POLITECNICO DI TORINO

Master's Degree in Biomedical Engineering

Politecnico di Torino

Development of a closed-loop neurostimulation system for stress reduction

Supervisors

Prof. Luca Mesin

Ing. Matteo Raggi

Ing. Giovanni Chiarion

Candidate

Francesco Ambrosi

Academic Year 2023/2024

Summary

Stress is defined as a physiological response to mental, physical, or emotional stimuli and affects many people all over the world compromising their health. This study aims to develop a closed-loop neurostimulation system based on binaural beats (BB) for stress reduction exploiting heart rate variability (HRV) features and chest acceleration data collected with the Polar H10 chest strap. Twenty-two healthy participants took part in a protocol consisting of four separate acquisitions. The first was aimed at training an artificial intelligence model for real-time stress recognition, the successive two evaluated the effects of a fixed stimulation at 16 Hz and our proposal, the last session was performed to evaluate any habit effects of the protocol. During a session, each candidate performed an arithmetic test, then watched a sequence of relaxing, stressful, and relaxing videos, and performed a Stroop test. At the beginning of each session, candidates observed three minutes of rest (baseline). Six types of artificial intelligence models were trained, each using multiple feature extraction strategies. The best-trained model was the Random Forest (RF) classifier which achieved more than 84% accuracy in recognizing three levels of stress. However, in subsequent sessions, the accuracy decreased significantly (to 44%, 41%, and 49%, respectively), probably due to the normalization of the data, based on the specific baseline of each session instead of that used in the initial training. A statistical analysis of HRV features showed that the protocol was designed properly to induce stress. In contrast, the use of BBs did not show significant effects, contrary to the literature. This result may be due to the heterogeneity of the studied group and the complexity of the proposed cognitive tests. In conclusion, this study explored an innovative neurostimulation technique exploiting BBs and low-cost portable devices. Future work will allow to optimize the performance of this system extending its application also to a wider population.

Acknowledgements

I would like to express my deep gratitude to my supervisor Luca Mesin and my correlators Matteo Raggi and Giovanni Chiarion for the opportunity and advice that contributed significantly to the improvement of the work done.

I would like to thank all the people who kindly agreed to participate as subjects in the research conducted. Without their contribution, the achievement of this milestone would not have been possible.

Special thanks go to all my family members, girlfriend and friends for their constant support even in the most challenging times.

Finally, I would like to thank Politecnico di Torino for providing me with the necessary resources to conduct my research effectively.

Thank you all sincerely.

Francesco

Table of Contents

List of Tables

List of Figures

Acronyms

ACTH Adrenocorticotropic hormone

ANS Autonomic nervous system

AV Atrioventricular node

BB Binaural beats

BPM Beats per minute

CRH Corticotropin-releasing hormone

DT Decision Tree

ECG Electrocardiogram

EDA Electrodermal activity

HPA Hypothalamic–pituitary–adrenal

HRV Heart rate variability

kNN k-Nearest Neighbors

LA Left arm

LL Left leg

MLP Multilayer perceptron

PCA Principal component analysis

PCHIP Piecewise Cubic Hermite Interpolating Polynomial

PNS Parasympathetic nervous systems

PSD Power density spectrum

RA Right arm

RF Random forest

RL Right leg

SA Sinoatrial node

SNS Sympathetic nervous systems

SVM Support vector machine

Chapter 1

Introduction

Mental stress is an issue that compromises the mental and physical health of people worldwide. The increasing complexity of social, economic and occupational pressures, has led to an increase in the manifestations of stress and related disorders. Chronic stress has been linked to numerous health problems such as hypertension, coronary artery disease, high heart rate, high blood pressure, muscle tension, depression, anxiety, and irritability [1, 2, 3]. A report by EU-OSHA reported the estimated costs due to work stress-related illnesses for each state: estimates differ among countries but the costs are in the millions and even billions of euros [4]. For these reasons, measuring mental stress has become crucial to fully understand its impacts on human health and to develop effective prevention and management strategies.

Stress can be described as the body's response to physical, mental or emotional stimulus [5]. This unexpected stimulus causes the activation of a cascade of biological responses (stress response) in the brain and in the body [6]. The current gold-standard for stress assessment is the Cortisol response magnitude [6], but in [7] has been found that an indicator based on HRV (Heart Rate Variability) is consistent with the results obtained from Cortisol response magnitude. In addition, HRV has the advantage of being very easy to record, being noninvasive and being able to obtain it from wearable devices.

This master's thesis aims to develop a method for real-time stress detection using HRV and accelerometer data, and to react through binaural beats (BB) modulated stimulation with the aim of stress reduction. Since the most widely used stress detection methods in the literature exploit artificial intelligence [8], algorithms such as Support vector machines (SVM), Random forests (RF), and Multilayer perceptron (MLP) will also be explored in this thesis.

1.1 Heart Rate Variability (HRV)

1.1.1 Heart

The heart is a vital organ that acts as a pump for the circulatory system, ensuring the circulation of blood throughout the body. It is located in the mediastinum, i.e. the central compartment of the thoracic cavity. Within this organ there are myocardial cells, a type of muscle cell that can generate and conduct electrical impulses, enabling the heart to contract in a coordinated and rhythmic manner. Cardiac activity, therefore, does not depend directly on stimulation by the nervous system, however, the autonomic nervous system is able to regulate it, modulating the frequency and the force of contraction of the heart muscle [9].

The heart is divided into four chambers: the right and left atrium, located at the top, which receive blood from the veins; the right and left ventricle, located at the bottom, which pump blood into the arteries. The heartbeat consists in the rhythmic contraction (systole) and relaxation (diastole) of the entire muscle mass of the heart. Electrical activity begins in the sinoatrial node (SA) and propagates from cell to cell, contracting first the atria and then the ventricles [9]. See Fig. 1.1.

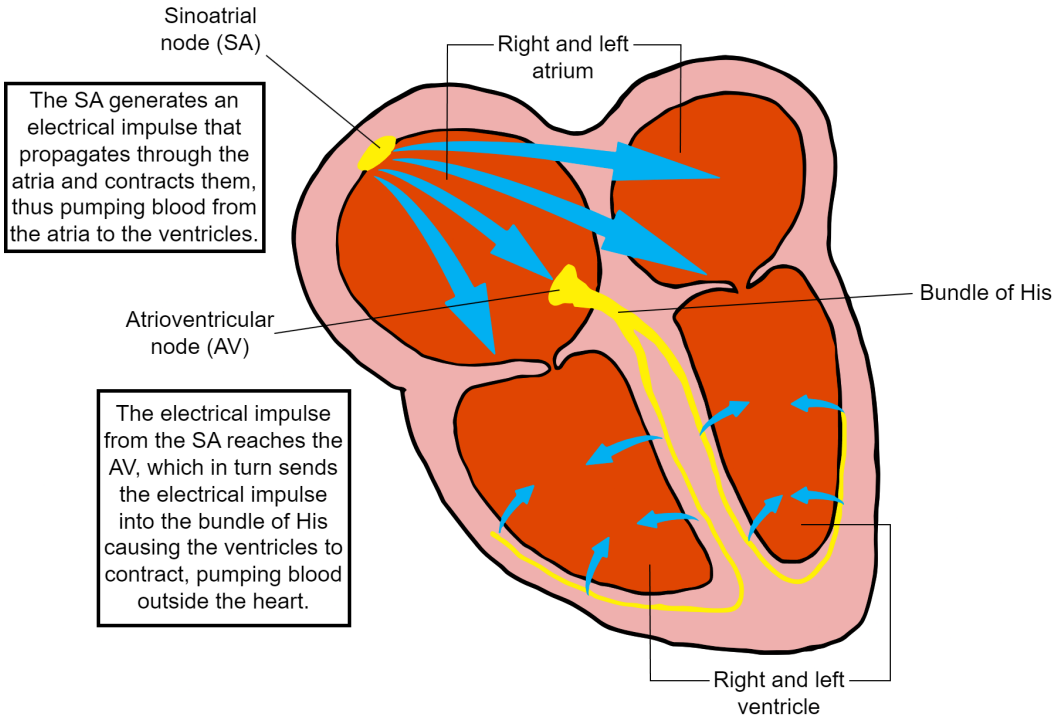

Figure 1.1: Heart's electrical conduction system.

The depolarization and the following repolarization of myocardial cells generates electric fields that are able to propagate from the heart to the epidermis. The electrocardiogram (ECG) is a noninvasive cardiological examination that allows to record this electrical activity by placing electrodes on the subject's skin. Analysis of the ECG tracing provides valuable information about cardiac function.

1.1.2 Electrocardiogram (ECG)

A "standard" ECG includes 12 different leads, each describing the same heartbeat at the same instant, but from different viewpoints. 12-lead ECG is obtained by connecting 10 electrodes to the patient: 1 electrode on each limb (Right Arm (RA), Left Arm (LA), Right Leg (RL), Left Leg (LL)) and 6 electrodes on the chest (V_1, V_2, V_3, V_4, V_5, V_6) [10].

There are three limb leads, see Fig. 1.2 (first row):

1. **Lead I**: measured between the positive electrode on the left arm (LA) and the negative electrode on the right arm (RA).

$$I = LA - RA$$

2. **Lead II**: measured between the positive electrode on the left leg (LL) and the negative electrode on the right arm (RA).

$$II = LL - RA$$

3. **Lead III**: measured between the positive electrode on the left leg (LL) and the negative electrode on the left arm (LA).

$$III = LL - LA$$

These 3 limb leads form the Einthoven's triangle, named after Willem Einthoven who invented the first practical ECG [10].

There are three Augmented limb leads, see Fig. 1.2 (second row):

4. **Lead augmented vector right (aVR)**: measured between the positive electrode on RA and the negative pole that is a combination of LA and LL.

$$aVR = RA - \frac{1}{2}(LA + LL)$$

5. **Lead augmented vector left (aVL)**: measured between the positive electrode on LA and the negative pole that is a combination of RA and LL.

$$aVL = LA - \frac{1}{2}(RA + LL)$$

6. **Lead augmented vector foot (aVF)**: measured between the positive electrode on LL and the negative pole that is a combination of RA and LA.

$$aVF = LL - \frac{1}{2}(RA + LA)$$

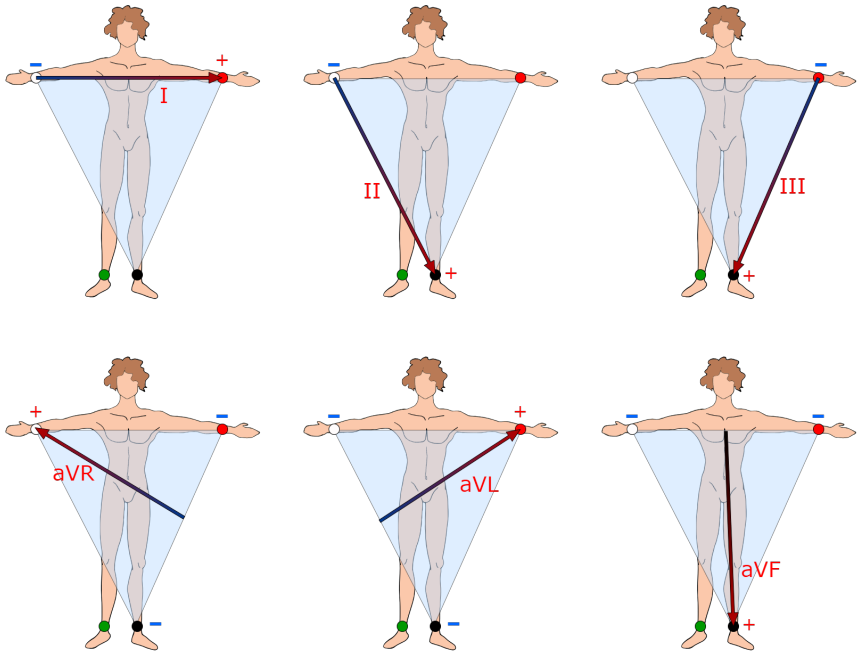

Figure 1.2: Limb leads and Augmented Limb Leads.

