in the range 0-0.1;
- The *z-score normalization* gives the highest values of standard deviation among the four normalizations; the majority is in the range of 0.6-0.8;

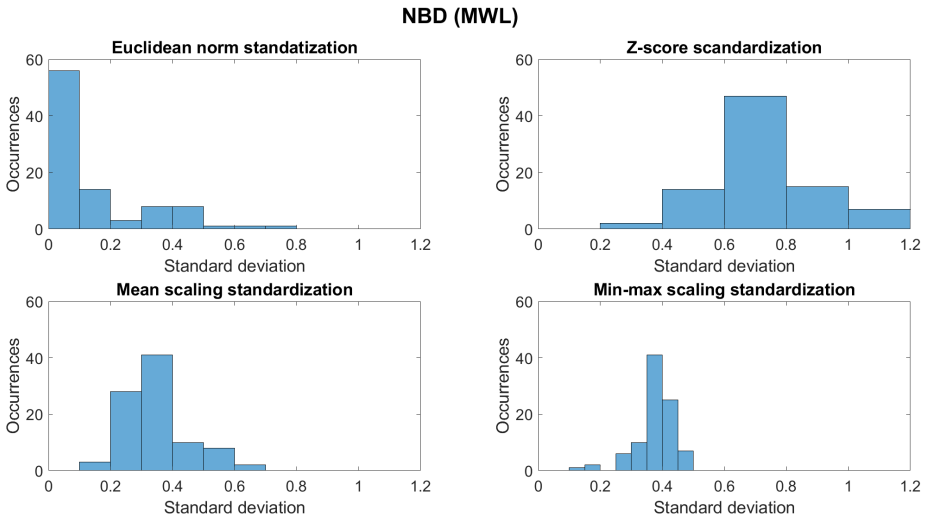

Figure 4.6: Dual N-Back Test: histogram of standard deviations

- The *mean scaling* gives values of standard deviation between the ones given by the *Euclidean norm* and the ones given by the *z-score*; the majority of standard deviations have values between 0.3 and 0.4;
- The *min-max scaling* gives values of standard deviation between the ones given by the *Euclidean norm* and the ones given by the *z-score*; the majority of standard deviations are around 0.4.

The normalization that permits the smallest error bars is the one that has a distribution of the standard deviation closer to zero; as shown in the pictures, the normalization done through *Euclidean norm*, in all four tests, is the one that satisfies this condition. This means that the error bars will be shorter if a normalization done through *Euclidean norm* is used.

The **normalization through *Euclidean norm*** is applied to each row, so to each participant, of each data set of all the features.

4.4 Dataset Investigation

A dataset investigation is carried out. Each matrix containing the values of each feature for all the participants is loaded one by one. The first step is the division of the matrix into four submatrices, or data sets, one for each test (Stroop, Visual N-Back, Auditory N-Back, and Dual N-Back). Each row corresponds to a participant, and in the first column, there are the values of the feature in the initial relax, while in the last three columns, there are the values of the feature in the three levels of

each test. A normalization through the *Euclidean norm* on each participant of each data set is carried out. Dataset investigation is performed on four classes according to the subjective perception in the participants themselves of the level of stress, caused by the Stroop Test, and MWL, caused by N-Back Tests, as explained in section 4.2. The first phase of relaxation is assigned to class 0, while the other three classes (1, 2, 3) are filled according to the participants' answers to the questionnaire.

They are plotted:

- The histograms that represent the distributions of classes in each phase related to Stress, for the Stroop Test, and MWL, for the N-Back Tests.
- The mean value of the normalized feature for each class and the standard deviation for each class, for each data set.
- The boxplots of the data in each class in each data set, to see how much the data are widespread and how many outliers there are.

4.4.1 Dataset distribution in classes

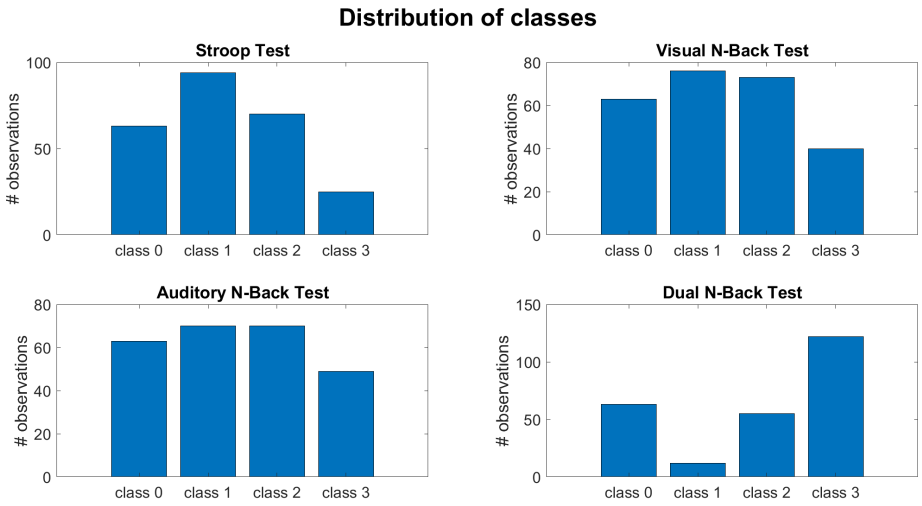

Figure 4.7: Distribution of classes

The histograms in Fig. 4.7 represent the distributions of classes in each phase related to Stress, for the Stroop Test, and MWL, for the N-Back Tests. Each class is well represented in every test, unless Class 1 in Dual N-Back Test. This is anyway considered suitable for this analysis. Class division is carried out for each feature for each dataset; in particular, the values of the feature in the first phase of

relaxation are assigned to class 0, while the other three classes are filled according to the answers of the participants to the questionnaire.

4.4.2 Dataset investigation results

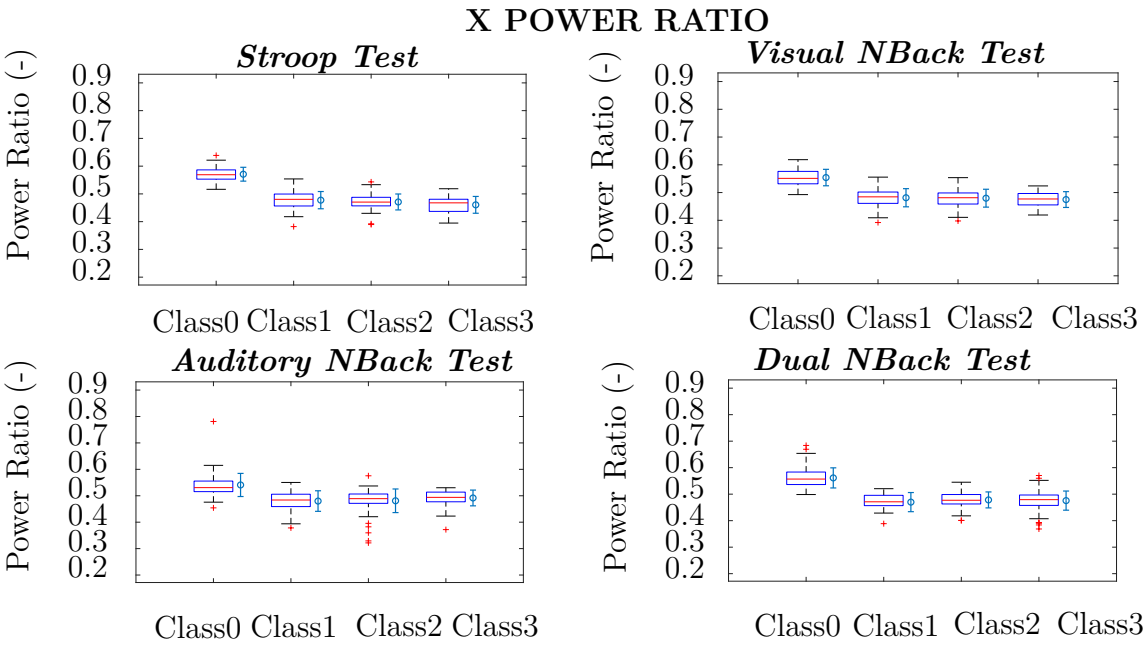

Figure 4.8: Data set investigation: Ratio of powers in horizontal signal

.

As said previously, for each feature, and for each dataset, the mean value and the standard deviation value, according to the four classes, are represented. Above two features that qualitatively show a significant difference between rest state and altered state, and other two features that qualitatively don't show any difference between the two states are shown. A statistical analysis must confirm these qualitative observations.

Fig. 4.8 and 4.9 show a sharp distinction between the values of the class 0 and the ones of the other three classes; this leads to the hypothesis that these features could be used for a binary classification between rest and altered states for all the four tests.

In Fig. 4.10 and 4.11, no difference can be graphically noticed between the four classes; this leads to the hypothesis that these features couldn't be used for a binary classification between rest and altered states. As said previously, this hypothesis must be confirmed during the statistical analysis. Further plots of mean and standard deviation and of boxplots of the features are in Section B.

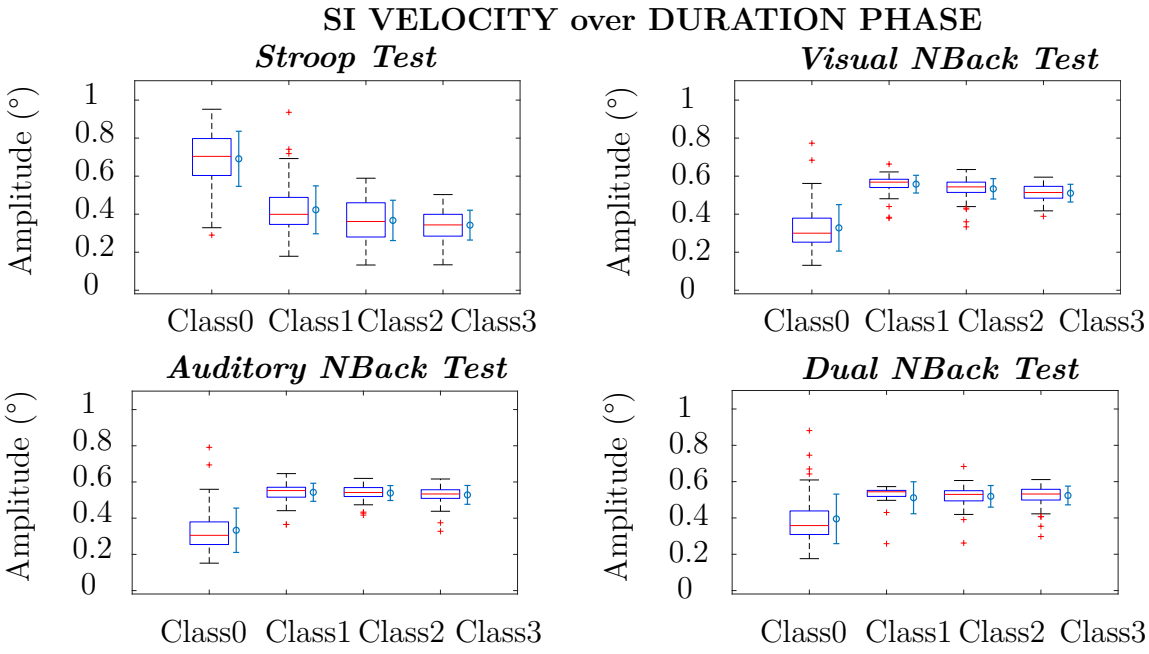

Figure 4.9: Data set investigation: single SI velocity over phases duration

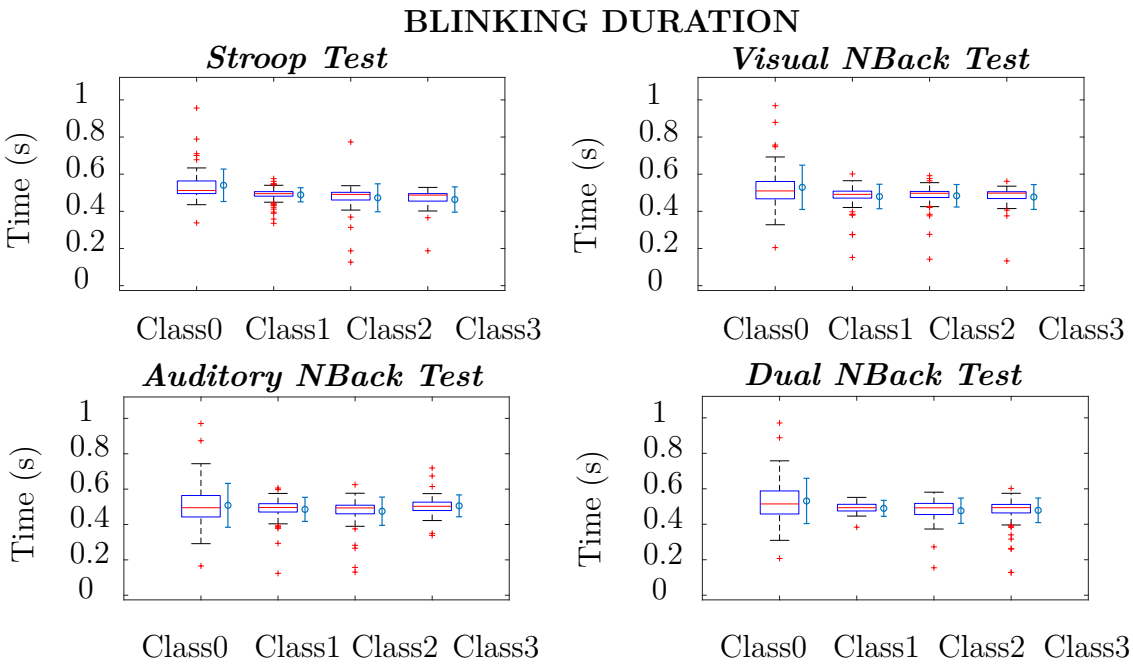

Figure 4.10: Data set investigation: Blinking duration

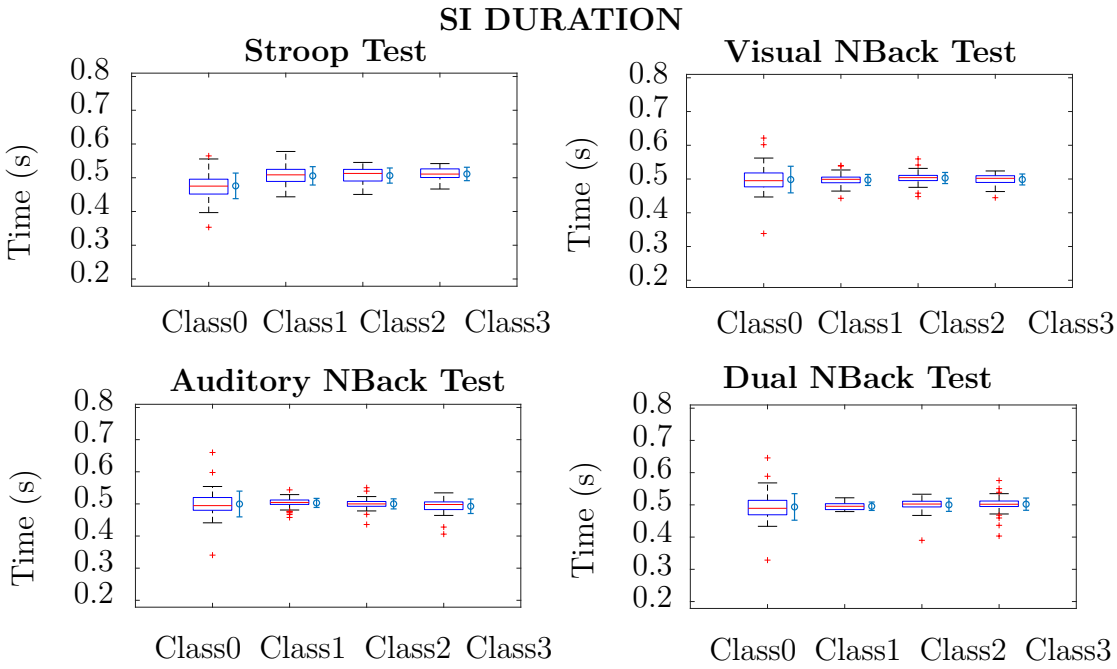

Figure 4.11: Data set investigation: SI duration

4.5 Comparison with Tobii Pro Lab

Two different kinds of comparison of the two methods, Tokuda et al. [25] algorithm and Tobii Pro Lab, could be done. First, the comparison of saccade and fixation points found through the first method and the ones identified by the second one. Then, the comparison between the features found through Tokuda et al. [25] algorithm and the ones found by Tobii Pro Lab.