Six **Precordial Leads** (V_1, V_2, V_3, V_4, V_5, V_6), see Fig. 1.3, where V_i is the positive electrode and the common virtual electrode also known as Wilson's central terminal (V_w) is the negative pole.

$$V_W = \frac{1}{3}(RA + LA + LL)$$

The limb leads and Augmented limb leads are used to study the sagittal plane of the heart, while the precordial leads allow the study of the transverse plane of the heart. It is important to notice that there is also an electrode on the Right Leg

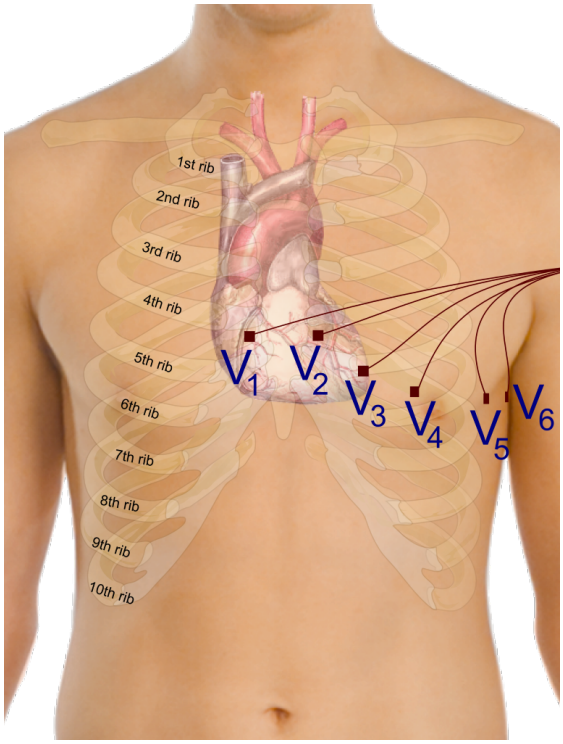

Figure 1.3: Precordial leads placement. *(Mikael Häggström, CC0, via Wikimedia Commons)*

(RL), but it is used as a grounding lead that helps minimize ECG artifact, therefore it is not present in ECG reading [10]. In the following I will refer exclusively to the limb lead I signal.

One heartbeat creates an ECG signal like in Fig. 1.4. All the waves in the ECG have a physiological meaning:

- **P wave**: it is the first wave of the heartbeat, it represents depolarization of the atria after SA node activation.
- **PR interval**: it is the time the electrical impulse takes to travel from the SA through the AV node.
- **QRS complex**: represents the rapid depolarization of the right and left ventricles. Since ventricle muscles are stronger than atria muscles, QRS complex usually has a much larger amplitude than the P wave.
- **ST segment**: it is the period when the ventricles are depolarized.
- **T wave**: it represents the repolarization of the ventricles.

Atria repolarization is not visible in the tracing as it is obscured by the depolarization of the ventricles.

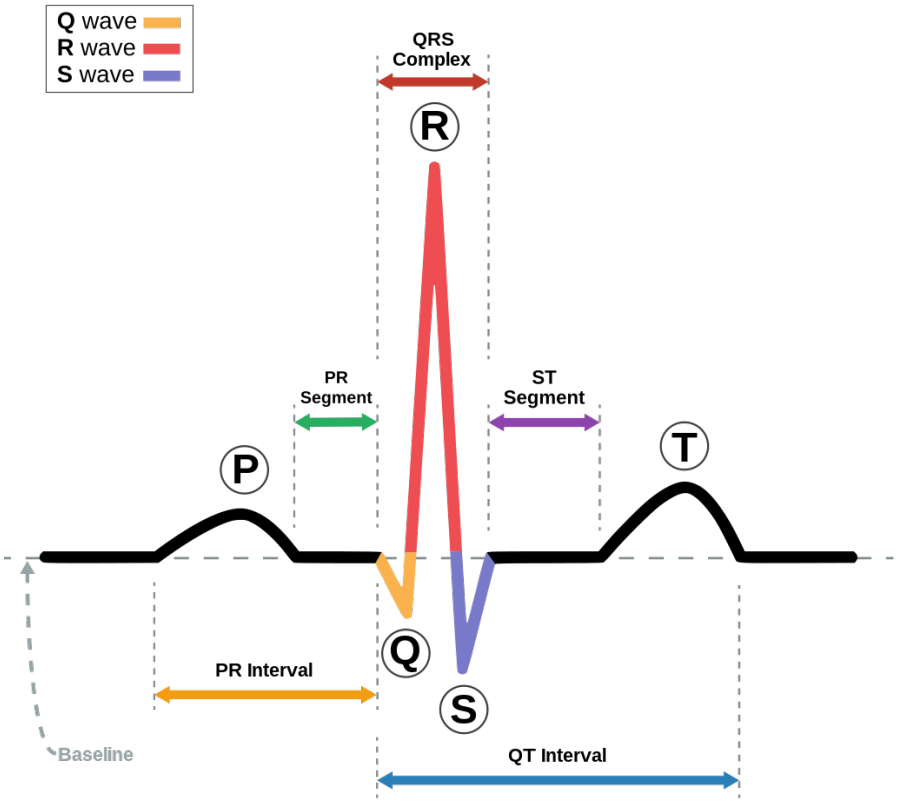

Figure 1.4: ECG of a heartbeat in normal sinus rhythm. *(Created by Agateller (Anthony Atkielski), converted to svg by atom., Public domain, via Wikimedia Commons)*

1.1.3 RR intervals

Heart Rate Variability is the variation in time interval between heartbeats [11] and it can be measured from the ECG. First the R peaks of each heartbeat are identified, either by algorithms such as Pan-Tompkins [12] or wavelet transforms [13] or by visual inspection. Once the R peaks are detected, RR intervals, i.e. the time intervals between successive R peaks, are calculated. Using the RR intervals, the tachogram is created: RR intervals series vs progressive number of beats [11]. RR intervals can also be referred to as NN (Normal to normal) intervals, i.e. intervals

between QRS complexes resulting from sinus node depolarizations. Example of tachogram in Fig. 1.5.

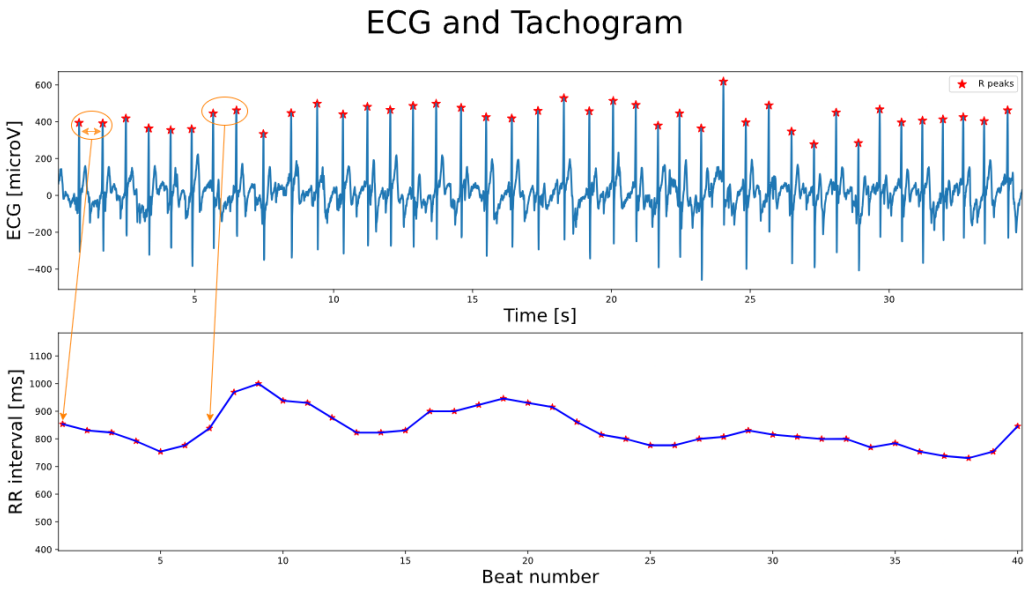

Figure 1.5: ECG and RR Tachogram example.

1.1.4 HRV measures

To study HRV, several parameters can be calculated from the tachogram. The 2 main methods of HRV analysis are:

- **Time Domain Methods** [11]: derived directly from the analysis of RR intervals in the tachogram. Most used parameters are described in Table 1.1.
- **Frequency Domain Methods** [11]: they derive from power density spectrum (PSD) analysis, which can be obtained by nonparametric methods (like FFT, simpler approach) or parametric methods (smoother spectra, but more complex to use), as shown in Fig. 1.6. Most used parameters are described in Table 1.2.

To properly analyze HRV using frequency domain parameters, it is necessary that the recording lasts 10 times the wavelength of the lowest frequency to be investigated. Thus, to assess HF components (lower frequency 0.15 Hz) is needed at least 1 min of recording ($1/0.15 * 10 \approx 67$ s), even more for LF components [11].

Variable	Description
$SDNN$	Standard deviation of RR intervals
$SDANN$	Standard deviation of the average NN intervals calculated over short periods, usually 5 minutes
$RMSSD$	Root mean square of successive differences, i.e. differences between adjacent NN intervals
$SDSD$	Standard deviation of successive differences
$NN50$	The number of pairs of successive NNs that differ by more than 50 ms
$pNN50$	The proportion of NN50 divided by total number of RR intervals
$NN20$	The number of pairs of successive NNs that differ by more than 20 ms
$pNN20$	The proportion of NN20 divided by total number of RR intervals

Table 1.1: HRV parameters for Time Domain methods.

Variable	Description	Frequency range
VLF	Power in very low frequency range	$0 \leftrightarrow 0.04$ Hz
LF	Power in low frequency range	$0.04 \leftrightarrow 0.15$ Hz
HF	Power in high frequency range	$0.15 \leftrightarrow 0.4$ Hz
Variable	**Description**	**Formula**
LF_{norm}	LF power in normalised units	$\frac{LF}{total\ power - VLF} * 100$
HF_{norm}	HF power in normalised units	$\frac{HF}{(total\ power - VLF)} * 100$
LF/HF	Ratio LF/HF	

Table 1.2: HRV parameters for Frequency Domain methods.

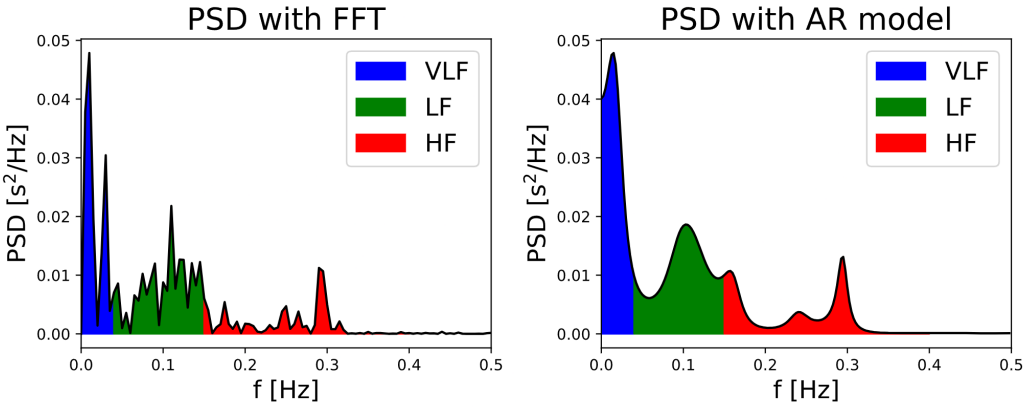

Figure 1.6: PSD estimation with FFT (non parametric) and order 20 autoregressive model (parametric).