Saccade and fixation samples detection

Saccades and fixations found with the Tokuda et al. [25] algorithm are compared with the ones found by the gold standard Tobii Pro Lab. In order to compare the same samples, only the instants present in both the Raw Data time axis extracted by Tobii Glasses 3, and used in eye movements detection in Tokuda et al. [25] algorithm, and the Data Export time axis from Tobii Pro Lab are compared. The parameters calculated are accuracy, sensitivity, specificity, PPV and NPV, and then the F1 score, already explained in Section 2.10. In Fig. 4.12 there is the specific confusion matrix with the values of TN, TP, FN, FP found for a participant, in the whole recording.

In Fig. 4.13 and 4.14, the parameters described related to saccades detection are calculated and shown in each macro-phase (Relax1, Stroop, Relax2, N-Back,

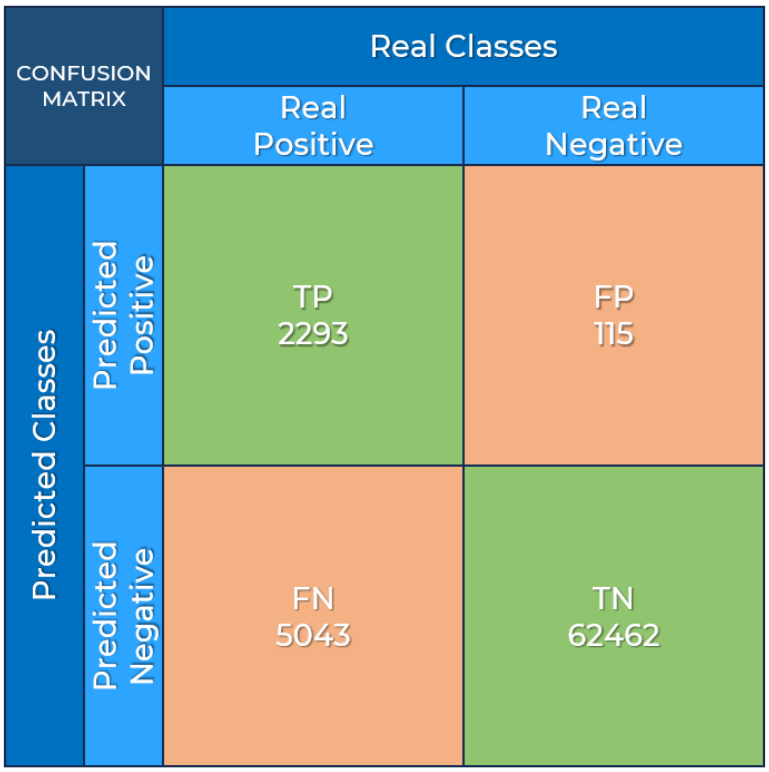

Figure 4.12: Confusion matrix for this use case

Relax3) and for the whole recording. The same is shown for fixations detection in Fig. 4.16 and 4.17. The F1 scores of saccades detection and fixation detection are shown in Fig. 4.15 and 4.18.

From the analysis of agreement between the two methods for *saccades detection* shown in Fig. 4.13 and Fig. 4.14, it could be concluded that:

- The sensitivity is around 0.4-0.5; this means that FN is high, and so the algorithm classifies fewer saccades than the gold standard;
- The specificity is high; this means that the FP are a few, and consequently, the algorithm classifies saccades well;
- The accuracy is over 0.9, which means that the sum of TP and TN is much higher than the sum of FP and FN; in particular, TN is very high being the sum of all the fixation, Unclassified, and EyesNotFound;
- Both PPV and NPV are very high and close to 1.

From Fig. 4.15, it can be concluded that the F1 score ranges from 0.5 to 0.6, which means that saccade detection is quite good.

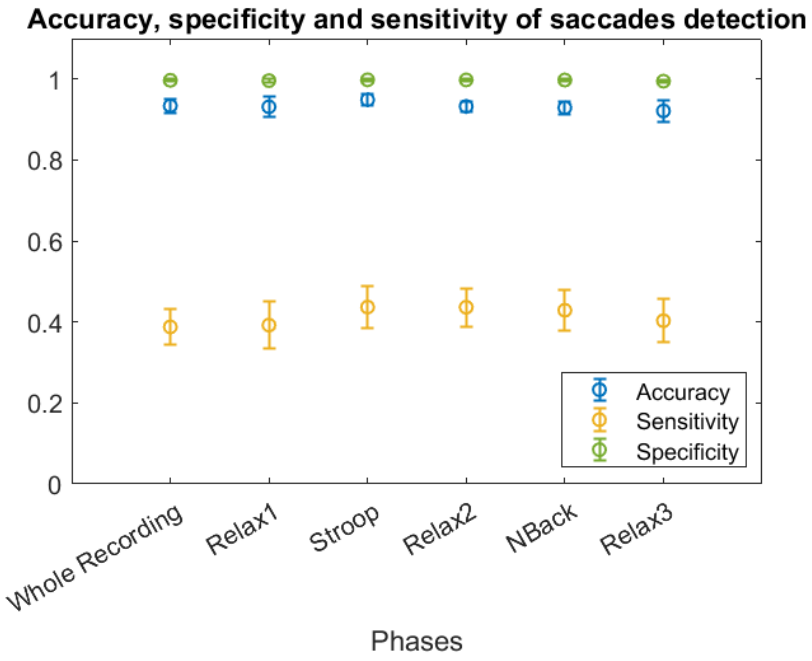

Figure 4.13: Accuracy, Specificity and Sensitivity of saccades detection

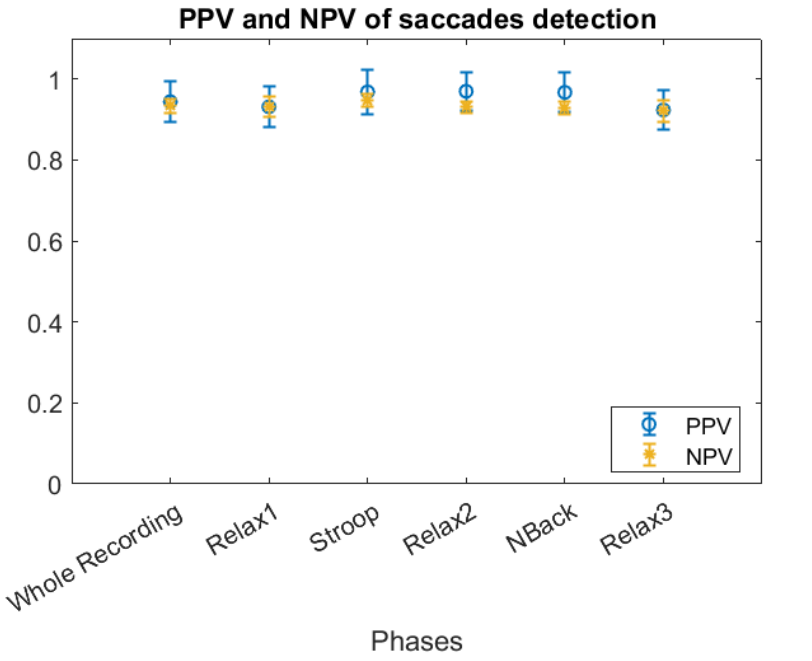

Figure 4.14: PPV and NPV of saccades detection

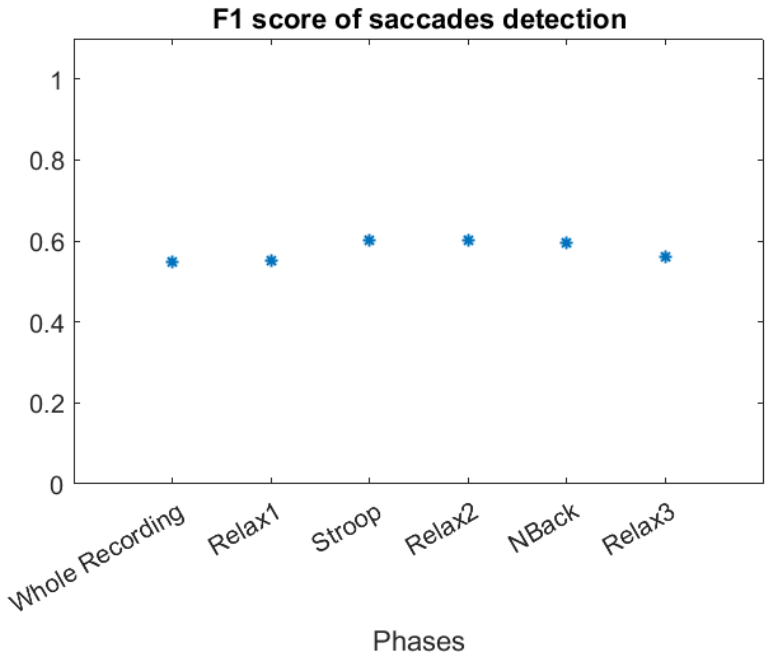

Figure 4.15: F1 score for saccade detection

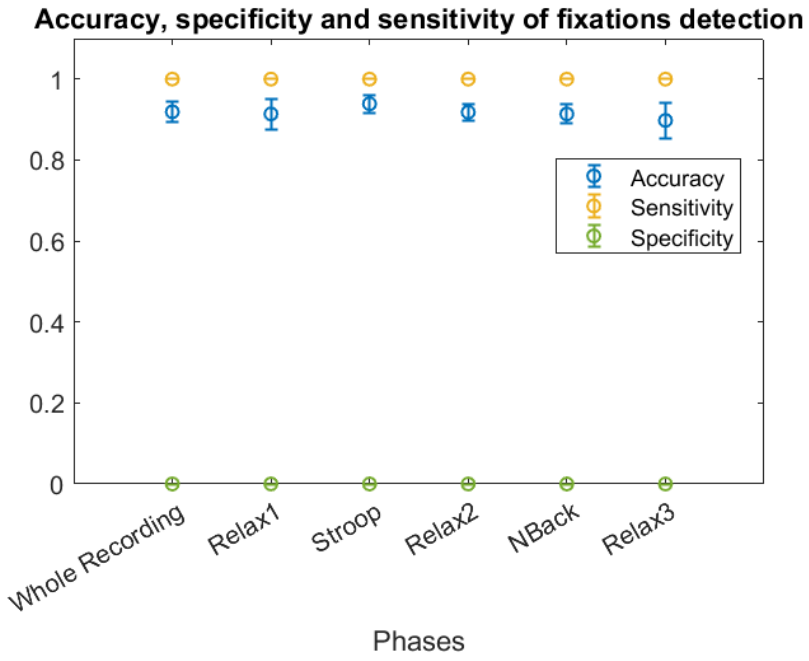

Figure 4.16: Accuracy, Specificity and Sensitivity of fixations detection

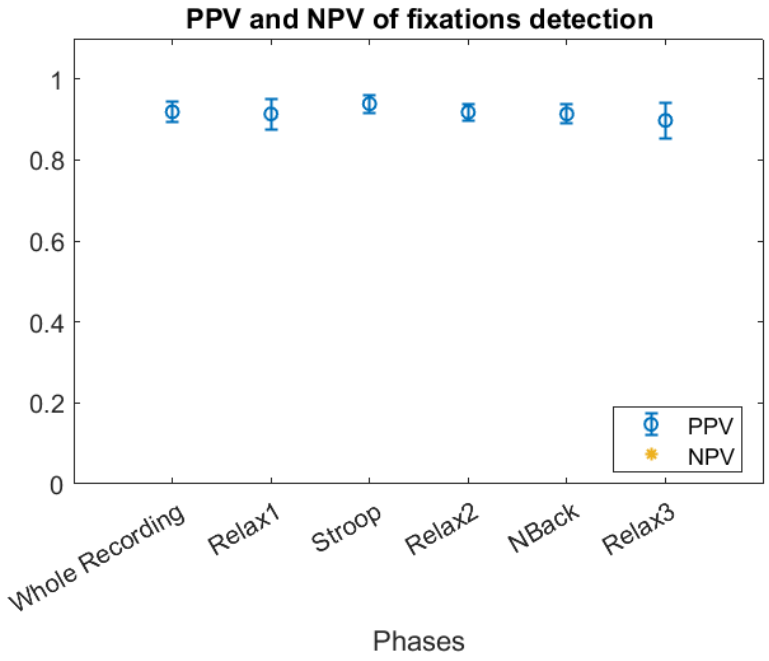

Figure 4.17: PPV of fixations detection; TN is equal to 0 so NPV can't be shown

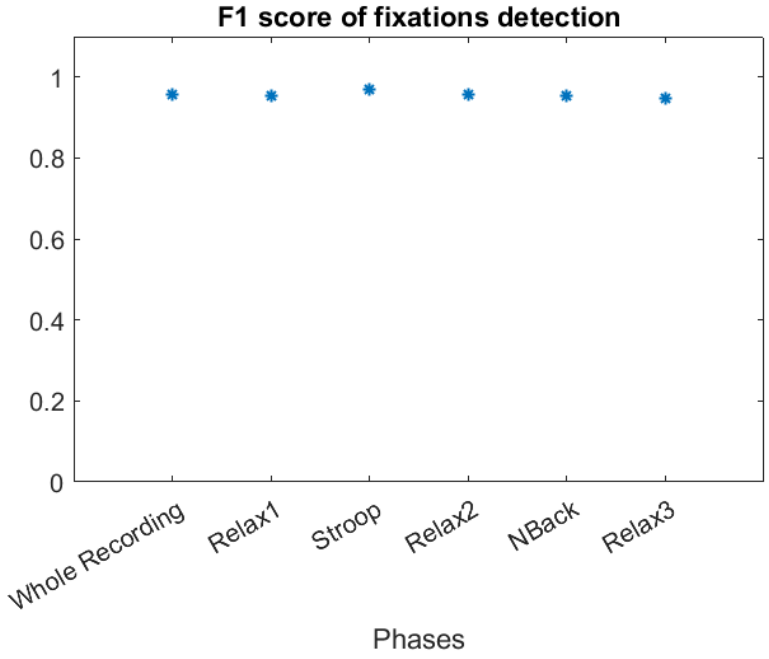

Figure 4.18: F1 score for fixation detection

From the analysis of agreement between the two methods for *fixations detection* shown in Fig. 4.16 and 4.17, it could be concluded that:

- The sensitivity is 1, which means that FN is equal to 0; this happens because the amount of fixations is much higher than the amount of all the other eye movements;
- The specificity is 0, which means that the TN is equal to 0; the reason is the same as described above;
- The accuracy is over 0.9, which means that the sum of TP and TN is much higher than the sum of FP and FN; in particular, TP is very high being the number of fixations much higher than the sum of all the other eye movements.
- NPV can't be calculated because TN is equal to 0. PPV values in all the phases are high and close to 1.

From Fig. 4.18, it can be concluded that the F1 score ranges from 0.9 to 1, which means that fixation detection is excellent; this happens also because the number of fixation samples are much higher than the number of saccade samples.

Features

The features found through Tokuda et al. [25] algorithm and the ones found by Tobii Pro Lab have been compared. This comparison is possible just for three features: *fixations duration*, *fixations frequency*, and *saccades frequency*. This is because they are the three features in common calculated by the two methods. The Metrics TSV files, explained in 3.3, of all the participants, obtained from Tobii Pro Lab, are imported. From these data, a table is created for each feature of interest. In particular, these matrices have on each column a phase of the tests, and each row contains the values of the feature of a participant. The comparison is carried out in two ways:

- The first one wants to calculate the percentage relative difference between the features calculated by Tobii and the ones calculated by the developed algorithm;
- The second one wants to create, for each phase, a graph with the values of Tobii features on the x-axis and the values of the algorithm features on the y-axis. In the graph, a point is plotted for each participant. Then, a linear fitting is carried out, and the Pearson Coefficient is calculated. The closer this coefficient is to 1 or -1, the better the correlation is. If the coefficient is 0, there is no correlation. Therefore, a Pearson Coefficient is calculated for each phase.

A figure is shown for each of the three features: duration fixations, frequency fixations, and frequency saccades. In each picture, the graph above represents the percentage of relative difference of the considered feature with a boxplot of the values. In contrast, the graph below shows the trend of the Pearson Coefficient in all phases.

- *Duration fixations*:

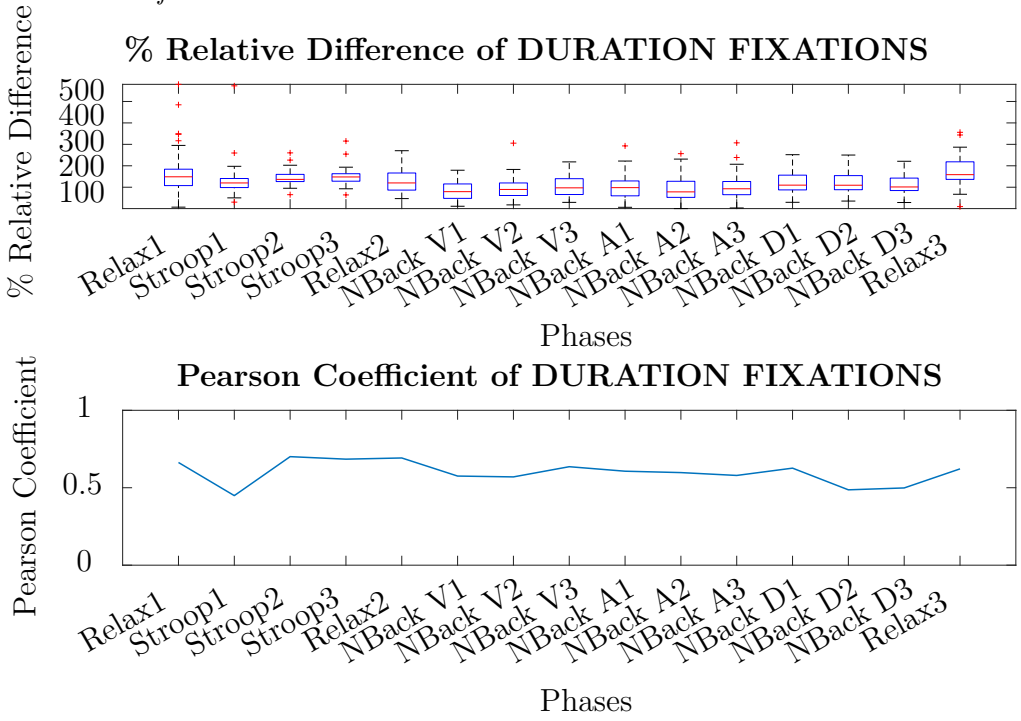

Figure 4.19: Comparison: Duration fixations

As shown in Fig. 4.21, the percentage relative difference is quite high; in the majority of phases, it is around 100-150%; this means that usually, the values of the feature calculated by the developed algorithm are, at least, the double, of the values of the same feature calculated by Tobii Pro Lab. This leads to the conclusion that the gold standard recognized more saccades than the algorithm, which is in line with the conclusions of the sensitivity of saccades detection. Moreover, the Pearson Coefficient is good since, for most of the phases it is higher than 0.5. As shown in the graph below of Fig. 4.21 the Coefficient is a bit lower than 0.5 just in the first level of the Stroop Test. This could be because during the Stroop Test phase the signal in both directions, horizontal and vertical, is more regular than in other phases. The developed algorithm, being quite conservative as regards the rules in amplitude and in time chosen, doesn't detect all the saccades in this part of the signal, and this

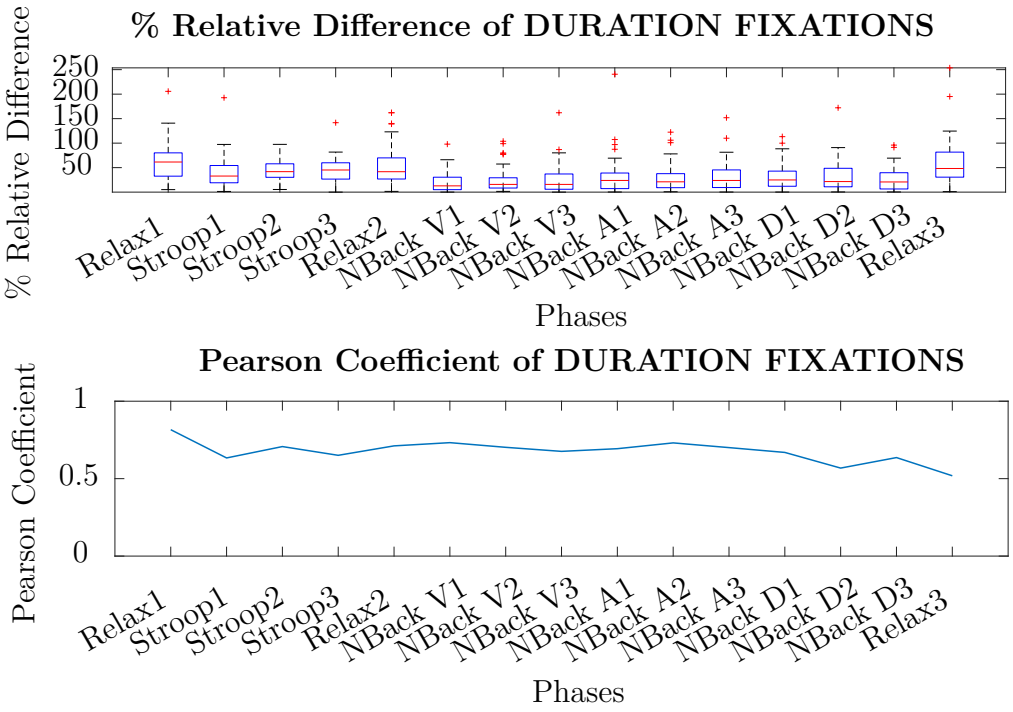

Figure 4.20: Comparison: Duration fixations with SI and saccades considered as saccades

leads to a different detection of fixations that will be longer. In fact, it could be that Tobii Pro Lab uses less restrictive rules in identifying saccades and so, what the algorithm classifies as SIs, which are the second eye movement for amplitude, the gold standard classifies as saccades. This hypothesis could be tested considering both the SI and the saccades found by the algorithm as saccades. As shown in Fig. 4.20, the percentage relative difference decreases under 50%, this is because there are more fixations that last less. The Pearson Coefficient increases above 0.6, compared to the previous situation, so the feature calculated with the two methods gives more correlated values.

- *Frequency fixations*:

  As shown in Fig. 4.21, the percentage of relative difference is less than 50%; this means that the algorithm calculates feature values higher than the ones calculated by the gold standard of less than the half of them. The Pearson Coefficient, in the majority of phases, is higher than 0.5. Just in the third relax it is next to zero, so there is no correlation between the two methods. Considering again the SIs found by the developed algorithm as saccades, it could be noticed, as expected, that the frequency of fixations, and so the percentage of relative difference, increases. The Pearson coefficient remains

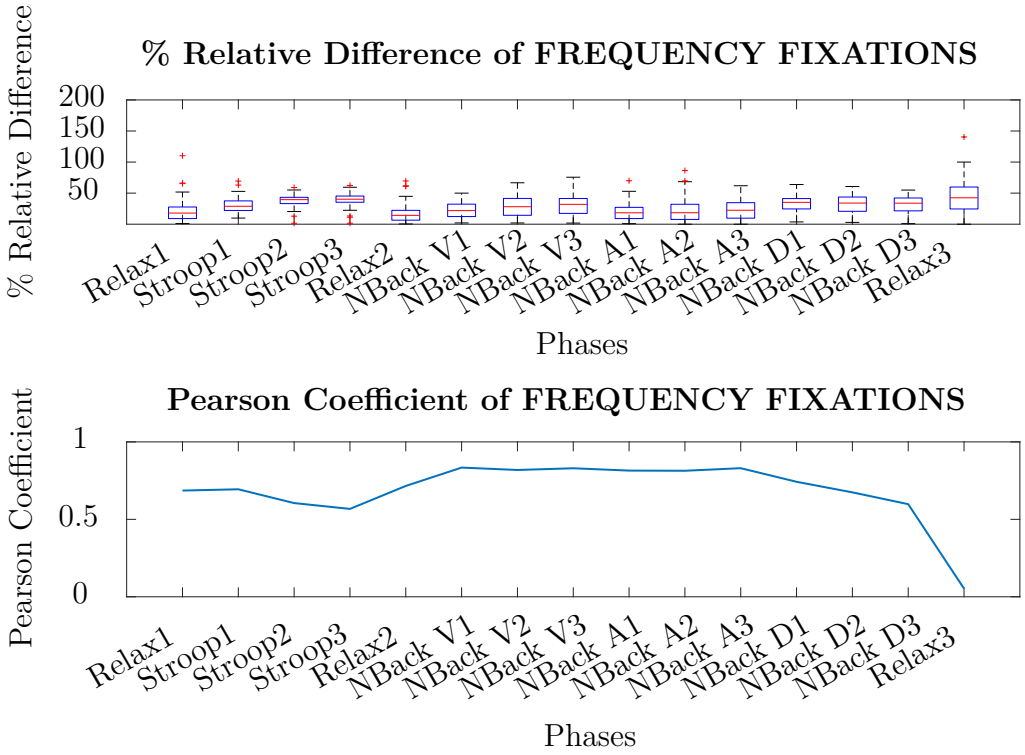

Figure 4.21: Comparison: Frequency fixations

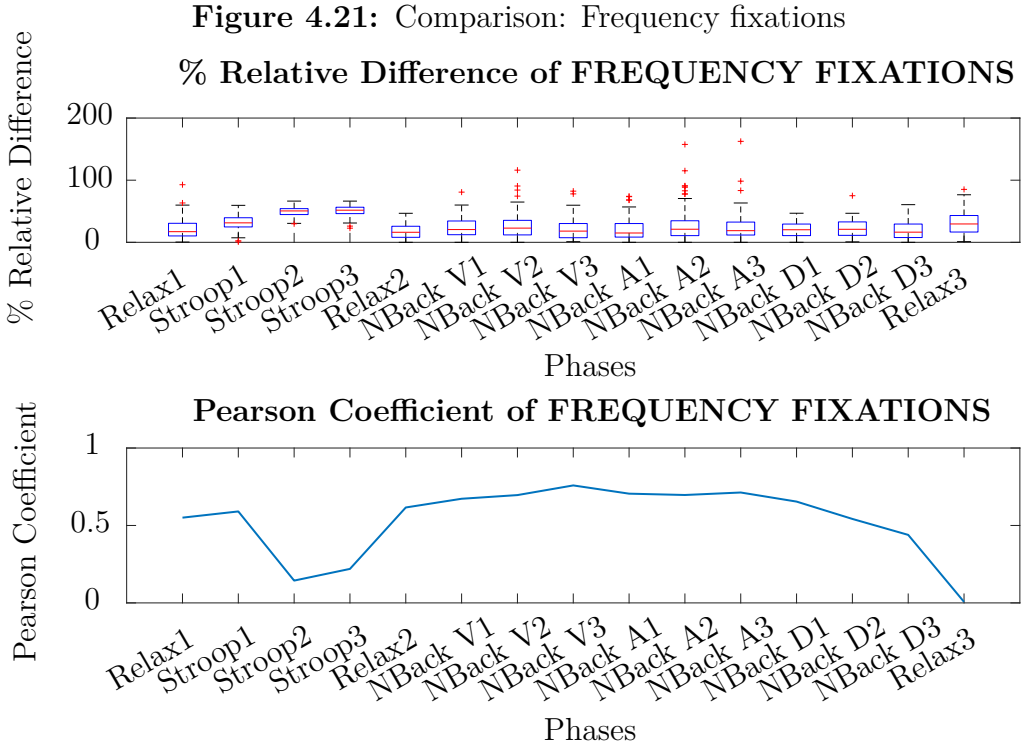

Figure 4.22: Comparison: Frequency fixations with SI and saccades considered as saccades

mainly equal to the one in the situation before: just for the second and the third level of the Stroop Test, it drops; this is probably due to the more conservative characteristic of the developed algorithm with respect to the gold standard, as explained before. This observations are supported by Fig. 4.22.

- *Saccades frequency*:

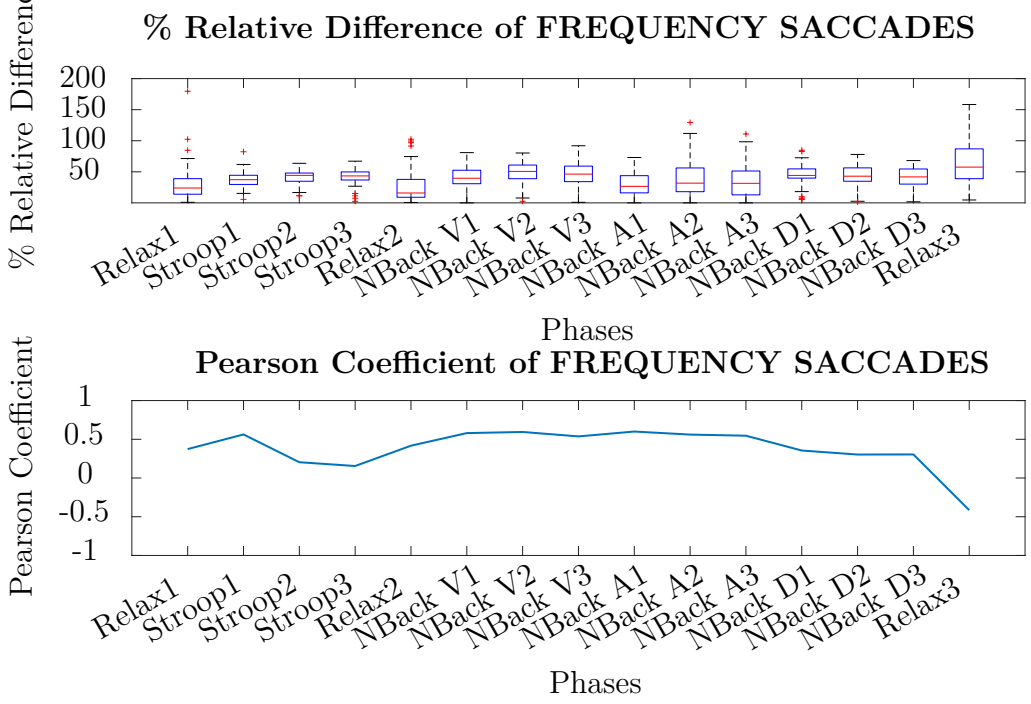

Figure 4.23: Comparison: Frequency saccades

As shown in Fig. 4.23, the percentage of relative difference is less than 50%, while the Pearson Coefficient is around 0.5 in all the phases unless the second and the third level of Stroop Test and all the levels of Dual N-Back where it is between 0 and 0.5. In the third relax phase, the Pearson Coefficient becomes negative around -0.5; this means that the relationship is inverse, but the two methods are correlated anyway.