1.1.5 How HRV is linked to stress?

HRV is the temporal variation between heartbeats, a phenomenon underlying heart rate that reflects the balance between the sympathetic (SNS) and parasympathetic (PNS) nervous systems [14]. The biological stress response is characterized by the synchronized activity of the autonomic nervous system (ANS) and hypothalamic–pituitary–adrenal (HPA) axis.

Stimulation of the ANS immediately activates the sympathetic nervous system branch, resulting in the releasing of epinephrine and norepinephrine [11] that cause increased heart rate, blood pressure, and secretion of salivary alpha-amylase. Activation of the slower acting HPA axis results in the secretion of corticotropin-releasing hormone (CRH), adrenocorticotropic hormone (ACTH), and cortisol [15]. Instead, the parasympathetic influence on heart rate is mediated via release of acetylcholine by the vagus nerve [11].

HRV thus represents the ability of the heart to respond to physiological and environmental stimuli. A low HRV is due to a very regular heartbeat and is associated with impaired regulatory functions of the ANS, indicating low adaptive capacity. In contrast, high HRV is associated with increased activity of the PNS and reduced activity of the SNS, which is cardioprotective [16, 17]. To summarize, during a stressful stimulus, heart rate increases, i.e. RR intervals decrease, and HRV decreases leading to lower values of SDNN and RMSSD, decreasing HF-band power, increasing LF-band power. Normative data of some HRV parameters [18, 19] are shown in table 1.3.

Parameter	Range
RR interval	650 $\leftrightarrow$ 1000 ms
SDNN	32 $\leftrightarrow$ 93 ms
RMSSD	19 $\leftrightarrow$ 75 ms
LF	193 $\leftrightarrow$ 1009 ms^2
HF	83 $\leftrightarrow$ 3630 ms^2

Table 1.3: Normative data of RR interval, SDNN, RMSSD, LF and HF power.

1.2 Accelerometer

An accelerometer is a sensing device used to measure acceleration, i.e. the change in the velocity of an object as a function of time. This makes them useful for monitoring the motion of an object. The working principle of an accelerometer is based on the use of a mass inside a device, connected through a spring to the outer case (Fig. 1.7).

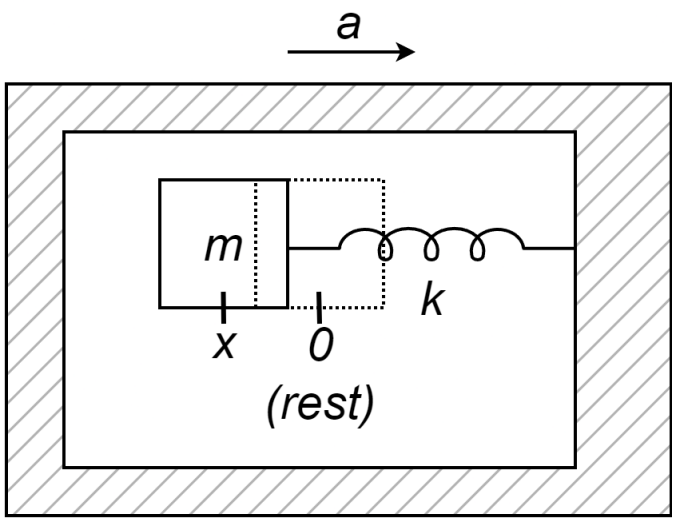

Figure 1.7: Working principle of an accelerometer.

The external case will be applied to the object whose acceleration has to be studied. When the object moves, the mass inside the accelerometer is subjected to an inertia force that will oppose the movement (1.1a). The inertia is balanced by the force exerted by the deformed spring (1.1b), therefore obtaining (1.2).

$$F = ma \tag{1.1a}$$

$$F = kx \tag{1.1b}$$

$$ma = kx \tag{1.2}$$

From (1.2) it is possible to derive (1.3), thus by knowing k and m and measuring x (the deformation of the spring), it is possible to obtain the acceleration of the object.

$$a = \frac{k}{m}x \tag{1.3}$$

The system described is only able to detect acceleration along one axis, but to obtain the acceleration in three-dimensional space it is possible to replicate the same device 3 times, placing each device along 3 perpendicular spatial directions (Fig. 1.8).

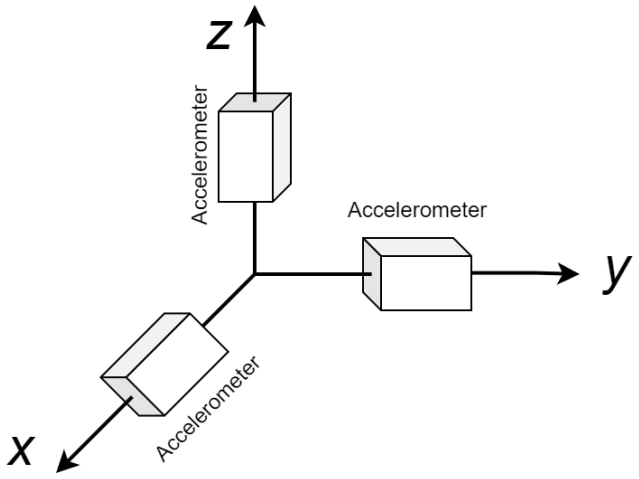

Figure 1.8: Example of a triaxial accelerometer.

For stress detection may be useful to measure acceleration of the subject chest in order to obtain:

- chest movement with respect to the body trunk, i.e. subject's breathing rate [20];
- body trunk movement.

1.3 Binaural Beats

When two sinusoids of constant amplitude but slightly different frequencies are added together, they interfere creating a new wave whose amplitude varies over time with a frequency equal to the frequency difference between the sinusoids Δf

(Fig. 1.9). When this sound is heard through ears, the perception is to hear beats at Δf frequency, and those are called "Acoustic beats" [21].

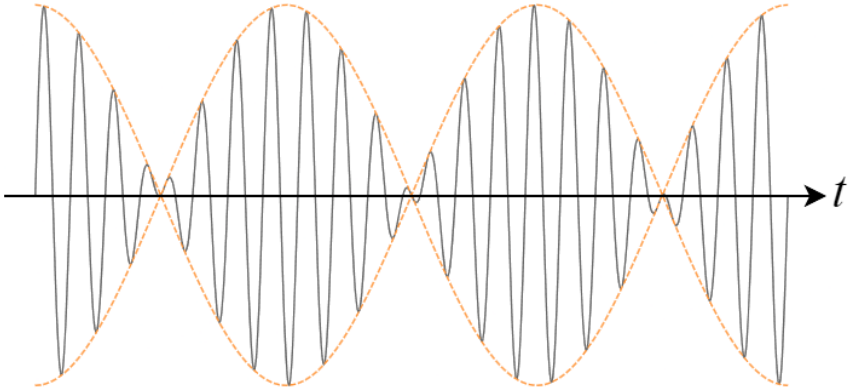

Figure 1.9: Example of an acoustic beat.

Instead, when a sinusoid with constant amplitude and frequency f_0 is presented to one ear, and another sinusoid with constant amplitude and frequency $f_0 + \Delta f$ is presented to the other ear, even though both sinusoids are steady there is a perception of amplitude modulated beats, with frequency Δf located in the center of the head. Those beats are called "Binaural Beats" (BB) [21], as shown in Fig. 1.10. The reason why Binaural Beats can be perceived is due to auditory pathways. It shows that the discharge of the neurons in the auditory nerve preserve the acoustic wave phase information, and in the point where auditory system from both ears are combined, the signals are combined together with their respective phase [22]. In [23] has been found that Binaural beats are present when the stimulating tones frequency are less than 1000 Hz and Δf is less than 40 Hz.

The electroencephalogram shows 5 main frequency bands for neural activity: delta band ($<$ 4 Hz), theta band (4 - 8 Hz), alpha band (8 - 13 Hz), beta band (13 - 30 Hz), gamma band ($>$ 30 Hz). BB waves can be used to stimulate neuronal activity of a specific band by exploiting the Δf frequency of the beats. A few examples in [24, 25, 16].

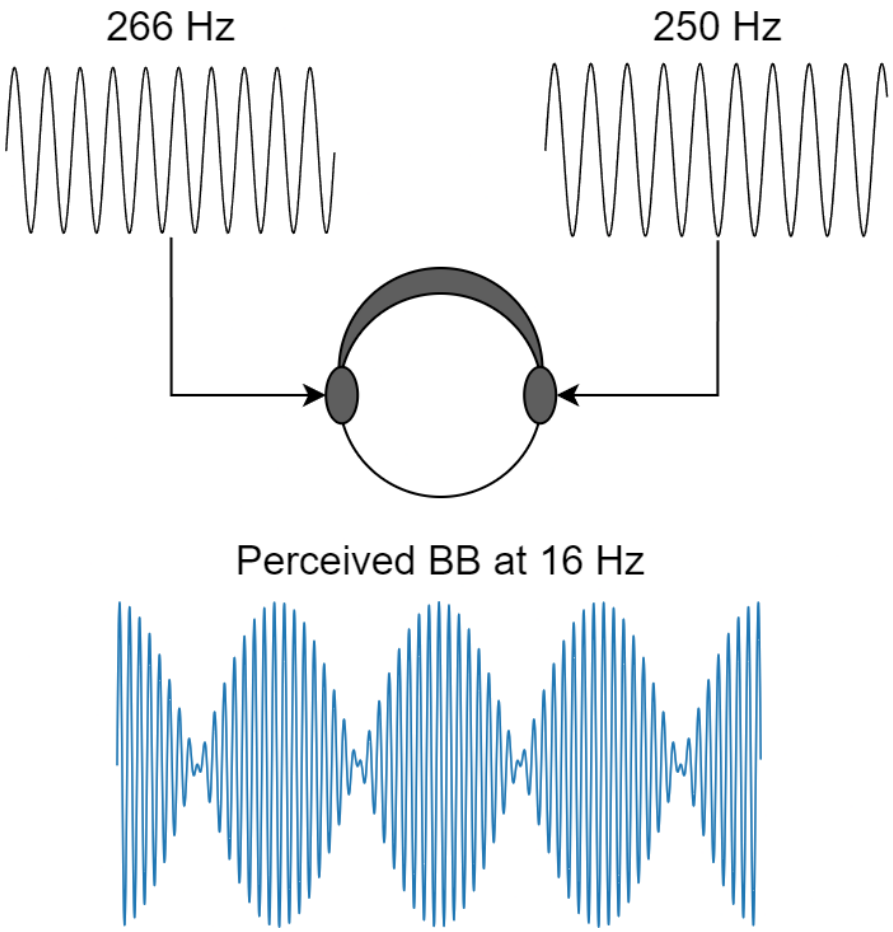

Figure 1.10: Example of Binaural Beats.