4.6 Statistical Analysis results

The choices made about which kinds of statistical tests to carry out are explained in Section 4.6. For each feature, the outputs considered for the Kruskal-Wallis H Test are the *p-value* and the *H* statistic, while the output considered for the Mann-Whitney U Test is the *p-value*. For each feature, the summary parameters are saved in a CSV file. These results are imported and summarised. In particular,

two tables are created: each one with four columns, one for each test (Stroop Test, Visual N-Back Test, Auditory N-Back, Dual N-Back). The column related to Stroop Test comes from the stress analysis, while the three columns related to N-Back Tests from MWL analysis:

- In Table 4.3 overall statistical results are shown: in the first row, the number of features, in each phase, with a *p-value* smaller than 0.05 in the Kruskal-Wallis H Test are counted. Considering that for the Mann-Whitney U Test, there are six possible couples of classes (Class 0 - Class 1, Class 0 - Class 2, Class 0 - Class 3, Class 1 - Class 2 etc.), the next rows indicate the number of features with a number of couples with a *p-value* smaller than 0.05 bigger than n, with n going from 1 to 5 going through the table.
- In Table 4.1, the aim is to understand if there is a statistical difference between the relaxed state or the altered state; this means that there are three possible couples of classes(Class 0 - Class 1, Class 0 - Class 2, Class 0 - Class 3). The first row is the same as the previous table, while the sequent ones refer to the number of features with a number of couples with a *p-value* smaller than 0.05 bigger than n, with n going from 0 to 2 going through the table.

By analyzing Tables 4.1 and 4.3, it is possible to observe that the first row of both tables indicates if there are at least a couple of classes among the four possibles with a *p-value* smaller than 0.05. If it is present, for the feature analyzed, and for the data set taken in count, there is at least a significant difference between two classes. This statement is insufficient for drawing conclusions; that's why the Mann-Whitney U Test is performed.

4.6.1 Binary Classification Results

Table 4.1 resumes the statistical analysis of binary classification without any threshold on H value.

Table 4.1: Resume of binary classification results

	Stroop	**Visual N-Back**	**Auditory N-Back**	**Dual N-Back**
Test Significance	20	19	16	19
Feature Significance (>0)	20	19	15	19
Feature Significance (>1)	20	17	15	16
Feature Significance (>2)	18	13	13	12

Analyzing the binary classification of Table 4.1, the most significant features that allow for differentiation between a rest state and an altered state are the ones

with all the three couples of classes with a *p-value* smaller than 0.05. In other words, Class 1, Class 2, and Class 3, which represent the altered state, are statistically different from Class 0, which represent the rest state.

In particular, over 23 total features, for the Stroop Test, about 80% of them result to be significative, so able to detect a difference between rest state and stressed state. Then, for the Visual N-Back and the Auditory N-Back, the significant features are almost 60%, while for the Dual N-Back, they are over 50%; these selected features can distinguish between a rest state and a state at a high MWL. Considering the significant features in the four tests, they are 83% of the total; this means that most of the extracted metrics are able to perfectly distinguish the state of relaxation of the subject with its stressed or at high MWL state. They are, therefore, significant for at least one of the four tests. The 40% of the total number of features extracted are the resulting significative features common to all four kinds of tests. In other words, they show simultaneously sensitivity to stress and cognitive load variations. Therefore, they allow a binary classification between rest state and altered state, which could be both, stressed and with high MWL. They are listed below:

- Frequency saccades,
- Single SI velocity,
- Single SI velocity over phase duration,
- Frequency fixations,
- Power provided by HF bandwidth in horizontal signal,
- Power provided by LF bandwidth in horizontal signal,
- Ratio of powers in horizontal signal,
- Power provided by HF bandwidth in vertical signal,
- Ratio of powers in vertical signal.

Boxplots and plots of these features' mean and standard deviation are shown in the section B.

In particular, concerning feature trends, the altered state has smaller values than the relaxed state for saccades and fixations frequency, for the ratio of powers in both vertical and horizontal signals, and for single SI velocity. While the altered state has higher values than the relaxed state for the powers provided by HF and LF bandwidth in both, vertical and horizontal signals and for the single SI velocity over phase duration.

Moreover, here it could be confirmed the significance of the ratio of powers in horizontal signal and of the SI velocity over phases duration used as examples previously and shown in Fig. 4.8 and 4.9. These two features permit a perfect distinction between the relaxed state and the altered states. Instead, the features about the SIs duration and the duration of blinking, shown in Fig. 4.11 and 4.10, don't permit a full distinction between the two states; the number of couples (Class 0 - Class 1, Class 0 - Class 2, Class 0 - Class 3) with a *p-value* smaller than the level of significance is not equal to three, so the maximum, but lower.

H statistic

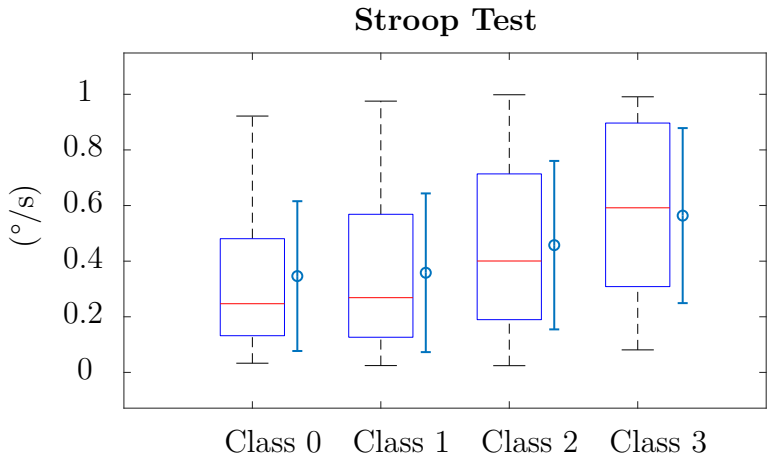

Figure 4.24: SI value in Stroop Test dataset, low *H* statistic

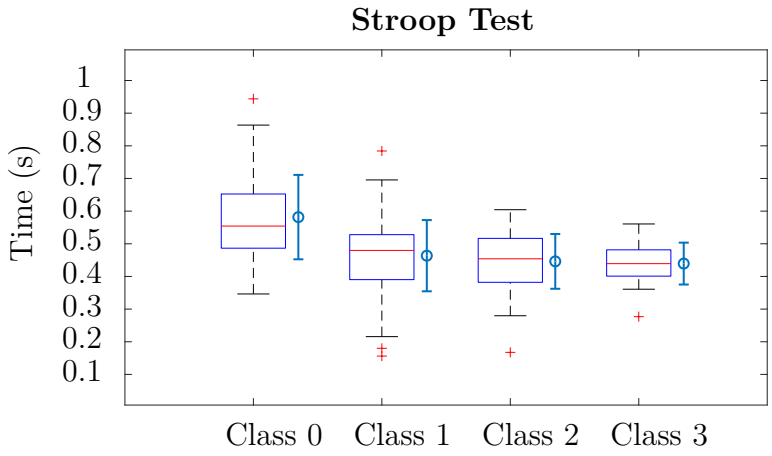

Figure 4.25: Fixation lasting in Stroop Test dataset, medium *H* statistic

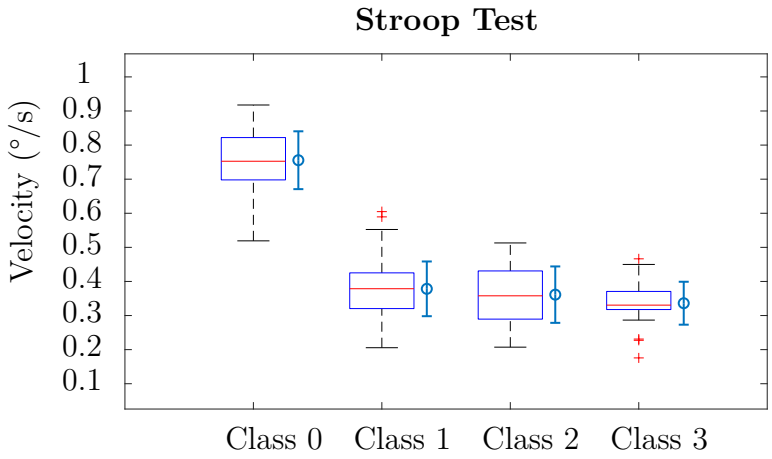

Figure 4.26: Single SI velocity in Stroop Test dataset, high H statistic

In order to enhance the statistical evaluation level, it is decided to refine the thresholds of the H statistic, the second output of the Kruskal-Wallis H Test. This would lead to the discard of features, significant for the Kruskal-Wallis H Test, but not so highly related to variation in stress and MWL. In Fig. 4.24, 4.25 and 4.26 are shown, as examples, three features for the Stroop Test dataset, with different values of H statistic in order to show their graphical differences.

It could be noticed in Fig.4.24 that the SI value among the four classes results to be significant for the Kruskal-Wallis H Test but with an H statistic equal to 14; this means that the difference between the four classes is not very high. For the research, a feature with this value of H statistic must be discarded in order to be as strict as possible. In Fig. 4.26, it can be seen that the single SI velocity has the values of class 0 and of the other three classes in two completely different bandwidths. In fact, the H statistic value confirms this high significance of the feature being equal to 143. All the features with this kind of significance must be kept for the statistical analysis. Fig. 4.25 shows the fixation lasting; the H statistic value is 48. It is decided to keep the features with this range of H values because, even if there isn't a sharp distinction between rest state and altered states, there are anyway mostly different values. After observing all the features with these more dubious values, it is decided to use a threshold on the H statistic values of 44.

Table 4.2 is created, with the same structure of 4.1, but with the imposition of the threshold on the H statistic. So, the first row will be about the features that result to be significant at the Kruskal-Wallis H Test and overcome the threshold on the H statistic. The other rows are about the results of the Mann-Whitney U Test applied just to the features "saved" in the H statistic threshold application. So, Table 4.2 shows the results of binary classification but just for the features that overcome the Kruskal-Wallis H Test with the threshold on the H statistic.

Table 4.2: Resume of binary classification results with H statistic threshold

	Stroop	**Visual N-Back**	**Auditory N-Back**	**Dual N-Back**
Test Significance	14	10	6	8
Feature Significance (>0)	14	10	6	8
Feature Significance (>1)	14	10	6	7
Feature Significance (>2)	14	10	6	6

In particular, the first row can be compared to the first row of Table 4.1: it is shown that more or less half of the features that result to be significant at the test have a H statistic smaller than 44. Just for the dataset related to the Stroop Test, three-quarters of the features are still significant. It can be noticed that, even if for some datasets there is a drop in the number of features that result to be significant in the Mann-Whitney U Test going through the table, the number of features in the fourth row (Feature Significance (>2)) are very close to the number of features in the first row (Test Significance at the Kruskal-Wallis H Test). This means that the threshold put on the H statistic saves the features most significant, with a sharper difference between the rest state and altered states. The 17% of the total number of features extracted are the resulting significative features common to all four kinds of tests with the application of this threshold. So, they show simultaneously a very high sensitivity to stress and cognitive load variations. Therefore, they allow a binary classification between rest state and altered state, which could be both, stressed and with high MWL. They are listed below:

- Single SI velocity,
- Single SI velocity over phase duration,
- Ratio of powers in horizontal signal,
- Ratio of powers in vertical signal.

4.6.2 Overall Statistical Results

In order to increase the granularity of stress and MWL conditions prediction, it is important to analyze Table 4.3. From the fourth row (Feature Significance (>3)) could be observed a sharp drop of the number of features with a number of couples between classes with a *p-value* smaller than 0.05 in the Mann-Whitney U Test. This means it is difficult to distinguish the four states one from another perfectly. Moreover, the majority of features that allow a higher granularity of prediction

Table 4.3: Resume of overall statistical results

	Stroop	**Visual N-Back**	**Auditory N-Back**	**Dual N-Back**
Test Significance	20	19	16	19
Feature Significance (>1)	20	19	15	18
Feature Significance (>2)	19	17	14	14
Feature Significance (>3)	14	5	2	6
Feature Significance (>4)	6	1	0	1
Feature Significance (>5)	0	1	0	0

(Feature Significance (>3)) also permit a binary classification between relaxed and altered states.

The features listed in Table 4.27 are the ones with a number of couples between classes with a *p-value* smaller than the level of significance in the Mann-Whitney U Test higher than three. The features colored in blue are the ones with a number of couples with a feature significance higher than four; they will be listed below. The only feature colored in yellow, single SI velocity over duration phase, is the one with all the couples showing a statistical difference among them, just for the Visual N-Back Test. Crossing the results of binary classification with the ones just shown, there are only two features that don't permit a binary classification but permit a beginning of a trend identification among the four classes: they are the SI value for the Stroop Test and the SIs frequency for the Visual N-Back Test. All the other features permit a binary classification between relaxed and altered states and moreover a beginning of a trend indentification. In particular, the features that permit, with a feature significance of the Mann-Whitney U Test higher than four couples, a good trend identification for the Stroop Test are:

- Interval blinking;
- Frequency saccades;
- frequency fixations;
- Frequency SIs;
- Single SI velocity over phase duration;
- Ratio of powers in vertical signal.

The feature that perfectly permits a trend identification for the Visual N-Back Test is:

STROOP	VISUAL N-BACK	AUDITORY N-BACK	DUAL N-BACK
Duration blinking	Frequency blinking	Interval blinking	Frequency saccades
Frequency blinking	Frequency fixations	Single SI velocity	Frequency fixations
Interval blinking	Frequency SI		Relative diameter right
Frequency saccades	Single SI velocity		Frequency SI
Frequency fixations	Single SI velocity over duration phase		HF Y
SI value			Ratio Y
LF X			
Ratio X			
Duration saccades			
Frequency SI			
Velocity SI			
Single SI velocity			
Single SI velocity over duration phase			
Ratio Y			

Figure 4.27: Overall results: features resulting to have a feature significance >3

- Single SI velocity over phase duration.

No features allow a trend identification with more than four couples of classes with a statistical difference in Auditory N-Back.

Finally, the feature that better permits a trend identification for the Dual N-Back Test is:

- Relative diameter of the right eye.

H statistic

Setting the threshold on the H statistic it could be seen that the first three rows of Table 4.4 are almost equal, that's why the feature selected are more significant and so they result to permit also an higher granularity.

Table 4.4: Resume of overall statistical results with H statistic threshold

	Stroop	**Visual N-Back**	**Auditory N-Back**	**Dual N-Back**
Test Significance	14	10	6	8
Feature Significance (>1)	14	10	6	8
Feature Significance (>2)	14	10	6	8
Feature Significance (>3)	11	3	2	3
Feature Significance (>4)	5	1	0	1
Feature Significance (>5)	0	1	0	0

4.6.3 Resume of statistical analysis results

In conclusion, we can therefore say that the extracted features are, for the most part, able to make a binary classification differentiating the state of relaxation from the altered one, which is due to stress or MWL. In particular, it is important to note that among the most significant features, there are many never investigated before in the literature, including those resulting from the frequency analysis of the signal. Almost all those extracted from the PSD of the signal, are metrics able, without any threshold imposition on H statistic, to identify in two distinct bandwidths the states of relaxation and altered. While, with the setting of this threshold, the ratio features on both the horizontal and the vertical signal allow a binary classification. Moreover, it could be seen that a greater granularity of results is more difficult.

Chapter 5

Conclusions

HMI system design focuses on the safety and reliability of the performance of the subject that interacts with the machine. He or she can be submitted to high levels of stress and MWL that the HMI system must be able to detect, measure and report in real-time. Many physiological signals have been seen to be sensitive to stress and MWL variations; one of them is the signal coming from eye movements. It seems to be promising in this field of research, but there is still no clear relationship in the literature between it and different stress and MWL levels. This study wants to investigate this relationship reaching all the objectives listed in Section 1.1.

At first, the definition of the best significant eye-tracking features to evaluate stress and mental workload variations has been carried out through a deep research in literature. Other features never investigated in literature concerning stress and MWL variations were added, almost the 50% of the totality, with a total number of 23. Then, the definition and implementation of tests to detect mental workload variations to gather a significant dataset has been carried out. The test session consisted of the performing of the Stroop Test, specifically created for generating stress conditions in the subject, and the N-Back Test, in three versions, to develop different levels of cognitive load in the participant, by the 64 participants enrolled. The signal acquired through Tobii Glasses 3 during these tests was then analyzed. The extraction of the selected 23 features from the data gathered from the tests is carried out. These metrics concern the ocular movements analyzed both, in the time and frequency domain, and the pupil diameter. Moreover, a questionnaire has been filled in by the subjects and, according to their answers to it, four classes have been created.

The main objective of this thesis study is the evaluation of the correlation between stress, mental workload variations, and eye tracking parameters variations. The achievement of this goal is supported by the results of the statistical analysis: it has been seen how 83% of the extracted features allow a binary classification between the relaxation state and altered state in at least one type of test. The

features that enable this type of classification for all four tests are the 40% on the totality. Moreover, in order to have rest state and altered states in two sharply separated bandwidths of values a threshold on the H statistic value, one of the two outputs of the Kruskal-Wallis H test, has been used. The number of significant features left in the binary classification decreases and the percentage of features extracted, in common to all four kinds of tests, resulting significative drops to 17%.

Speaking about the overall statistical result, with or without any threshold on the H statistic value, it could be concluded that it is difficult to distinguish the four states one from another perfectly, so a greater granularity of prediction of stress and MWL from the analysis of features is more complex.

In particular, the majority of the new features extracted from frequency domain analysis result to be significant in the binary classification without the H statistic threshold and the ratio of power on both signals, horizontal and vertical, is significant also with the threshold. This means that it detects finely the distinction between rest state and altered states.

To achieve these macro goals, other steps have been made and other conclusions have been drawn. The first one consists of ascertaining that the formulas for the conversion of gaze location from pixels to degrees, and between the two reference systems, are consistent in this use case. It was carried out through both Assisted and Manual Mapping of Tobii Pro Lab, and they give angles close to the ideal ones. This leads to the conclusion that the initial claim is verified. Another observation is that for most of all plots, the Assisted Mapping gives an error visibly higher than the Manual Mapping, this means that Manual Mapping better permits calculated angles closer to the ideal ones.

Another objective of the research is, after the identification of different kinds of eye movements with the developed algorithm, their comparison with the ones found by the gold standard in the field of eye-tracking. This comparison between the two methods of classification of eye movements has been carried out in two different ways: the first one is a comparison of the saccade points found, while the second one is the comparison of three features in common calculated by the two methods. Regarding the saccade points, looking to Fig. 4.13, it could be concluded that all the saccade points detected by the algorithm are also detected by the gold standard, but not all the gaze points classified by Tobii Pro Lab as saccades are also found by Tokuda's algorithm. Speaking about the second kind of comparison, it could be concluded that for all the three features, saccade frequency, fixation frequency, and lasting, the correlation between the two methods is quite good, having mainly a Pearson Coefficient higher than 0.5.

Finally, thinking about future studies in this field that should be carried out, it must be noticed that among the features extracted those not yet investigated in the literature give encouraging results. In particular, metrics regarding frequency analysis are promising and should be deepened. Moreover, it is demonstrated that

the eye signal is very sensitive to changes in stress and cognitive load and, therefore, deserves further research in this field.

With the aim of the creation of safer HMI systems with risk reduction for humans, the future steps to carry out in this field are, at first, the implementation of results with other biosignals. The objective is to create a model for discrimination of stress and MWL states from rest state with just the most significant features. Then, the development of a wearable device for stress and MWL control by the acquisition of some selected biosignals should be carried out.