1.4 Machine Learning Algorithms

In this section [26] is used as a reference. Machine learning is a field of study in artificial intelligence concerned with developing methods to enable computers to learn without being explicitly programmed. Classification and regression are two common task in machine learning. Classification is the problem of identifying to which category (or class) an observation belongs to. Regression is the estimating of the relationships between a dependent variable and one or more independent variables. In our case observations or independent variables will be referred to as "features", while the dependent variable or category will be referred to as "label".

Machine learning approaches are usually divided into three categories:

- **Supervised learning**: is the task of learning a general rule that maps an input (features) to an output (labels) based on numerous examples of input-output

pairs.

- **Unsupervised learning**: unlike supervised learning, no labels are given to the learning algorithm. Algorithms are free to discover data structures data on their own.
- **Reinforcement learning**: the algorithm interacts with a dynamic environment in which it must perform a certain goal. As it navigates its problem space, the algorithm receives feedback, analogous to rewards, which it tries to maximize.

There are several algorithms that enable machine learning. In this thesis 3 supervised learning algorithms are explored: Support Vector Machine, Random Forest and Multilayer Perceptron.

1.4.1 Support Vector Machines

The following information come mainly from [27]. Support Vector Machines (SVM) are supervised learning models able to perform classification and regression. In order to understand the basics, let's consider a 2 class classification problem. SVM maps training examples to points in N-dimensional space, where N is the number of feature used as input, and its goal is to separate such points with an hyperplane, that is a linear surface whose dimension is N-1. For this reason SVMs are considered linear classifiers. Teoretically there are many hyperplanes that might classify the data but a reasonable choice is the one that represents the largest separation, or margin, between the two classes, i.e. the hyperplane that maximizes the distance between itself and the nearest data points on each side. An example in Fig. 1.11.

It may happen that the sets to discriminate are not linearly separable in that space. The solution proposed is the "kernel trick": a mathematical trick that takes advantage of a kernel function which projects data from a low-dimensional space to a space of higher dimension [28]. Simply put, it is like adding extra features obtained from combination of starting feature, thus increasing the dimensionality of the space. This allows to separate data in a non-linear way in the original space. An example in Fig. 1.12.

The SVM algorithm used in this thesis is the one implemented within the scikit-learn library for Python [29, 30].

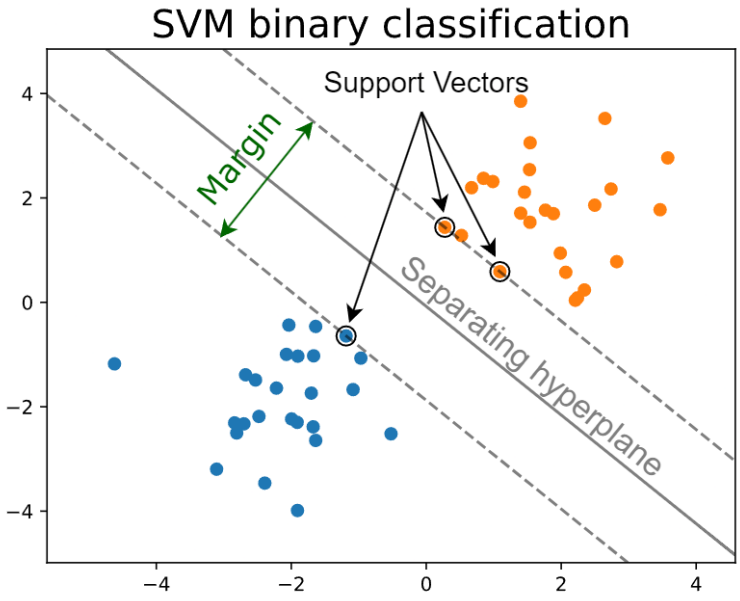

Figure 1.11: SVM binary classification example.

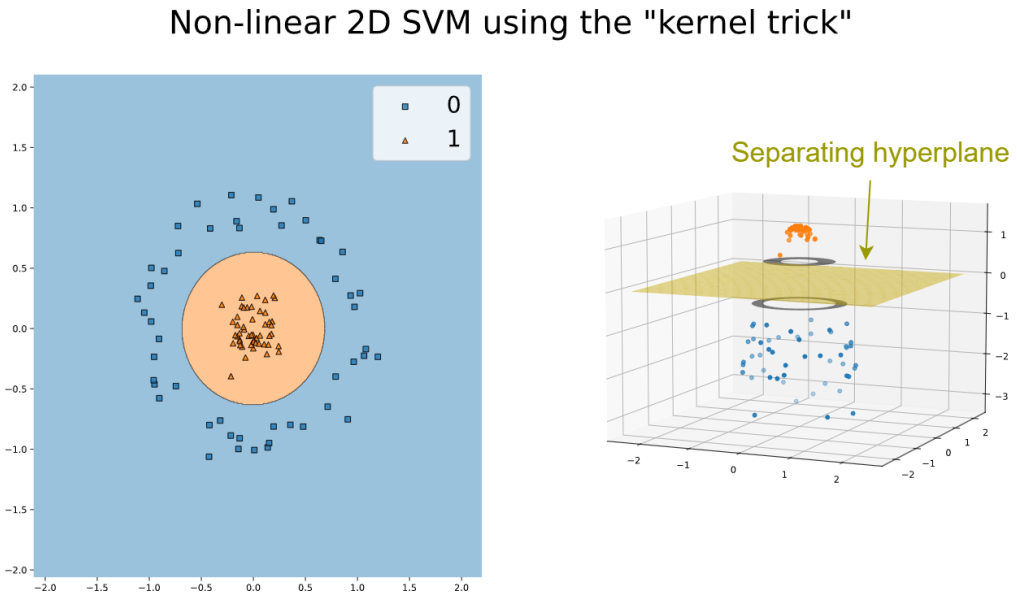

Figure 1.12: Example of kernel trick on a 2D non-linear classification problem.

1.4.2 Random Forest

Before understanding random forest it is important to understand what is a Decision Tree (DT). Decision tree is a supervised learning approach used for classification and regression. A general structure of a DT is shown in Fig. 1.13.

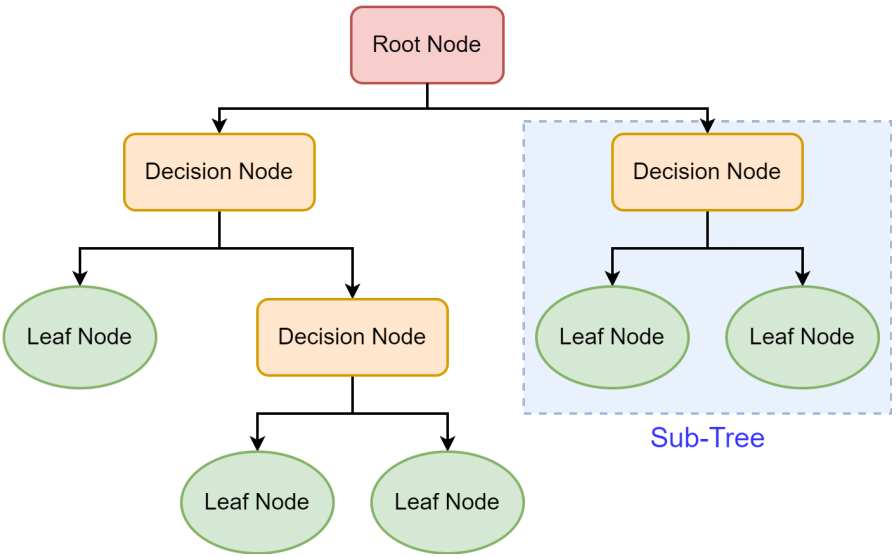

Figure 1.13: Structure of a Decision Tree.

A tree is built by splitting the dataset, represented by the root node, into subsets, thus creating the successor children. This process is repeated on each derived subset in a recursive manner (Decision Node). The recursion ends when the subset at a node has all the same values of the target label or when splitting no longer improve prediction performance (Leaf Node). There are several splitting rules based on different metrics and the best one may vary from case to case. Some metrics examples are: Gini impurity, Information gain, Variance reduction [31].

Decision trees that are grown very deep tend to overfit their training sets. It is possible to overcome this issue by using random forest. Random Forests (RF) or random decision forests is an ensemble learning method that can be used for classification and regression. An ensemble learning method use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone [32]. For classification tasks, the output of the random forest is the class selected by most trees, while for regression tasks, the mean prediction of the individual trees is returned. Random forest is thus a set of decision trees, hence the name "forest". While "Random" is what really takes care of the overfitting issue. Two source of randomness are used:

- **Bagging** (or bootstrap aggregating): consists of giving to each tree of the

forest only a subset of the original dataset. Random sampling with replacement is used, thus allowing to select multiple times the same sample while creating the subset. Then each tree is trained and the results are aggregated.

- **Random features subset**: each tree uses a different subset of the features used in the dataset.

This randomness is what ensures that the trees are highly differentiated from one another [32]. A single tree can overfit on its subset, but the whole forest won't.

The RF algorithm used in this thesis is the one implemented within the scikit-learn library for Python [29, 30].

1.4.3 Multilayer Perceptron

A multilayer perceptron (MLP) is a feedforward artificial neural network, consisting of a system of simple interconnected neurons, as illustrated in Fig. 1.14. The nodes (neurons) of each layer are fully connected with each node of the next layer. Each connection has its weight. Within a neuron, the weighted inputs are added together and an activation function is applied to the sum. The result obtained is the output of the neuron (see Fig. 1.15) [33]. The "Activation function" is the function that links the input of neurons to their output. Using simple non-linear function on lots of neurons enables the multilayer perceptron to approximate extremely non-linear functions [33]. Most commonly used activation function are "Sigmoid" and "ReLU" (Fig. 1.16).

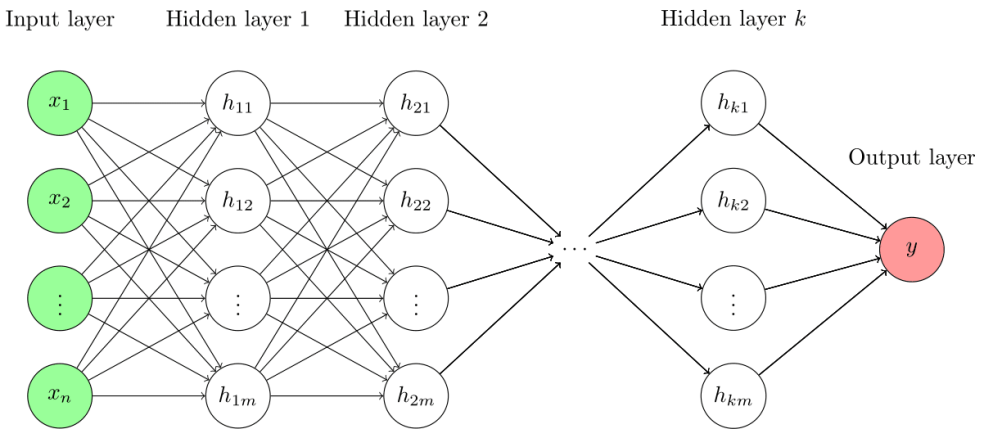

Figure 1.14: Structure of the MLP with k hidden layers.