Appendix A

Pixels-degrees conversion validation

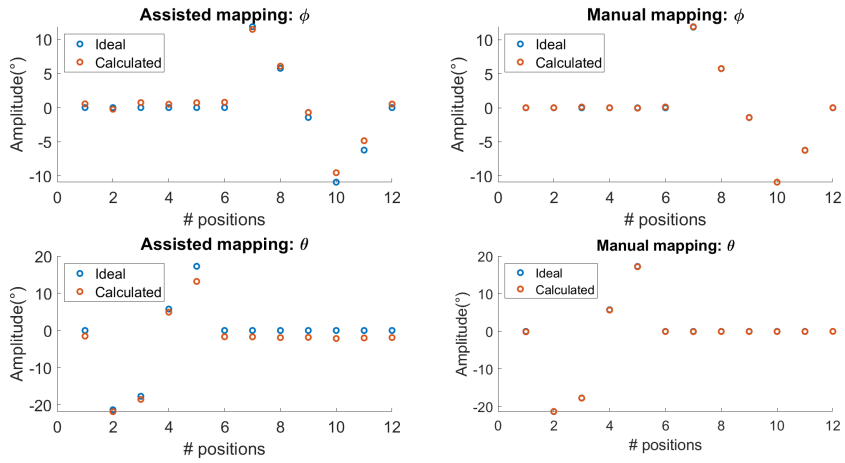

Figure A.1: Comparison of ideal and calculated angles: Recording 2

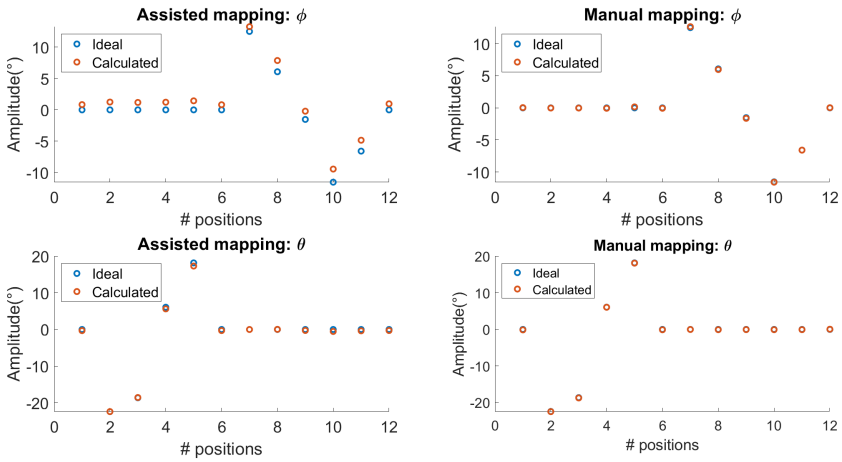

Figure A.2: Comparison of ideal and calculated angles: Recording 3

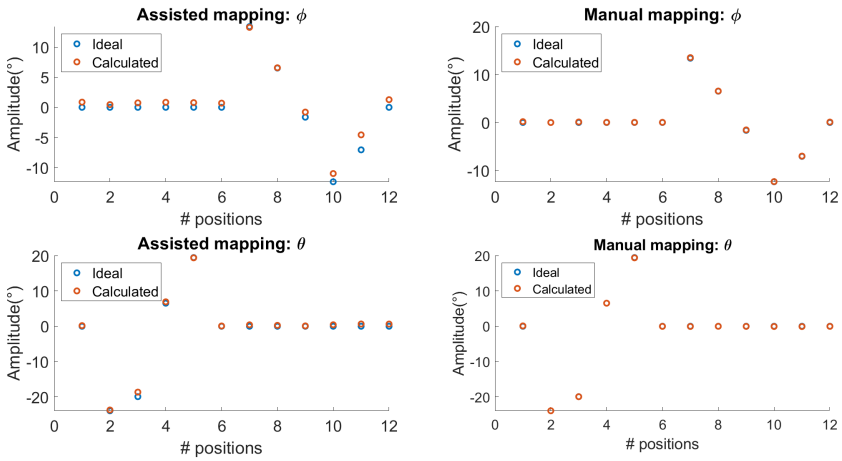

Figure A.3: Comparison of ideal and calculated angles: Recording 4

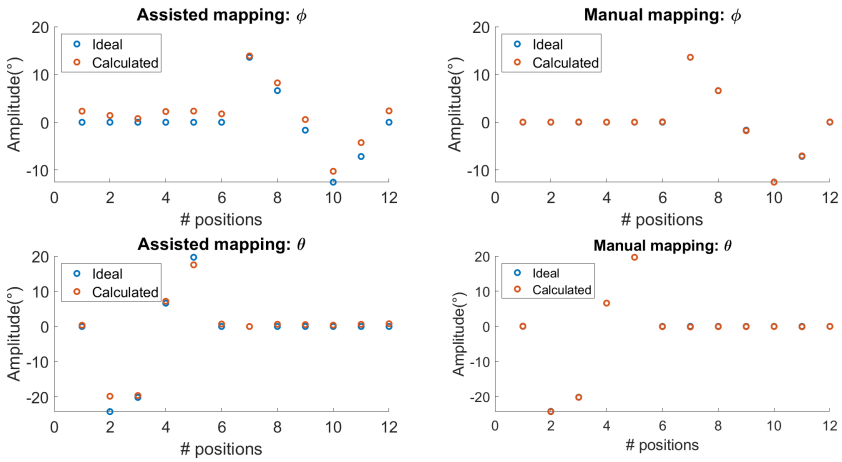

Figure A.4: Comparison of ideal and calculated angles: Recording 5

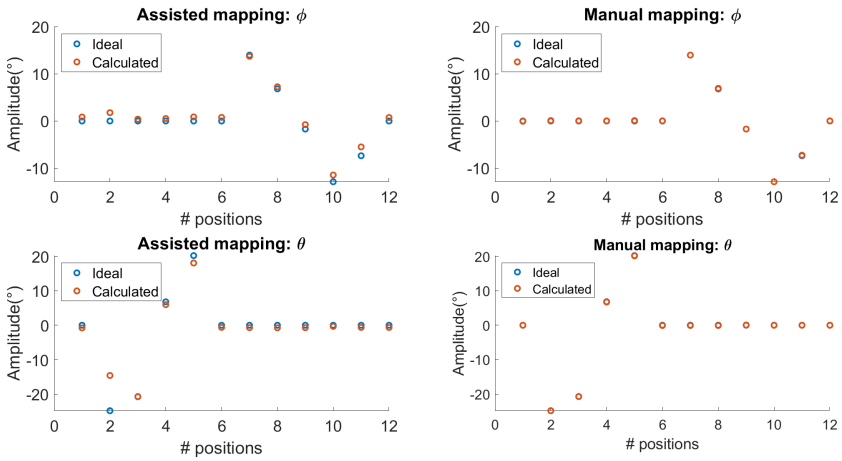

Figure A.5: Comparison of ideal and calculated angles: Recording 6

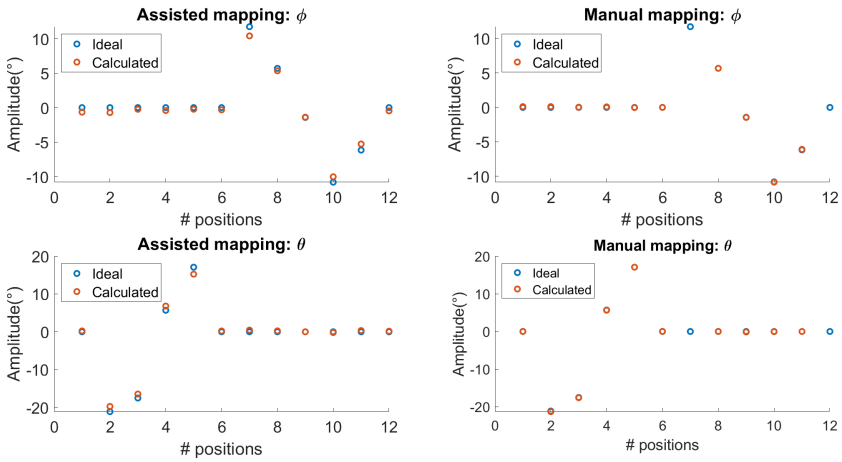

Figure A.6: Comparison of ideal and calculated angles: Recording 7

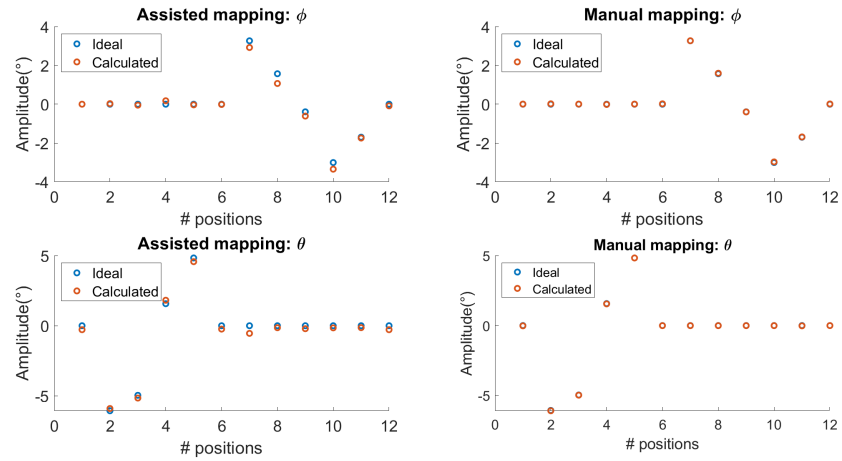

Figure A.7: Comparison of ideal and calculated angles: Recording 8

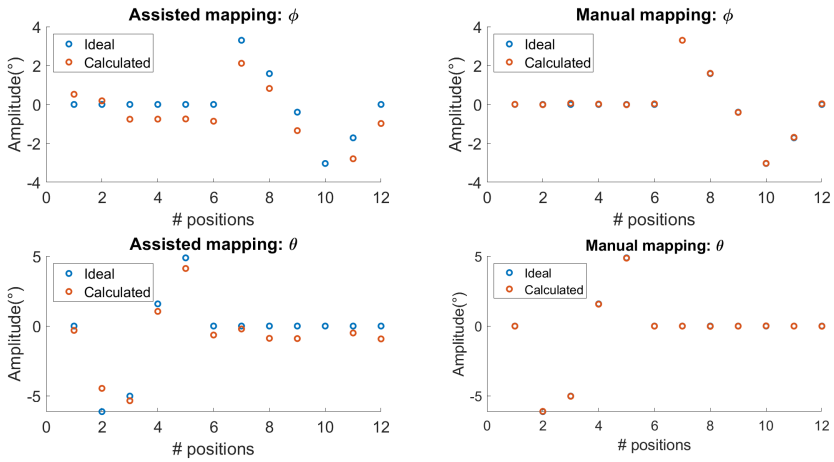

Figure A.8: Comparison of ideal and calculated angles: Recording 9

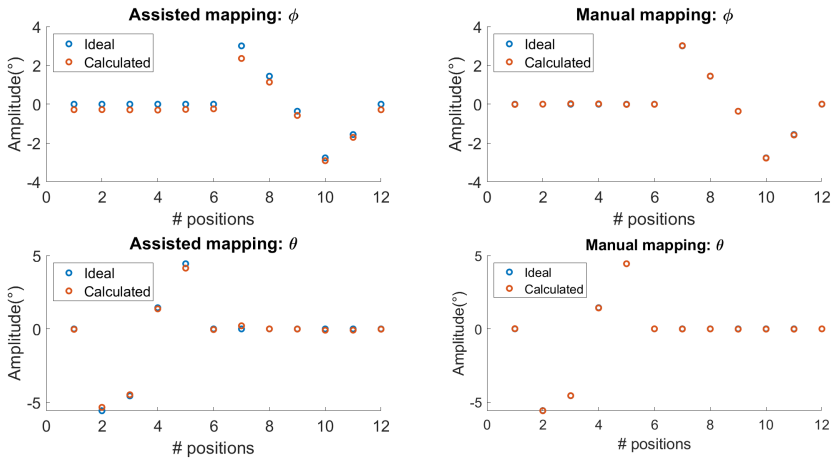

Figure A.9: Comparison of ideal and calculated angles: Recording 10

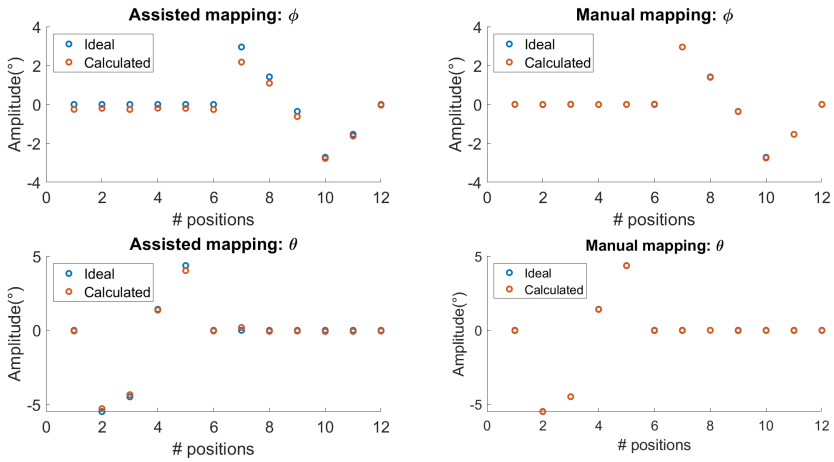

Figure A.10: Comparison of ideal and calculated angles: Recording 11

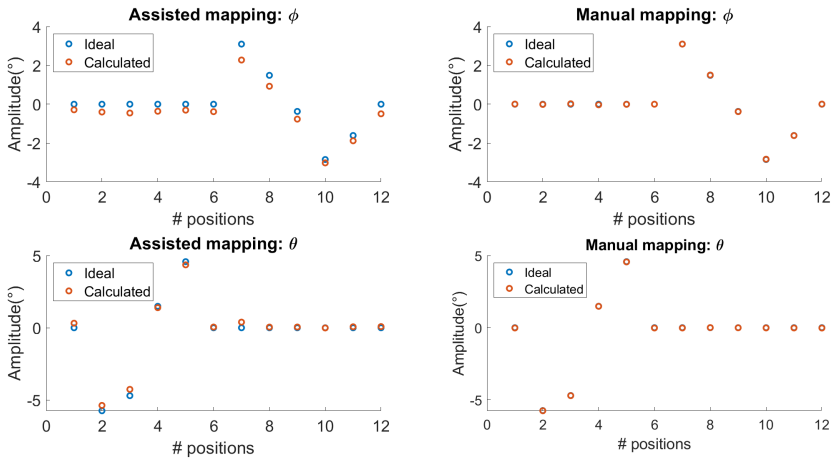

Figure A.11: Comparison of ideal and calculated angles: Recording 12

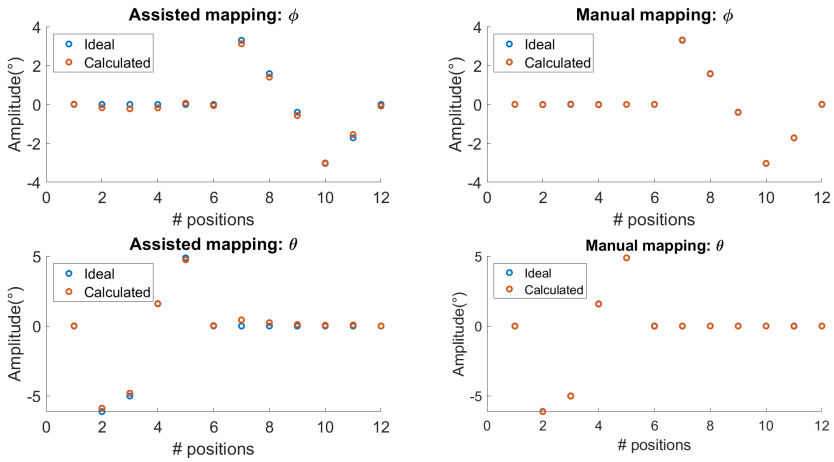

Figure A.12: Comparison of ideal and calculated angles: Recording 13

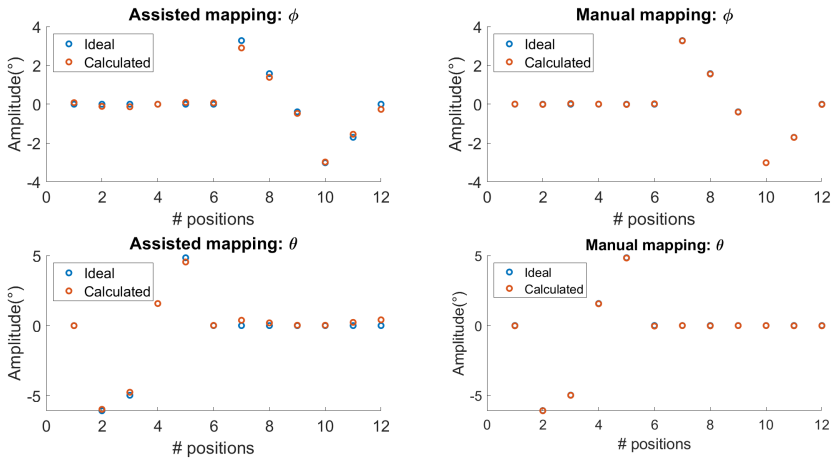

Figure A.13: Comparison of ideal and calculated angles: Recording 14

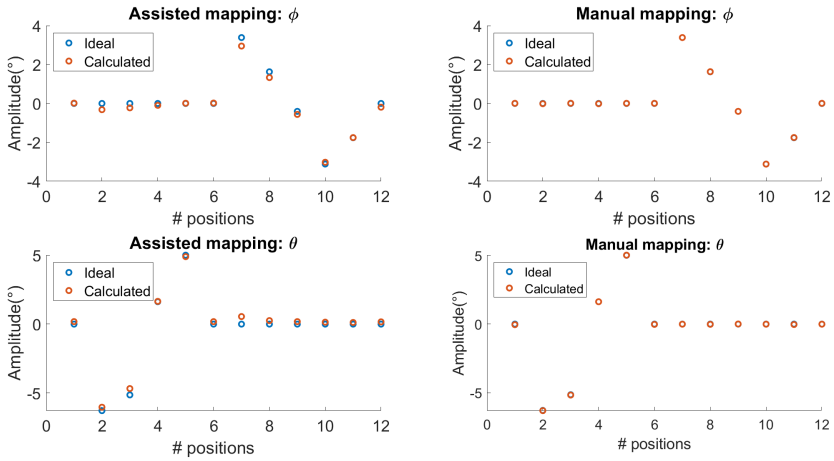

Figure A.14: Comparison of ideal and calculated angles: Recording 15

Appendix B

Features boxplots

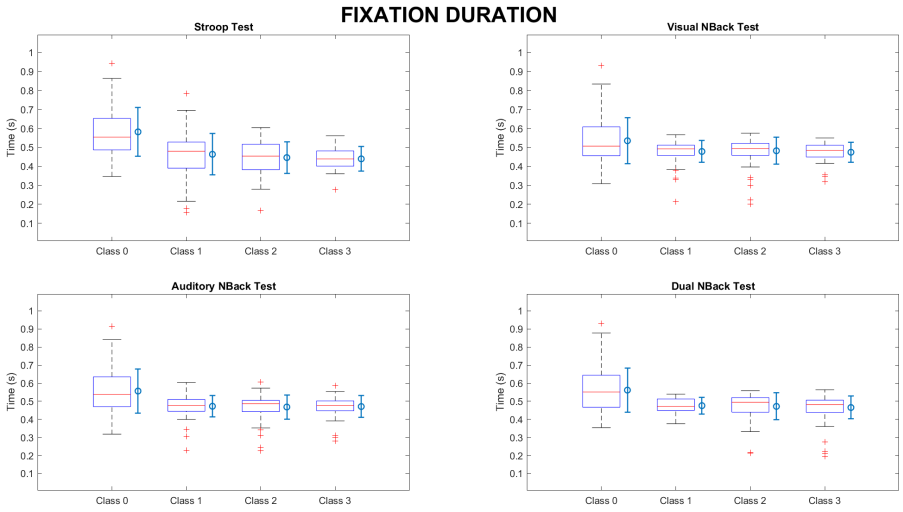

Figure B.1: Duration fixations

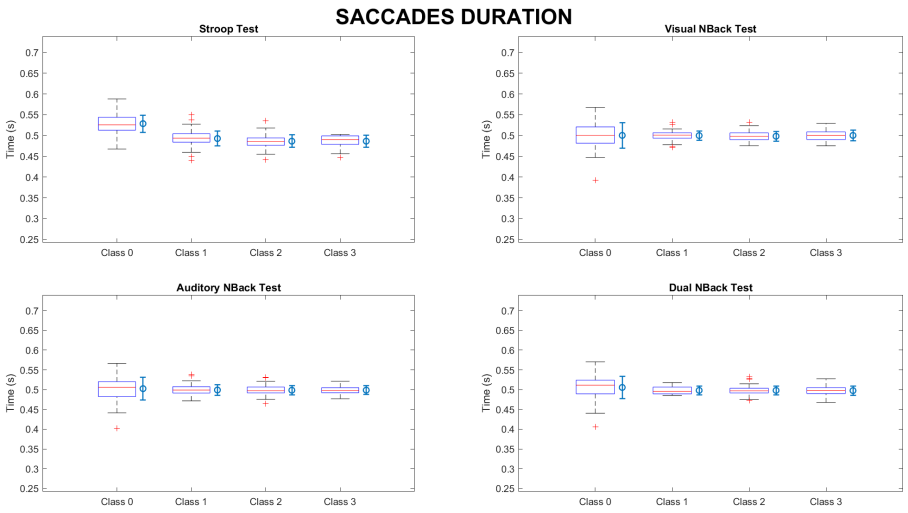

Figure B.2: Duration saccades

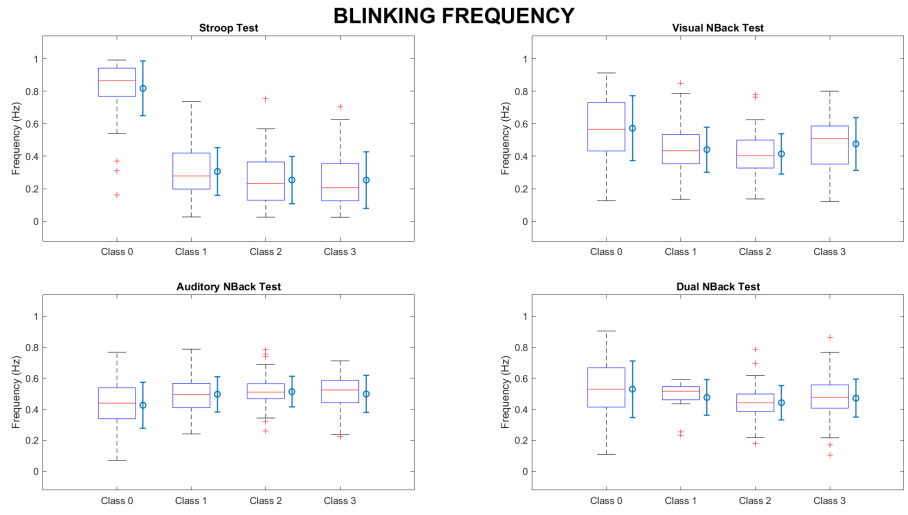

Figure B.3: Frequency blinking

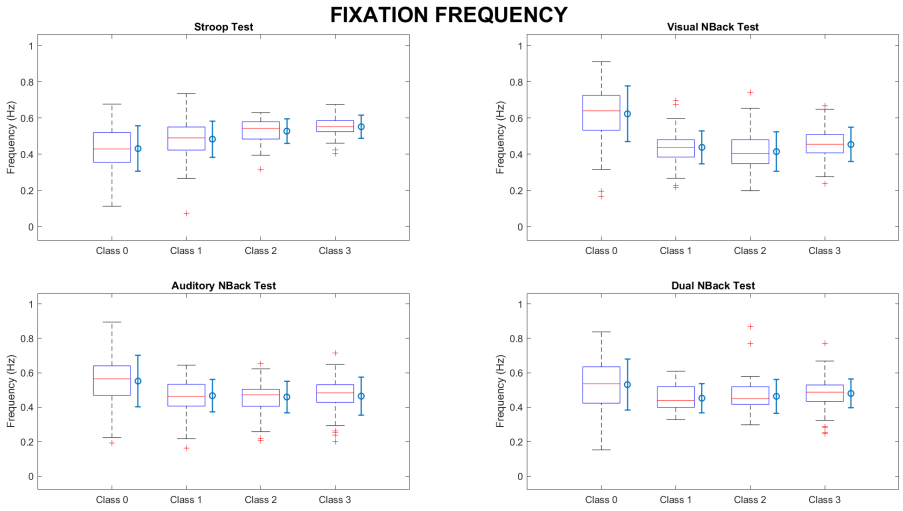

Figure B.4: Frequency fixation

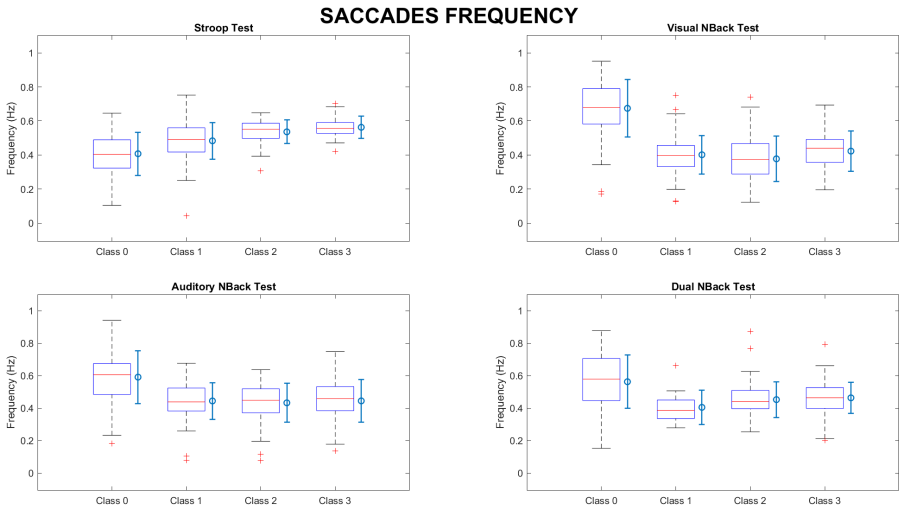

Figure B.5: Frequency saccades

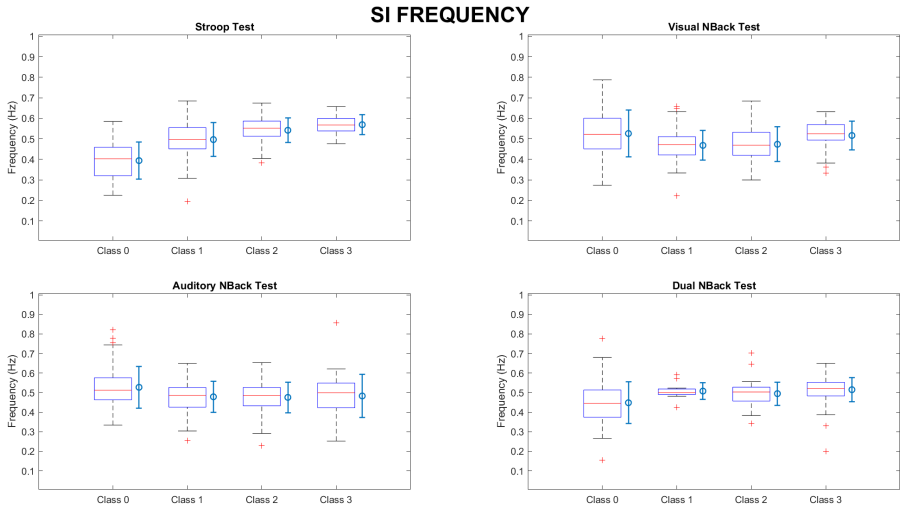

Figure B.6: Frequency SI

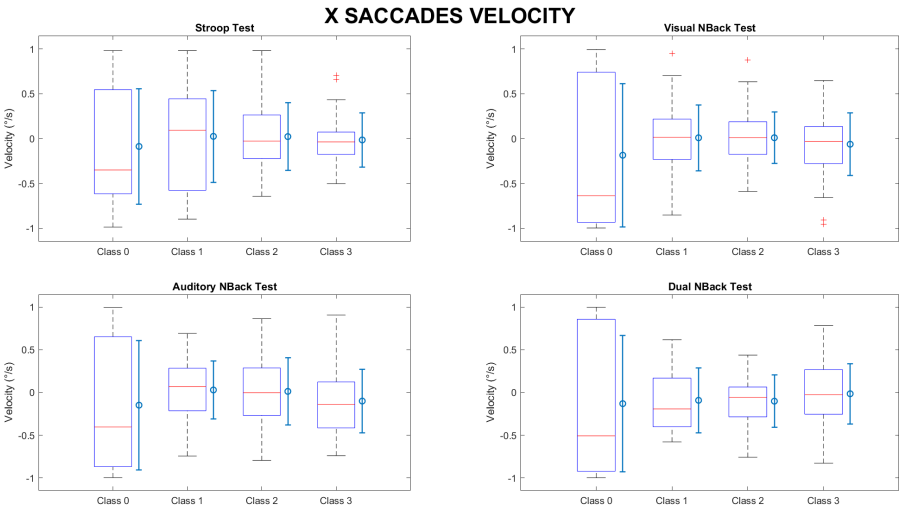

Figure B.7: X velocity

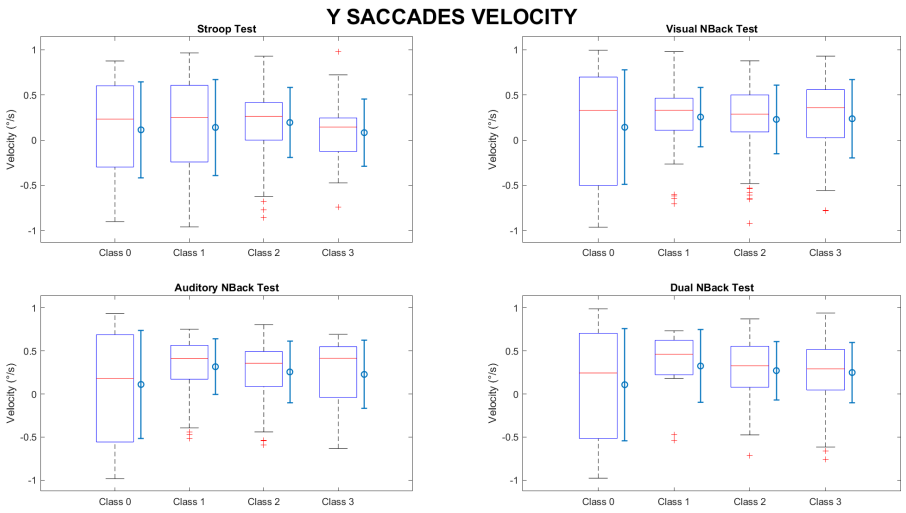

Figure B.8: Y velocity

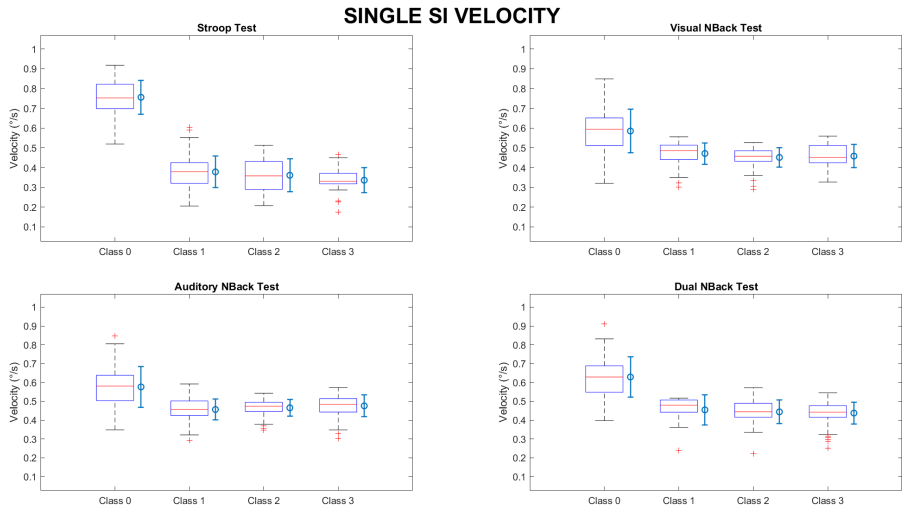

Figure B.9: Single SI velocity

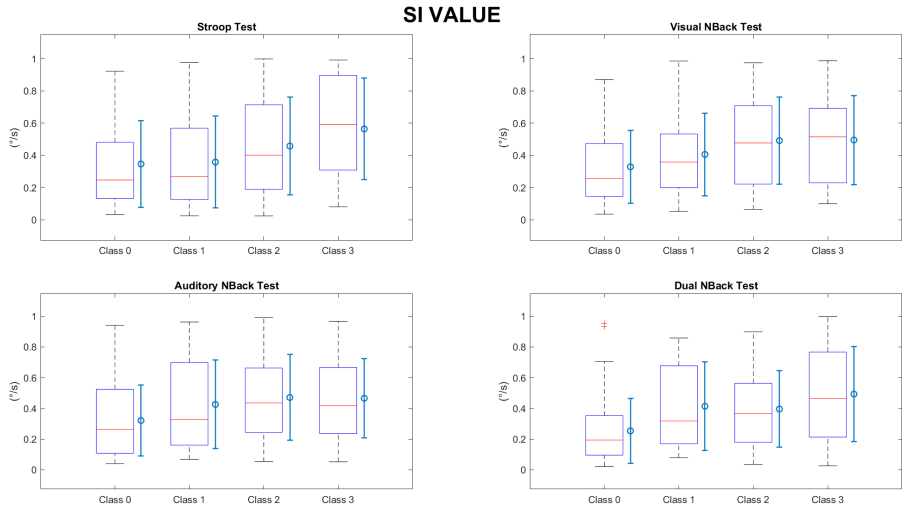

Figure B.10: SI value

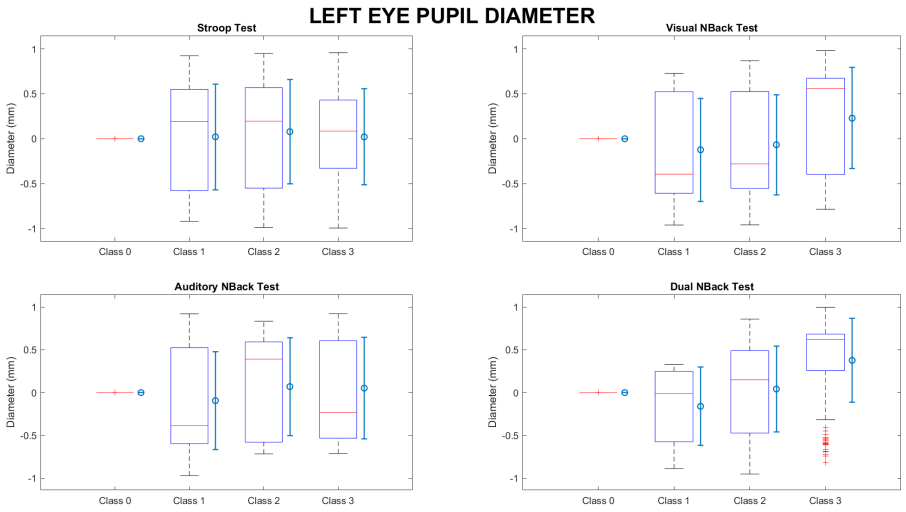

Figure B.11: Left eye relative diameter

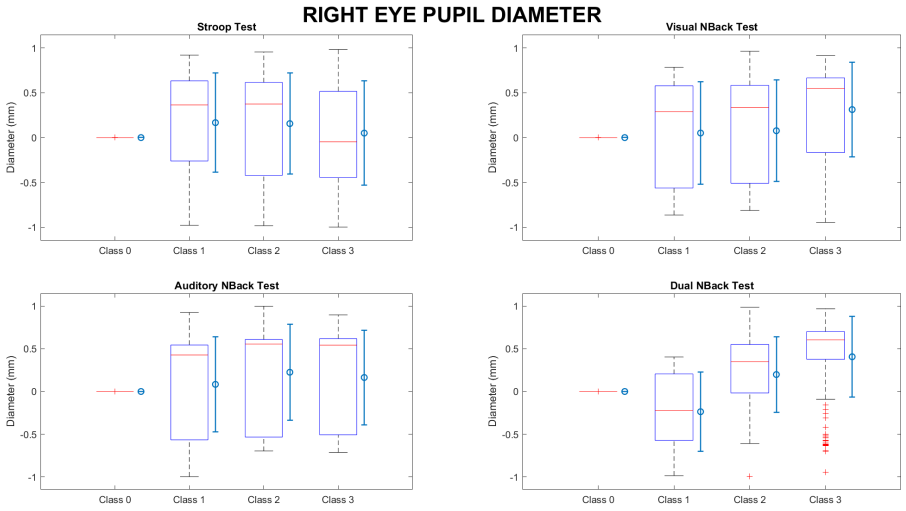

Figure B.12: Right eye relative diameter

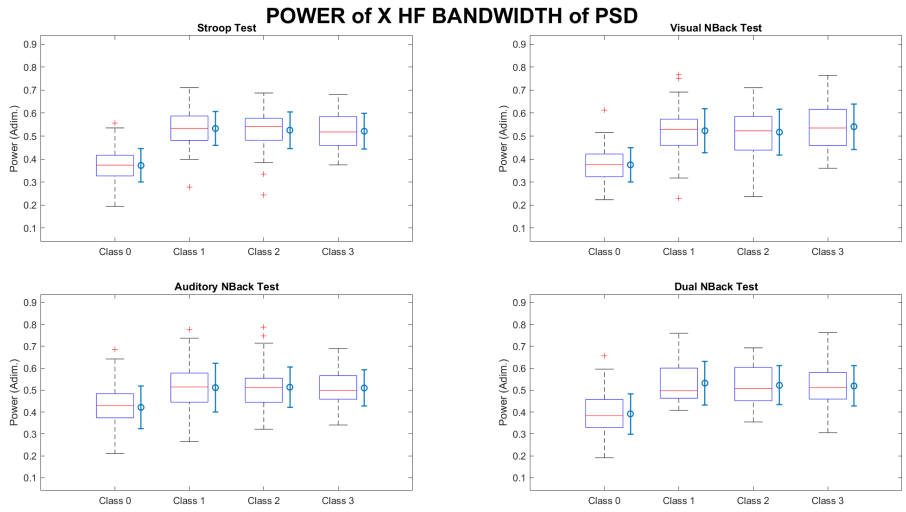

Figure B.13: X HF bandwidth

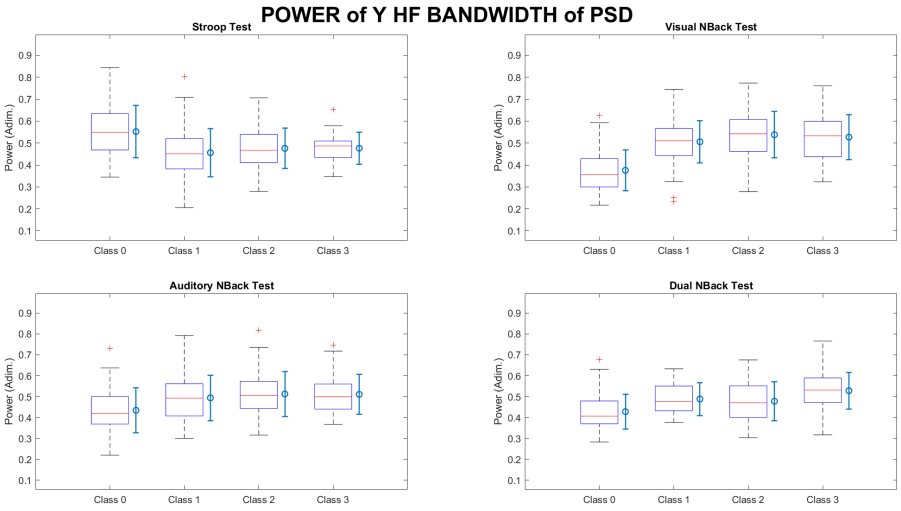

Figure B.14: Y HF bandwidth

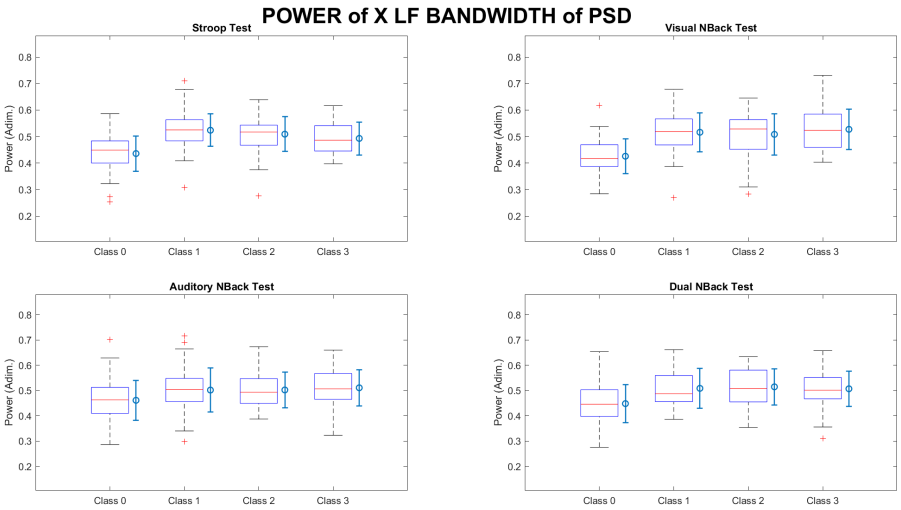

Figure B.15: X LF bandwidth

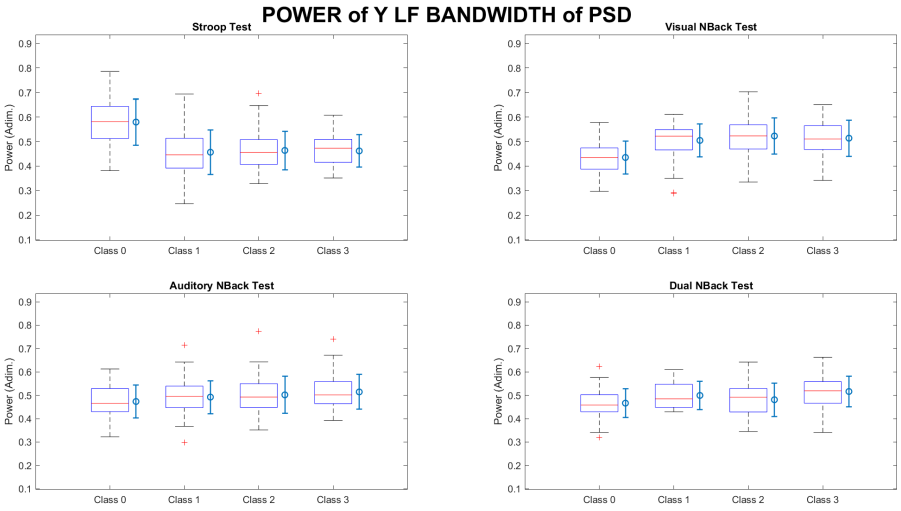

Figure B.16: Y LF bandwidth

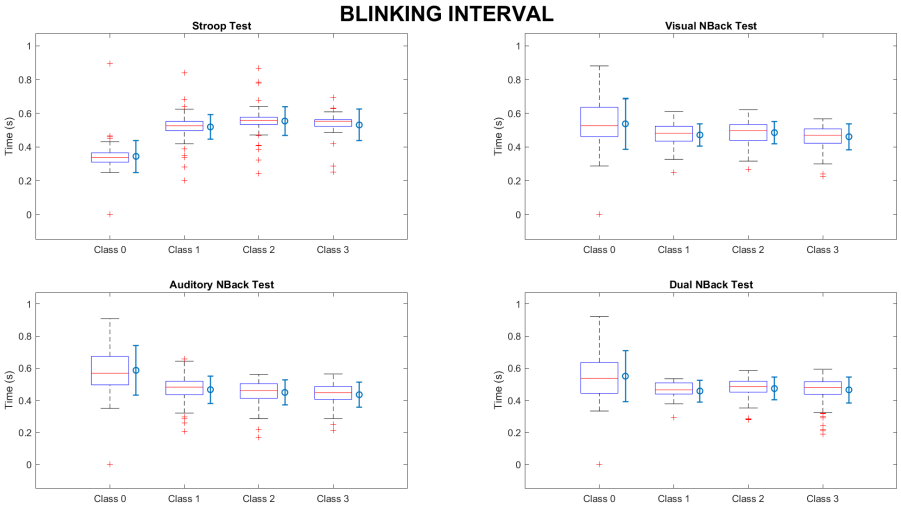

Figure B.17: Interval blinking

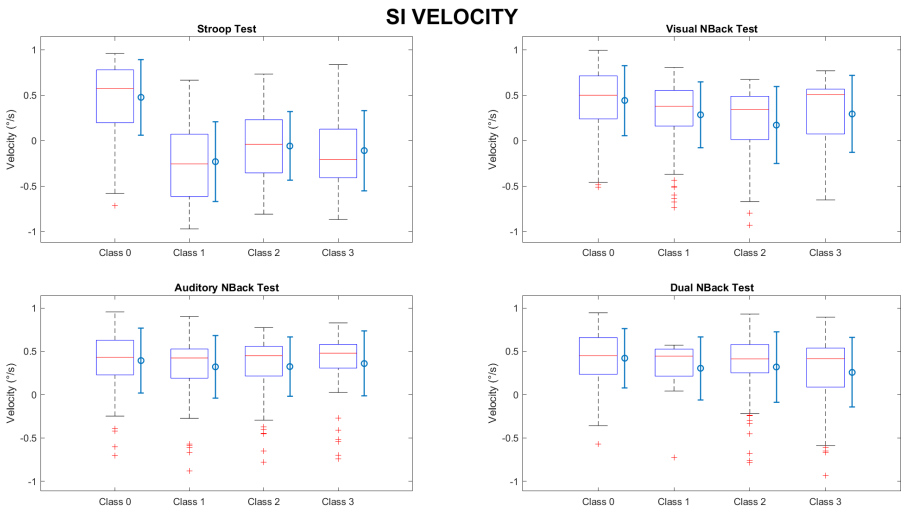

Figure B.18: SI velocity

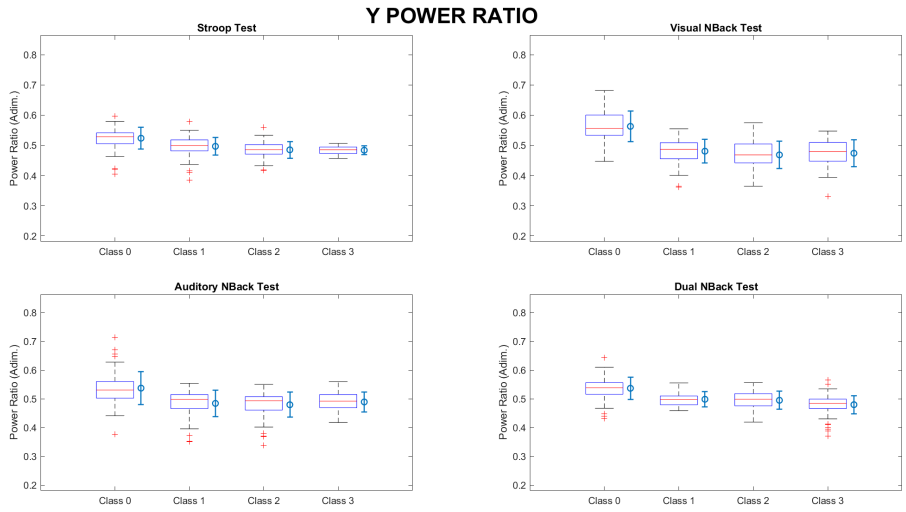

Figure B.19: Y ratio

Bibliography

[1] Weiwei Yu, Dian Jin, Xinliang Yang, Feng Zhao, Haiyang Wang, Ran Peng, et al. «Performance index based on predicted auditory reaction time analysis for the evaluation of human-machine interface in flight control». In: *Computational and mathematical methods in medicine* 2022 (2022) (cit. on p. 1).

[2] Harpreet Pal Singh and Parlad Kumar. «Developments in the human machine interface technologies and their applications: a review». In: *Journal of medical engineering & technology* 45.7 (2021), pp. 552–573 (cit. on p. 1).

[3] Bruce Mehler, Bryan Reimer, and Marin Zec. «Defining workload in the context of driver state detection and HMI evaluation». In: *Proceedings of the 4th international conference on automotive user interfaces and interactive vehicular applications*. 2012, pp. 187–191 (cit. on p. 1).

[4] Mark S Young, Karel A Brookhuis, Christopher D Wickens, and Peter A Hancock. «State of science: mental workload in ergonomics». In: *Ergonomics* 58.1 (2015), pp. 1–17 (cit. on p. 1).