Modern feedforward networks are trained using the backpropagation method [33]. In supervised learning the learning process is possible by changing connection

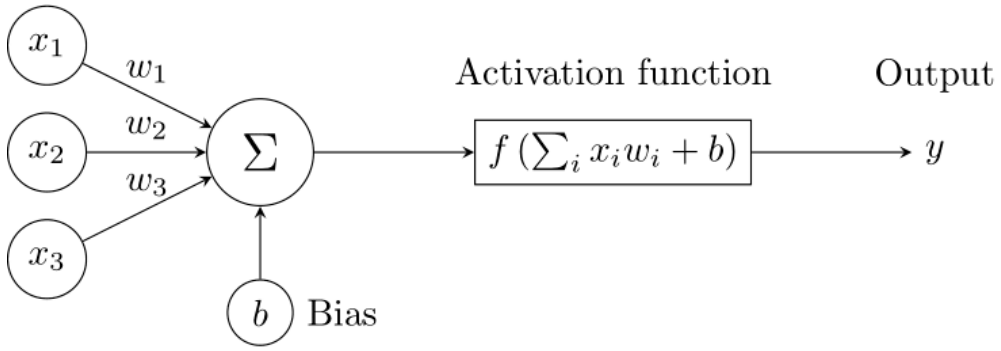

Figure 1.15: MLP neuron structure.

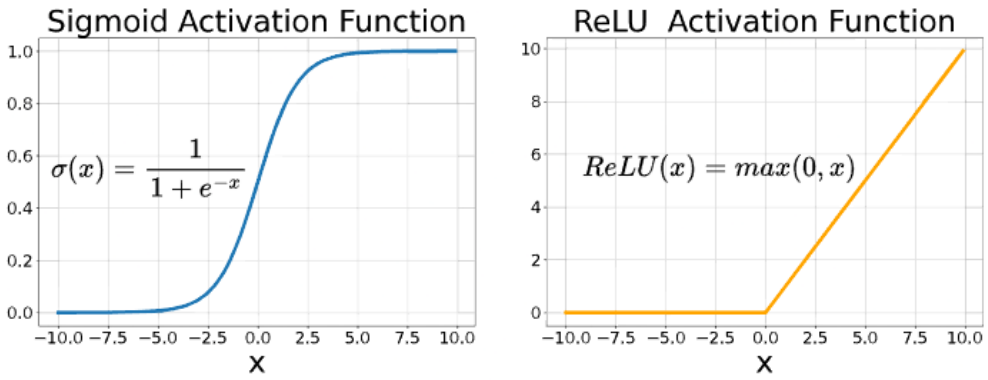

Figure 1.16: Sigmoid and ReLU activation functions for MLP neuron.

weights of every neuron after each piece of data is processed, based on the amount of error in the output compared to the expected result, as illustrated in 1.17.

The MLP algorithm used in this thesis is the one implemented within the scikit-learn library for Python [29, 30].

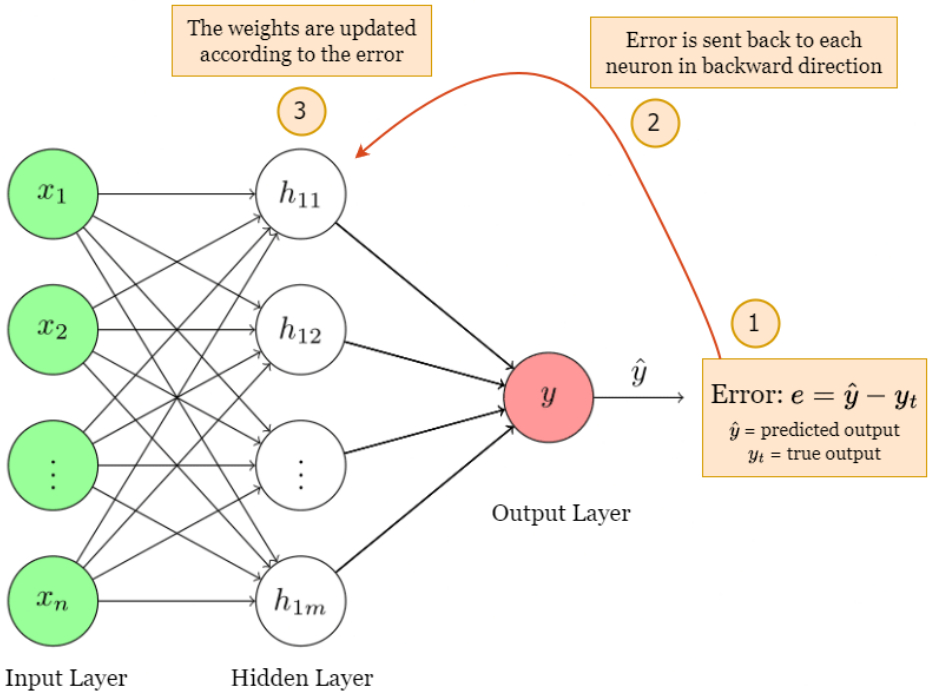

Figure 1.17: Backpropagation in MLP.

1.5 Overview of Previous Studies

The scientific literature has extensively explored the topic of stress through the analysis of HRV as a physiological indicator. The aim of this thesis is to broaden its understanding by examining the effects of BB modulation on stress. The review conducted by Vos et al. [8] highlights the presence of numerous scientific articles that identify stress in a binary manner using machine learning techniques such as SVM, RF, MLP, k-Nearest Neighbors (kNN), and Principal component analysis (PCA), after signal acquisition. Regarding BB stimulation, studies by Katmah et al. [5] and Al-Shargie et al. [24] are of interest, as they observed reduced mental stress and improved performance in cognitive tasks through the use of BB at a fixed frequency and volume. The work of Choi et al. [34] deserves special attention, as they developed a system based on wearable sensors for the simultaneous collection of various physiological signals, including HRV and Electrodermal Activity (EDA). The experimental protocol used, similar to the one adopted in this research, involves the application of various stressful tasks such as the Stroop test, public speaking and memory search, alternating each stressful task with more relaxing activities, such as deep breathing exercises.

1.6 Objective

The goal of this study is to create a closed-loop system that can modulate a BB stimulation in real-time with the aim of reducing detected stress. To achieve the goal 3 steps are required:

1. creation of a protocol that is able to induce both stress and relaxation in subjects, so that the model can learn from data of different types;
2. creation of a model that can estimate the stress level in real-time;
3. development of a method to adapt binaural beat stimulation according to the estimated stress.

Chapter 2

Methodology

2.1 Participants

Twenty-two healthy subjects (13 males and 9 females, age 35 ± 16 years old) participated in this study. Inclusion criteria were normal hearing, no heart problems, not color blind and accustomed to use a computer. All inclusion criteria were based on the subject's self-report. Participants were asked not to take coffee, alcohol and smoke 3 hours before the test. Participants are aware of collected data and have signed an informed consent. The study protocol was prepared following the principles of ethical research as of the declaration of Helsinki.

2.2 Collected data

The following data were collected from each participant:

- age;
- biological sex;
- RR intervals tachogram;
- tri-axial accelerometer;
- outcome responses during arithmetic and stroop tests;
- response times during arithmetic and stroop tests;
- estimation of self-reported stress level at the end of each task.

Age and biological sex allow us to understand how stress may vary among different demographic groups. RR intervals and tri-axial accelerometer measurements are used to train and use the artificial intelligence model that will detect, in real-time, the participant's stress level. Outcome responses, thus the percentage of correct answers, and response times during the arithmetic and stroop tests are used to assess any improvements due to the effect of binaural stimulation or learning. Finally, estimation of self-reported stress level at the end of each task enables labeling of data that will be used to train and test the machine learning model (see section 2.6.6).

2.3 Equipment

Equipment used includes:

- Polar H10 heart rate sensor;
- ASUS laptop computer;
- Mi In-ear Headphones Basic with jack connector.

The Polar H10 heart rate sensor is a chest strap, which records and transmits data in real-time to the computer (see section 2.3.1). The computer is needed throughout the test (see section 2.6) and is where relaxing and stressful videos will be shown and arithmetic and Stroop tests will be performed, while headphones are necessary to be able to listen to BBs and videos while isolating outside noise. Fig. 2.1 shows the equipment used during the protocol.

2.3.1 Polar H10

The Polar H10 chest strap, in Fig. 2.2, is a heart-rate monitoring device specialized in fitness and physical activity monitoring. This chest strap is designed to accurately measure heart rate during physical activity and send data to compatible devices via Bluetooth and ANT+ technology [35]. The PolarH10 sensor is equipped with a comfortable elastic band that does not bother during physical activity and it is adjustable to fit different chest sizes. It is water-resistant and has a long battery life. In addition, the PolarH10 band also has a triaxial accelerometer inside that can be streamed in real-time [36]. How RR intervals are extrapolated from the heart signal is not a public information, but the resolution appears to be less than 2 ms and tests with ECGs and other wearable devices show that the Polar H10 chest strap is an excellent sensor for HRV measures both at rest and during physical activity [37, 38].

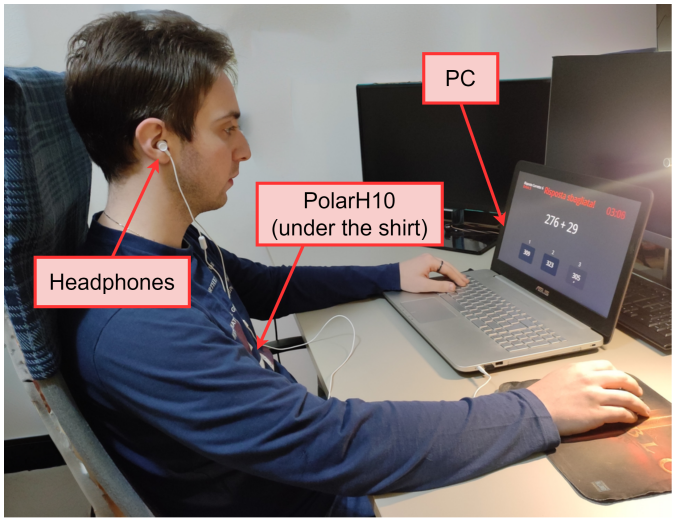

Figure 2.1: Equipment setup for protocol.

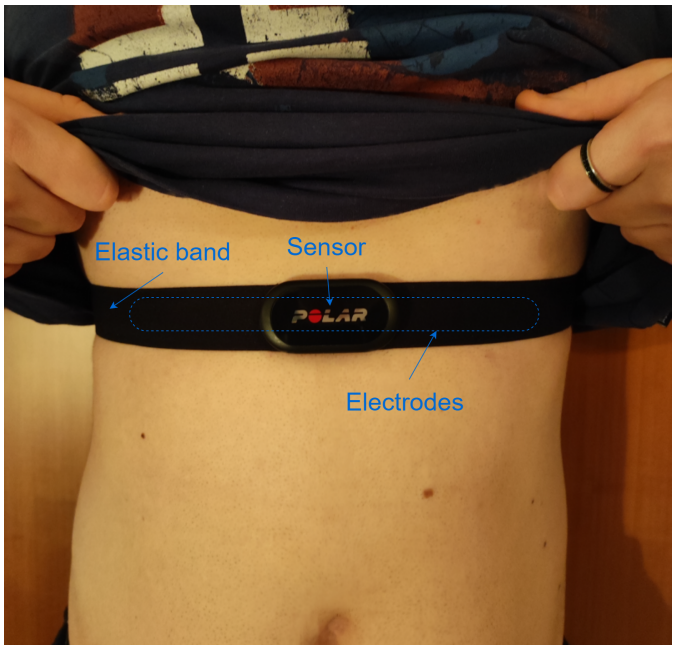

Figure 2.2: Polar H10 chest strap.

2.4 Feature Extraction

Since one of the objective of this thesis is to develop a system that can estimate stress in real-time, the sliding window technique is exploited, i.e. only the last few seconds of the recorded signal are analyzed to extract features of interest.

Extracted feature for RR intervals are in table 2.1. Extracted feature for the acceleration are in table 2.2.

Variable	Description	Formula
$SDNN$	Standard deviation of RR intervals	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(RR_i - \overline{RR})^2}$
$RMSSD$	Root mean square of successive differences	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(RR_{i+1} - RR_i)^2}$
$\overline{RR}$	Mean value of RR intervals	$\frac{1}{N}\sum_{i=1}^{N} RR_i$
$pNN50$	The proportion of NN50 divided by total number of RR intervals	$\frac{Count((RR_{i+1} - RR_i) > 50ms)}{N-1}$
$pNN20$	The proportion of NN20 divided by total number of RR intervals	$\frac{Count((RR_{i+1} - RR_i) > 20ms)}{N-1}$
last RR	Most recent RR value in the window	RR_N

Table 2.1: Tachogram and HRV features extracted from the window. The window has N RR intervals. RR_i is the i-th interval in the window.

Variable	Description	Formula
$\overline{Acc}$	Mean value of the acceleration	$\frac{1}{N}\sum_{i=1}^{N} Acc_i$
$SDAcc$	Standard deviation of the acceleration	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(Acc_i - \overline{Acc})^2}$
change Acc	Number of sign changes of the derivative ($Dacc$) of acceleration	$Count((Dacc_{i+1} \cdot Dacc_i) < 0)$
$H_s(ACC)$	Shannon Entropy of the acceleration	$-\sum_i p(Acc_i) \cdot \ln p(Acc_i)$

Table 2.2: Acceleration features extracted from the window. Each feature is extracted on all 3 signals of the accelerometer (x-axis, y-axis, z-axis). Acc_i is the i-th sample in the window. $p(Acc_i)$ is the probability of a specific sample of the signal and it is simply calculated as the number of time Acc_i appears in the window divided the length of the window.

It is important to note that since the length of the window used is at most 10 seconds (section 2.7.1), an analysis in the frequency domain is not possible, since the frequency resolution obtained is, at best, $1/10 = 0.1$ Hz. Such low resolution does not allow for proper frequency analysis. A simple workaround used for acceleration is the *change Acc* feature, that tries to give a little information about the rate of

change of the signal. A visual example is shown in Fig. 2.3

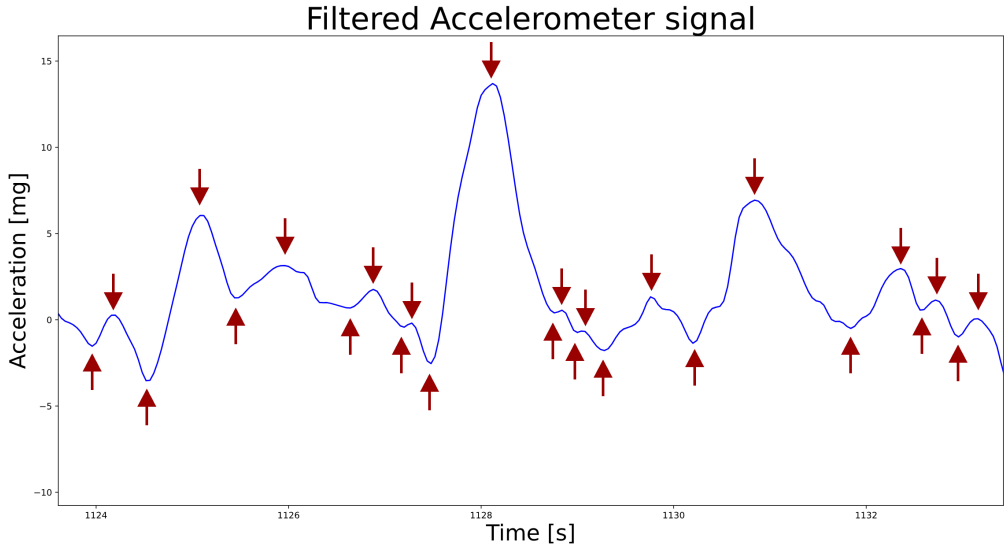

Figure 2.3: Visual example of *change Acc* feature. In this window there are 26 direction changes of acceleration (red arrows).

2.5 Data Pre-processing

The PolarH10 heart rate sensor connects via Bluetooth to the PC and allows real-time streaming of RR intervals and tri-axial acceleration. The pre-processing is done on the analysis window. The typical range of RR intervals is between 300 ms and 1300 ms range of 300 ms, equivalent to a range of 46 to 200 beats per minute (bpm) [39], therefore RR intervals outside this range are discarded and replaced by interpolation with Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) technique. If more than 30% of RR intervals in the analyzed window are discarded in this way, the whole window is discarded.

The accelerometer is sampled at 25 Hz. The accelerometer signal is filtered in real time using data from the last 60 seconds (less if a minute is not passed). This is to reduce the distortion effects that filtfilt filtering introduces both because of the filter transient and because of the inability to set the initial filter conditions in the scripts used. Fig. 2.4 shows filtering from the recorded accelerometer signal over a random window of 10 and 60 seconds. This figure emulates the filtering that occurs in real-time, using the last 10 and 60 seconds. The full filtered signal has been included as a reference.

The signal is high-pass filtered at 0.04 Hz (second-order Butterworth filter) to

remove the component due to gravity: gravity causes a constant acceleration along the axes into which it can be decomposed, but a simple removal of the mean value does not take into account a possible slight rotation, during the acquisition, of the accelerometer reference system that leads to changes in the decomposition of the force of gravity in the 3 axes. The signal is low-pass filtered at 2.5 Hz (second-order Butterworth filter) to reduce noise; the threshold at 2.5 Hz was chosen by looking at the power spectrum of the signal and trying to preserve the highest frequency that had a significant contribution. The goal is to record the respiratory signal, whose normal bandwidth is $\approx [0.1 \leftrightarrow 0.5]$ Hz [40], and torso movements as well. The accelerometer signal is filtered in the same way for all 3 axes.

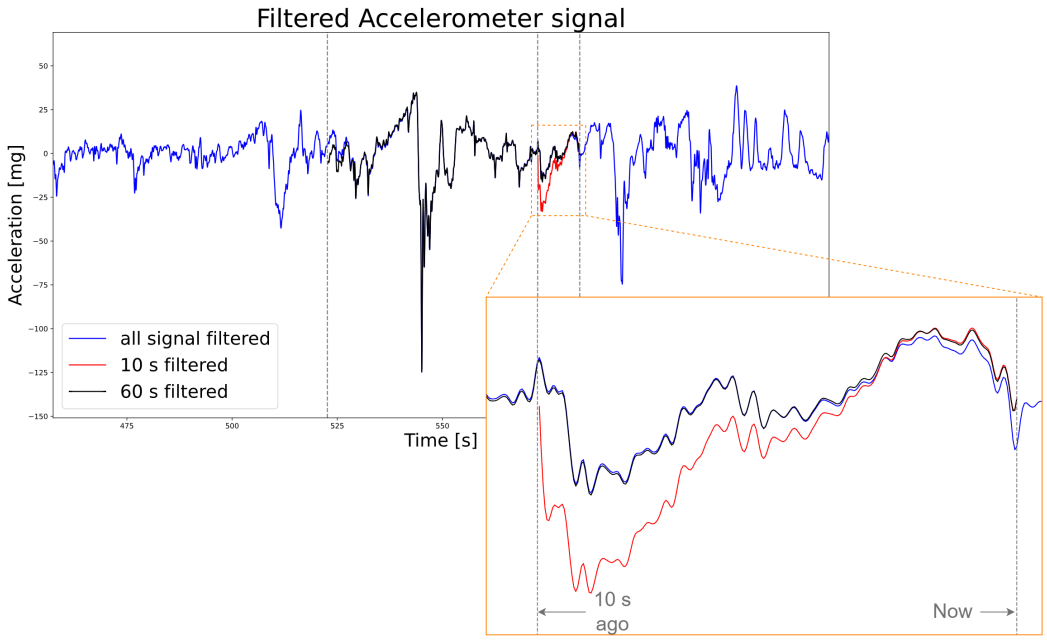

Figure 2.4: Simulation of real-time filtered accelerometer signal. In blue the entire signal is filtered (reference), in black the last 60 seconds are filtered, in red the last 10 seconds are filtered. The black signal is closer to the reference, except for the end of the window where it is closer to the red one.

2.6 Protocol

The participant wears the PolarH10 chest strap, sit in front of the computer, in a dimly lit room, and wears headphones. The protocol consists of 4 acquisitions, to be made on different days, waiting at least 5 days between each acquisition. The first acquisition has the purpose of acquiring data to build an artificial intelligence

model that can estimate stress in real time. In the second acquisition, half of the participants will be randomly selected and will be exposed, throughout the test (baseline excluded), to binaural beats (BB) modulated according to the stress detected by the trained model. The other half will be exposed to fixed binaural beats. The third acquisition is the opposite of the second: participants who had been exposed to modulated BB will now be exposed to fixed BB and vice versa. Finally, the fourth acquisition is used to verify that any reductions in stress are actually due to BB rather than a protocol habit effect.

A single acquisition involves performing 6 tasks, lasting approximately 3 minutes each. The tasks are: baseline, arithmetic test, relaxing video, stressful video, relaxing video and Stroop test. At the end of each task, the participant is asked to rate the average stress level perceived (self-reported stress level). Between acquisitions, the order of the 6 tasks is randomized, but there are some constraints to be met:

- the baseline must be first: it serves as a reference to normalize all data acquired in subsequent tasks.
- the last task should be either the arithmetic test or the Stroop test: these tasks stimulate a sense of challenge within people and can help keep the participant's interest high until the end.
- there must be alternation between stressful task and relaxing task: this allows the participant to recover from the stressful state.

Before starting the acquisition, the subject is required to adjust the computer brightness and volume to be comfortable. Participants are informed of the type of tasks in the protocol, but not the order in which they will be performed. The task order for each acquisitions is shown in Fig. 2.5.

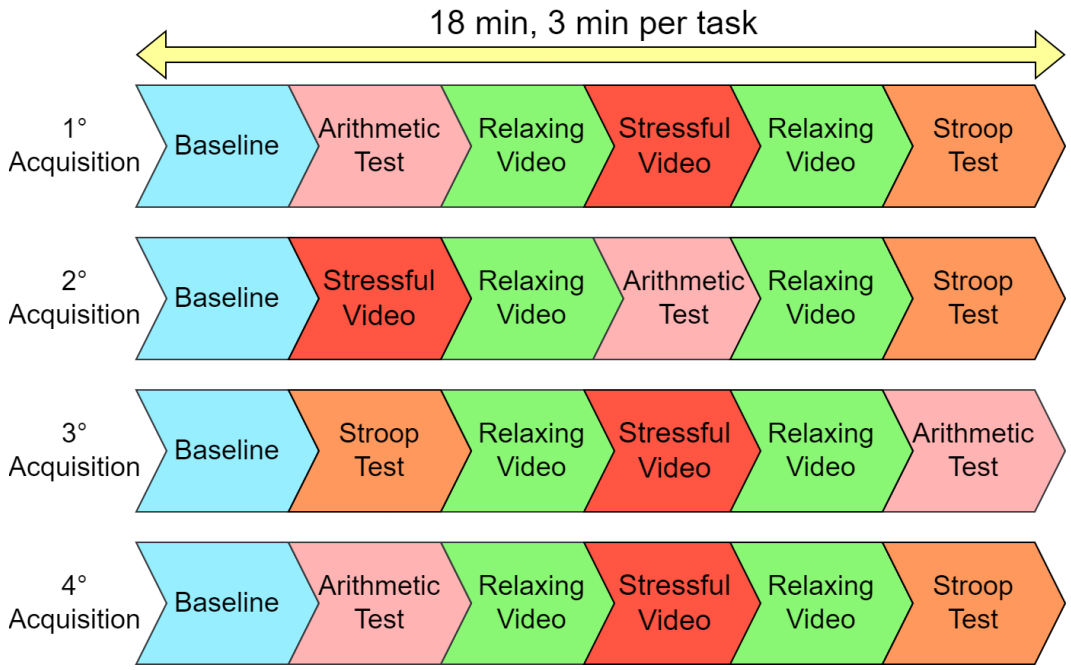

Figure 2.5: Task order in each acquisition. Used videos are listed in appendix A.