[5] Changxu Wu and Yili Liu. «Development and evaluation of an ergonomic software package for predicting multiple-task human performance and mental workload in human–machine interface design and evaluation». In: *Computers & Industrial Engineering* 56.1 (2009), pp. 323–333 (cit. on p. 1).

[6] Luca Longo, Christopher D Wickens, Gabriella Hancock, and Peter A Hancock. «Human mental workload: A survey and a novel inclusive definition». In: *Frontiers in psychology* 13 (2022), p. 883321 (cit. on pp. 1, 4).

[7] Norah H Alsuraykh, Max L Wilson, Paul Tennent, and Sarah Sharples. «How stress and mental workload are connected». In: *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 2019, pp. 371–376 (cit. on pp. 1, 4, 5).

[8] Gabriele Luzzani, Irene Buraioli, Danilo Demarchi, and Giorgio Guglieri. «Preliminary Study of a Pilot Performance Monitoring System Based on Physiological Signals». In: *Aerospace Europe Conference 2023 – 10 EUCASS – 9 CEAS*. 2023, pp. 154–162. DOI: 10.13009/EUCASS2023-349 (cit. on pp. 2, 4, 5).

[9] Alžběta Brychtová, Stanislav Popelka, and V Vozenilek. «The analysis of eye movement as a tool for evaluation of maps and graphical outputs from GIS». In: *Proceedings of the 19th International Conference on Geography and Geoinformatics: Challenge for Practise and Education. Brno, Czech Republic*. 2012, pp. 154–162 (cit. on p. 2).

[10] Hans Selye. *Stress*. W/W Recording Services, 1972 (cit. on p. 5).

[11] George Fink. «Stress: Definition and history». In: *Stress science: neuroendocrinology* 3.9 (2010), pp. 3–14 (cit. on p. 5).

[12] Timothy C Barnum and Starr J Solomon. «Fight or flight: Integral emotions and violent intentions». In: *Criminology* 57.4 (2019), pp. 659–686 (cit. on p. 5).

[13] JA Veltman and AWK Gaillard. «Physiological workload reactions to increasing levels of task difficulty». In: *Ergonomics* 41.5 (1998), pp. 656–669 (cit. on pp. 6, 10).

[14] Jeffrey B Brookings, Glenn F Wilson, and Carolyne R Swain. «Psychophysiological responses to changes in workload during simulated air traffic control». In: *Biological psychology* 42.3 (1996), pp. 361–377 (cit. on pp. 6, 10).

[15] Nadja Herten, Tobias Otto, and Oliver T Wolf. «The role of eye fixation in memory enhancement under stress–an eye tracking study». In: *Neurobiology of learning and memory* 140 (2017), pp. 134–144 (cit. on p. 6).

[16] Xueli He, Lijing Wang, Xiaohua Gao, and Yingchun Chen. «The eye activity measurement of mental workload based on basic flight task». In: *IEEE 10th international conference on industrial informatics*. IEEE. 2012, pp. 502–507 (cit. on pp. 6, 11, 12).

[17] Joseph Coyne and Ciara Sibley. «Investigating the use of two low cost eye tracking systems for detecting pupillary response to changes in mental workload». In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 60. 1. SAGE Publications Sage CA: Los Angeles, CA. 2016, pp. 37–41 (cit. on pp. 6, 11).

[18] Christian Hirt, Marcel Eckard, and Andreas Kunz. «Stress generation and non-intrusive measurement in virtual environments using eye tracking». In: *Journal of Ambient Intelligence and Humanized Computing* 11 (2020), pp. 5977–5989 (cit. on pp. 6, 12).

[19] Miguel Ángel Recarte, Elisa Pérez, Ángela Conchillo, and Luis Miguel Nunes. «Mental workload and visual impairment: Differences between pupil, blink, and subjective rating». In: *The Spanish journal of psychology* 11.2 (2008), pp. 374–385 (cit. on pp. 6, 11).

[20] Bilge Pehlivanoğlu, Nezih Durmazlar, and Dicle Balkancı. «Computer adapted Stroop colour-word conflict test as a laboratory stress model». In: *JOURNAL OF CLINICAL PRACTICE AND RESEARCH* 27.2 (2005), pp. 58–63 (cit. on p. 6).

[21] Mahdi Malakoutikhah, Reza Kazemi, Hadiseh Rabiei, Moslem Alimohammadlou, Asma Zare, and Soheil Hassanipour. «Comparison of mental workload with N-Back test: A new design for NASA-task load index questionnaire». In: *International Archives of Health Sciences* 8.1 (2021), pp. 7–13 (cit. on p. 6).

[22] Stephen RH Langton, Roger J Watt, and Vicki Bruce. «Do the eyes have it? Cues to the direction of social attention». In: *Trends in cognitive sciences* 4.2 (2000), pp. 50–59 (cit. on p. 8).