2.6.1 Baseline

The baseline consists of a video without audio. Kuijsters et al. [41] suggest the use of kaleidoscope videos instead of a blank screen to avoid boredom and annoyance to participants. Different videos are used for different acquisitions. Used videos are listed in appendix A. This task serves to acclimatize the participant and to have an initial reference to normalize the data of subsequent tasks. From the baseline, all features necessary for the artificial intelligence model are extracted and are normalized using the Standard Scaler algorithm of the scikit-learn library [29, 30]. The algorithm applies the transformation:

$$z = \frac{(x - u)}{s}$$

where x is the sample of the baseline, u is the mean and s is the standard deviation of the samples, thus removing the mean and scaling to unit variance. As indicated by the scikit-learn library, this step is a necessary pre-processing to be applied to all input data before feeding the model.

2.6.2 Arithmetic Test

Arithmetic test is one of the most widely used mental stressor [42, 43, 44, 45], but in this protocol it was slightly modified to further increase the time pressure. This test involves answering a series of multiple-choice mathematical calculations within a time limit. The time limit is reduced by 0.5 s for each correct answer and increased by the same amount for each error. The minimum time is 3 s and the maximum time 5 s. The timer is always visible during the task. As soon as the subject gives an answer, a new random question is immediately generated, and in case of an error in the previous answer, the text "Risposta sbagliata!" or "Risposta non data!" is shown in red on the screen for 1 s. A screenshot is shown in Fig. 2.6.

Figure 2.6: Screenshot of the Arithmetic Test. Time left in the upper right corner.

Each question has the following structure:

$$N_1 \; operator \; N_2 = Result$$

where N_1, N_2 and $Result$ are integers and $operator$ is one of the four basic arithmetic operators (addition +, subtraction -, multiplication $\cdot$, division /). In the case of addition and subtraction, N_1 is a random number between 200 and 2000, while N_2 is a random number between 10 and 100. In the case of multiplication, N_1 is a random number between 11 and 20, while N_2 is a random number between 5 and 15. In the case of division, N_2 is a random number between 5 and 9, while $Result$ is a random number between 5 and 15, consequently N_1 is obtained from $Result \cdot N_2$. To further increase the difficulty, wrong answers are taken within a range of ± 20 in the case of addition, subtraction and multiplication, while the

range is reduced to ±5 for division.

At the end of the test, the result obtained is shown compared with the ones obtained by other participants, thus stimulating participants' competition. To have reproducibility among different subjects within the same acquisition, a random seed was set. The random seed is different between acquisitions.

2.6.3 Relaxing Video

The relaxing video task requires to watch a relaxing video. The goal is to induce a relaxation state in the participant in order to recover from the previous stressful task. Two videos per acquisition are shown, but different between acquisitions. Used videos are listed in appendix A.

2.6.4 Stressful Video

The stressful video requires to watch a horror video. The goal is to induce a stressful state in the participant. Different videos are used for different acquisitions. Used videos are listed in appendix A.

2.6.5 Stroop Test

Stroop test is often used as mental stressor [5, 24], but in this protocol it was slightly modified to further increase the time pressure. This test consists of a series of words written in different colors (e.g., the word "red" written in blue or the word "green" written in red) and the subject must respond with the color of the word, ignoring the meaning of the word itself. There are 5 possible answer (i.e. colors) to choose from, within a time limit. The time limit is reduced by 0.5 s for each correct answer and increased by the same amount for each error. The minimum time is 1 s and the maximum time 3 s. The timer is always visible during the task. As soon as the subject gives an answer, a new random question is immediately generated, and in case of an error in the previous answer, the text "Risposta sbagliata!" or "Risposta non data!" is shown in red on the screen for 1 s. A screenshot is shown in Fig. 2.7.

Furthermore, occasionally the response mode changes, thus having to respond with the word itself, ignoring its color. This is useful in reducing task habit effects, in which simply stopping reading is enough to success: the change of mode forces the mind to have to readapt each time. The difficulty of the test lies in the fact that the human brain tends to automatically process the meaning of words, which can interfere with the ability to identify colors.

Five colors are used: red, blue, green, yellow and purple. One random color and one random word are selected. There is a 30% chance that the color and the word

Figure 2.7: Screenshot of the Stroop Test.

match. This, too, aims to reduce task habit: if the word and color never match, the subject may be inclined to immediately rule out what he or she reads, whereas if they may happen to coincide, one is less likely to discard that information.

Before the next question is generated, if sufficient time has passed, the mode is changed and the subject is warned with an on-screen message for 1 s (see Fig. 2.8). The "Color" mode, that is, the one in which one has to answer with the color of the answer, can last a random time between 30 and 45 seconds. The "Word" mode, that is, the one in which one has to answer with the word itself, can last a random time between 15 and 20 seconds.

Figure 2.8: (Stroop test) Change warning on screen.

At the end of the test, the result obtained is shown compared with the ones obtained by other participants, thus stimulating participants' competition. To have reproducibility among different subjects within the same acquisition, a random seed was set. The random seed is different between acquisitions.

2.6.6 Self-Reported Stress Level

At the end of each task, the participant is asked to rate the average stress level perceived. This value is used as the reference stress label for that specific task during training and testing sessions of the model. The answer is given as a number between -4 and +4, where -4 is high relaxation and +4 is high stress. Screen shown in Fig. 2.9.

Figure 2.9: Self-Reported Stress Level screen.

2.6.7 Binaural Beats Stimolation

Binaural Beats (BBs) at 16 Hz have been shown to be effective in increasing mental concentration and decreasing stress [5, 24]. Therefore BBs at 16 Hz will be used in this protcol. One of the goals of this thesis is to understand whether subject-adapted stimulation can be more effective in reducing stress and increasing concentration than fixed stimulation. For this reason, two types of BBs stimulation are provided:

- **Fixed** BBs: given at a constant 50% volume throughout the test, excluding baseline.
- **Modulated** BBs: given at a volume 0%, 50% or 100%. Two stress thresholds S_{50} and S_{100} are set. Every 5 seconds the stress detected in the previous 5 seconds is analyzed and if the stress is greater than S_{100}, the volume is raised to 100%. If the stress is greater than S_{50} but less than S_{100}, the volume is raised to 50%. If the stress is less than S_{50}, the volume goes down to 0%.

2.7 Model Training

After performing the first acquisition on all participants, the collected data are used to train a machine learning model to be able to predict stress in real-time. In practice, it is impossible to know a priori which parameters to choose in order to get the best possible model, so it is necessary to do a lot of training by varying the starting parameters each time.

2.7.1 Real-time Feature Extraction

When the model will be used in subsequent acquisitions, only the last few seconds of the recorded signals will be available to predict the subject's stress level. For this reason, the training will have to reflect this limitation.

Let's define some parameters for feature extraction:

- ***WL*** (window length) as the length of the time window.
- ***TBW*** (time between windows) as the time distance between two successive windows.
- ***overlap*** $= \dfrac{WL - TBW}{WL} \cdot 100$ as the overlap percentage of two successive time windows.

A graphic explanation is shown in Fig. 2.10. The chosen parameters are listed in the table 2.3.

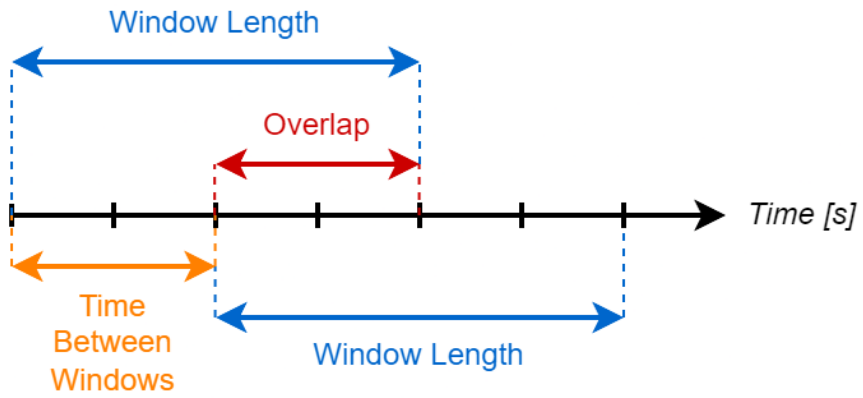

Figure 2.10: Window Length (WL), Time between windows (TBW) and Overlap.

WL [s]	***TBW*** [s]	***overlap***
5	1	80%
5	2	60%
10	1	90%
10	2	80%

Table 2.3: Explored parameters for feature extraction.

2.7.2 Labeling Protocol

As written in the subsection 2.6.6, 9 labels are extracted from the protocol (-4 $\leftrightarrow$ 4). In the literature, the most commonly used labeling protocol is binary labeling (stress or non-stress), but it is possible to find papers that use 3 classes (low, mid, high stress) [8]. It has been shown that a 3-class classification is significantly less accurate than a 2-class classification [46]. A high number of labels would allow more precise identification of the subject's true stress level, but there is a strong risk that the model would fail to have sufficiently high accuracy. Therefore, alongside the 9-class labelling protocol, the 3-class labelling protocol was also tested. The 3 classes are obtained by applying a simple transformation to the original labels:

9-class Label		**3-class Label**
(-4, -3, -2)	$\Rightarrow$	-1 (low stress)
(-1, 0, +1)	$\Rightarrow$	0 (mid stress)
(+2, +3, +4)	$\Rightarrow$	+1 (high stress)

2.7.3 Feature Matrix

Starting from the recorded signals, we slide by TBW seconds and extract the previous WL seconds. Let's define as t_0 the time at the beginning of the extracted window and as t_c (current time) the time at the end of the window. From this window all the features described in the section 2.4 are calculated, thus obtaining a vector 1×18, 6 feature deriving from HRV and 12 feature deriving from acceleration (4 features $\times$ 3 axis). The assigned label to this vector corresponds to the label the subject gave to the task performed at time t_c.

Repeating this step for all the windows that can be extracted from the individual subject and on all participants in the protocol, we construct the feature matrix $N\ samples \times 18\ features$ and the vector of labels $N\ samples \times 1$. Then the data are shuffled so that the machine learning algorithms are not affected by the order in which the data were extracted.

The resulting dataset is not balanced. It is therefore necessary to balance

it before training the model. The technique used is Random Undersampling, implemented in the python imbalanced-learn library [47]. Random undersampling it is the simplest undersampling technique and it involves randomly deleting samples from the majority class. This technique reduces the bias in selecting which samples to delete, but may remove valuable information for training.