[23] Mandy Ryan, Nicolas Krucien, and Frouke Hermens. «The eyes have it: Using eye tracking to inform information processing strategies in multi-attributes choices». In: *Health economics* 27.4 (2018), pp. 709–721 (cit. on p. 8).

[24] Susana Martinez-Conde. «Fixational eye movements in normal and pathological vision». In: *Progress in brain research* 154 (2006), pp. 151–176 (cit. on p. 8).

[25] Satoru Tokuda, Goro Obinata, Evan Palmer, and Alex Chaparro. «Estimation of mental workload using saccadic eye movements in a free-viewing task». In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2011, pp. 4523–4529 (cit. on pp. 8, 9, 11, 12, 39, 42, 44, 51, 52, 65, 70).

[26] Ralf Engbert. «Microsaccades: A microcosm for research on oculomotor control, attention, and visual perception». In: *Progress in brain research* 154 (2006), pp. 177–192 (cit. on p. 8).

[27] Susana Martinez-Conde, Stephen L Macknik, and David H Hubel. «The role of fixational eye movements in visual perception». In: *Nature reviews neuroscience* 5.3 (2004), pp. 229–240 (cit. on pp. 8–10, 50).

[28] Emma Gowen, RV Abadi, and E Poliakoff. «Paying attention to saccadic intrusions». In: *Cognitive Brain Research* 25.3 (2005), pp. 810–825 (cit. on p. 9).

[29] Satoru Tokuda, Evan Palmer, Edgar Merkle, and Alex Chaparro. «Using saccadic intrusions to quantify mental workload». In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 53. 12. SAGE Publications Sage CA: Los Angeles, CA. 2009, pp. 809–813 (cit. on pp. 9, 11, 39, 45, 50).

[30] Miguel A Recarte and Luis M Nunes. «Mental workload while driving: effects on visual search, discrimination, and decision making.» In: *Journal of experimental psychology: Applied* 9.2 (2003), p. 119 (cit. on p. 12).

[31] Maria Laura Mele and Stefano Federici. «Gaze and eye-tracking solutions for psychological research». In: *Cognitive processing* 13 (2012), pp. 261–265 (cit. on pp. 13, 14).

[32] Tohru Yagi. «Eye-gaze interfaces using electro-oculography (EOG)». In: *Proceedings of the 2010 workshop on Eye gaze in intelligent human machine interaction*. 2010, pp. 28–32 (cit. on p. 13).

[33] V Onkhar, D Dodou, and JCF de Winter. «Evaluating the Tobii Pro Glasses 2 and 3 in static and dynamic conditions». In: (2023) (cit. on pp. 14, 19).

[34] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. «Eye tracking for everyone». In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2176–2184 (cit. on p. 14).

[35] *Biosignalplux Professional KIT*. URL: `https://www.pluxbiosignals.com/products/professional-kit` (cit. on p. 17).

[36] *Tobii Glasses 3*. URL: `https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3` (cit. on p. 17).

[37] Chuang Zhang, JianNan Chi, ZhaoHui Zhang, XiaoLiang Gao, Tao Hu, and ZhiLiang Wang. «Gaze estimation in a gaze tracking system». In: *Science China Information Sciences* 54 (2011), pp. 2295–2306 (cit. on p. 18).

[38] *Tobii Pro Lab*. URL: `https://www.tobii.com/products/software/behavior-research-software/tobii-pro-lab` (cit. on p. 20).

[39] Eva Ostertagova, Oskar Ostertag, and Jozef Kováč. «Methodology and application of the Kruskal-Wallis test». In: *Applied mechanics and materials* 611 (2014), pp. 115–120 (cit. on p. 31).

[40] Patrick E McKnight and Julius Najab. «Mann-Whitney U Test». In: *The Corsini encyclopedia of psychology* (2010), pp. 1–1 (cit. on p. 31).

[41] Gabriel Diaz, Joseph Cooper, Dmitry Kit, and Mary Hayhoe. «Real-time recording and classification of eye movements in an immersive virtual environment». In: *Journal of vision* 13.12 (2013), pp. 5–5 (cit. on p. 32).

[42] Ignace TC Hooge, Diederick C Niehorster, Marcus Nyström, Richard Andersson, and Roy S Hessels. «Fixation classification: How to merge and select fixation candidates». In: *Behavior Research Methods* 54.6 (2022), pp. 2765–2776 (cit. on pp. 38, 44).

[43] Marcus Nyström and Kenneth Holmqvist. «An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data». In: *Behavior research methods* 42.1 (2010), pp. 188–204 (cit. on p. 38).

[44] Dario D Salvucci and Joseph H Goldberg. «Identifying fixations and saccades in eye-tracking protocols». In: *Proceedings of the 2000 symposium on Eye tracking research & applications*. 2000, pp. 71–78 (cit. on p. 38).