2.7.4 Machine Learning Training

As already mentioned in the subsection 1.4, three supervised learning algorithms are explored: SVM, RF and MLP. For all three, the scikit-learn library has already implemented the classifier and the regressor algorithms. Therefore, 3 classifiers and 3 regressors will be used as machine learning models. Before starting model training, the feature matrix (i.e., the starting dataset) is further divided into Training set (80% of the dataset) and Testing set (20% of the dataset). These machine learning algorithms need hyper-parameters, i.e. parameters that are not directly learned during training. To fine-tune the hyper-parameters, a grid search is performed, exploiting the GridSearchCV algorithm from the scikit-learn library [29, 30]. Once all the hyper-parameters to be explored have been chosen, the GridSearchCV algorithm uses every possible combination of the parameters and trains the models using 5-Fold Cross-Validation to reduce overfitting. The score used to evaluate the performance of a trained classifier is accuracy, while for a trained regressor R^2 is used.
Hyper-parameters searched for the SVM Classifier are listed in table 2.4.
Hyper-parameters searched for the RF Classifier are listed in table 2.5.
Hyper-parameters searched for the MLP Classifier are listed in table 2.6.
Hyper-parameters searched for the SVM Regressor are listed in table 2.7.
Hyper-parameters searched for the RF Regressor are listed in table 2.8.
Hyper-parameters searched for the MLP Regressor are listed in table 2.9.

SVM Classifier	
Parameters	**Search Values**
C	[0.01, 0.1, 1, 10, 20]
gamma	[0.001, 0.01, 0.1, 'scale', 'auto']
kernel	['linear', 'rbf']

Table 2.4: SVM Classifier hyper-parameters used in grid search.

RF Classifier	
Parameters	**Search Values**
n_estimators	[50, 75, 100, 150, 200, 500, 750, 1000]
criterion	["gini", "entropy", "log_loss"]
max_features	['sqrt','log2', None]
bootstrap	[True]
n_jobs	[-1]

Table 2.5: RF Classifier hyper-parameters used in grid search.

MLP Classifier	
Parameters	**Search Values**
hidden_layer_sizes	[(15), (15, 15), (20), (20, 20), (15, 12, 9, 6, 3, 2), (20, 15, 10, 5, 2), (15, 13, 11, 9, 7, 5, 3)]
activation	['logistic', 'tanh', 'relu']
solver	['adam']
alpha	[0.0001, 0.0005, 0.001, 0.05, 0.01, 0.05, 0.1, 0.5]
learning_rate	['constant', 'invscaling', 'adaptive']
max_iter	[1000]
early_stopping	[True]
n_iter_no_change	[20]
validation_fraction	[0.1]

Table 2.6: MLP Classifier hyper-parameters used in grid search.

SVM Regressor	
Parameters	**Search Values**
C	[0.001, 0.01, 0.1, 1, 10]
gamma	[0.0001, 0.001, 0.01, 0.1, 'scale', 'auto']
kernel	['linear','rbf', 'sigmoid']

Table 2.7: SVM Regressor hyper-parameters used in grid search.

RF Regressor	
Parameters	**Search Values**
n_estimators	[50, 75, 100, 200, 400, 700, 1000]
criterion	["friedman_mse", "squared_error"]
max_features	['sqrt','log2', None]
bootstrap	[True]
n_jobs	[-1]

Table 2.8: RF Regressor hyper-parameters used in grid search.

MLP Regressor	
Parameters	**Search Values**
hidden_layer_sizes	[(5, 5), (10, 10), (15, 12, 9, 6, 3, 2), (20, 15, 10, 5, 2)]
activation	['identity', 'logistic', 'tanh', 'relu']
solver	['adam']
alpha	[0.0001, 0.001,0.01, 0.1]
learning_rate	['constant', 'invscaling', 'adaptive']
max_iter	[1000]
early_stopping	[True]
n_iter_no_change	[20]
validation_fraction	[0.1]

Table 2.9: MLP Regressor hyper-parameters used in grid search.

2.7.5 Feature Selection

Feature selection will be applied only on the final trained model. The reason for this choice is simply due to the need of reducing computational costs, which are already very high due to grid search, feature extraction parameters, and different types of labeling. The method used for feature selection will depend on the winning model, i.e., the one with the highest accuracy or R^2.

Chapter 3

Results

3.1 Trained machine learning model

To search for the optimal model, several feature extraction and labeling strategies were employed on six machine learning models, divided between classifiers (SVM, RF and MLP) and regressors (SVM, RF and MLP). This approach resulted in a total of 48 trained models, each characterized by a GridSearchCV analysis to identify the most effective set of hyper-parameters (see section 2.7).

From each model, performance on the test set were evaluated. In table 3.1 is shown the accuracy of the 24 trained classifiers. In table 3.2 is shown the R^2 value for the 24 trained regressors.

Classifier - Accuracy (%)					
N° Label	**WL [s]**	**TBW [s]**	**SVM**	**RF**	**MLP**
3	5	1	57.81	76.42	49.65
		2	51.99	71.76	48.96
	10	1	71.48	**84.85**	58.11
		2	64.19	78.73	53.99
9	5	1	42.55	67.62	27.68
		2	36.67	63.50	25.80
	10	1	61.54	79.27	36.46
		2	45.96	69.42	32.47

Table 3.1: Classifier accuracy on test set.

The chosen model, highlighted in bold in table 3.1, is the random forest classifier with $WL = 10[s]$, $TBW = 1[s]$ and 3 Label. This model has the highest performance on test set. The hyper-parameters of the chosen model are given in table 3.3.

Regressor - R^2					
N° Label	**WL [s]**	**TBW [s]**	**SVM**	**RF**	**MLP**
3	5	1	0.2211	0.4600	0.1591
		2	0.1215	0.3569	0.1314
	10	1	0.3761	0.5547	0.1960
		2	0.3019	0.4285	0.1557
9	5	1	0.2138	0.4761	0.2101
		2	0.1226	0.3592	0.064
	10	1	0.4046	0.5670	0.1961
		2	0.2348	0.4563	0.1506

Table 3.2: Regressor R^2 on test set.

RF Classifier	
Parameters	**Values**
n_estimators	750
criterion	"entropy"
max_features	"log2"
bootstrap	True
n_jobs	-1

Table 3.3: RF Classifier hyper-parameters used in grid search.

It is important to point out that in 9-class labeling the random undersampling (see section 2.4) greatly reduces the number of dataset samples used in learning, so it may not be suitable for model generalization.

3.1.1 Feature selection

The random forest implemented in the scikit-learn library [29, 30] has the property "feature_importances_" and it is specified that the greater the value, the more important the feature. The importance of a feature is computed as the (normalized) total reduction of the criterion brought by that feature. The importance of each feature used in the chosen model are shown in the tables 3.4 and 3.5.

In order to reduce the number of features needed, one feature at a time was removed, starting with the least important ones and re-training an RF model each time. Although not the best approach, for computational cost issues the models were trained without the use of GridSearchCV, but directly using the hyperparameters from the table 3.3. No training was able to obtain better results than the model with 18 features, and since in the current protocol reducing the number of features brings no real advantages in terms of real-time processing, since

real-time analysis is performed once per second, it was decided to keep all the features.

HRV Feature	Importance (%)
$SDNN$	5.58
$RMSSD$	5.51
$\overline{RR}$	9.06
$pNN50$	5.27
$pNN20$	3.88
$last\ RR$	5.46

Table 3.4: HRV feature importance.

Acc Features	Importance (%)
$\overline{Acc_x}$	3.51
$SDAcc_x$	5.91
$change\ Acc_x$	5.11
$H_{sx}(ACC)$	6.97
$\overline{Acc_y}$	3.68
$SDAcc_y$	5.70
$change\ Acc_y$	4.81
$H_{sy}(ACC)$	6.93
$\overline{Acc_z}$	3.76
$SDAcc_z$	6.36
$change\ Acc_z$	5.52
$H_{sz}(ACC)$	6.98

Table 3.5: Acceleration features importance.

3.1.2 Performance on subsequent acquisitions

In order to evaluate the generalization capabilities of the RF model, the accuracy on the whole dataset was calculated for each of the four acquisitions. The results obtained indicate an accuracy of 77.94% on the first acquisition, 44.37% on the second, 41.90% on the third and 49.18% on the fourth. It can be observed that the accuracy obtained in the first acquisition is lower than that obtained in the test set because, in this case, no undersampling was applied. This approach favored a balanced training of the model, however, the starting dataset did not have intrinsic balance, which negatively affected the overall performance when the model was tested on the entire dataset. An example of data obtained during acquisition, along with real-time prediction, are shown in Fig. 3.1.

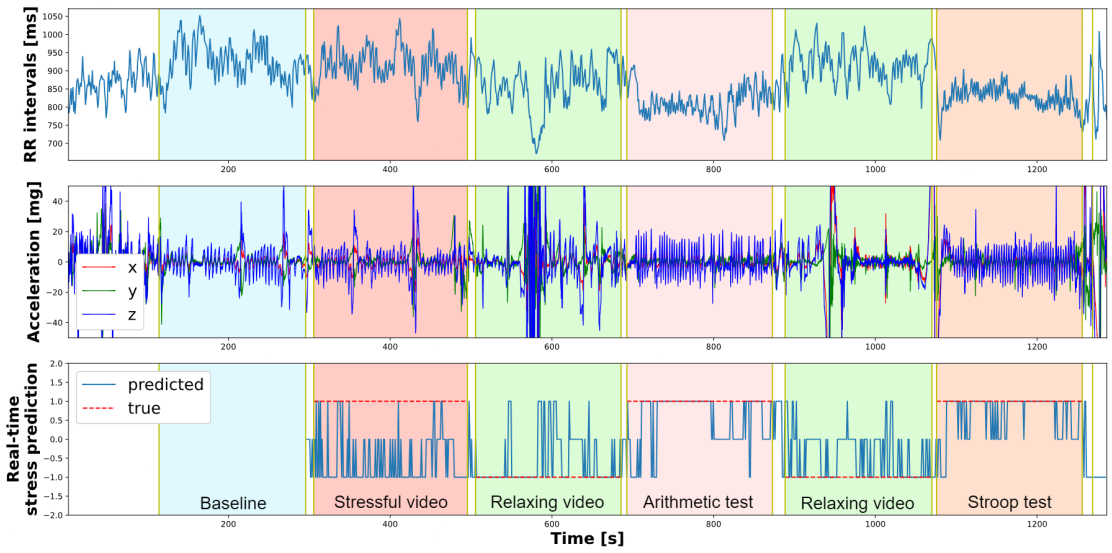

Figure 3.1: Data obtained from one acquisition: RR intervals, acceleration and real-time stress prediction. Gaps between tasks are the time periods in which the participant provides the self-reported stress level of the previous task.

3.2 Assessment of Stress Induction

It is important to verify that the protocol described in section 2.6 succeeds in inducing stress in participants. Three HRV parameters were used as physiological indicators of stress: SDNN, RMSSD and the mean value of RR intervals, hereafter referred to as MEAN RR. All three indicators decrease in presence of stress (see section 1.1.5). In Figure 3.2, individual indicators extracted during relaxation phases (baseline + two relaxing videos) and during stress phases (stressful video + arithmetic test + Stroop test) are compared. The statistical analysis performed involves checking the gaussianity of the distributions by Lilliefors test, followed by a paired sample t-test if both distributions were found to be gaussian, or a Wilcoxon signed-rank test if they were not. Lilliefors test was implemented via statsmodel library [48], while both Wilcoxon and t-test used are those implemented by scipy library [49].

In Fig. 3.2 can be observed that all 3 differences were statistically significant ($p < 0.001$), so it is likely to infer that the proposed protocol was able to induce stress within the participants.

Instead, comparing individual tasks with the baseline results in the graphs in Fig. 3.3 for stressful tasks and in Fig. 3.4 for relaxing tasks. SDNN, RMSSD and MEAN RR show statistically significant differences between each stressful task and baseline ($p < 0.001$), with the stressful video being the least impactful task

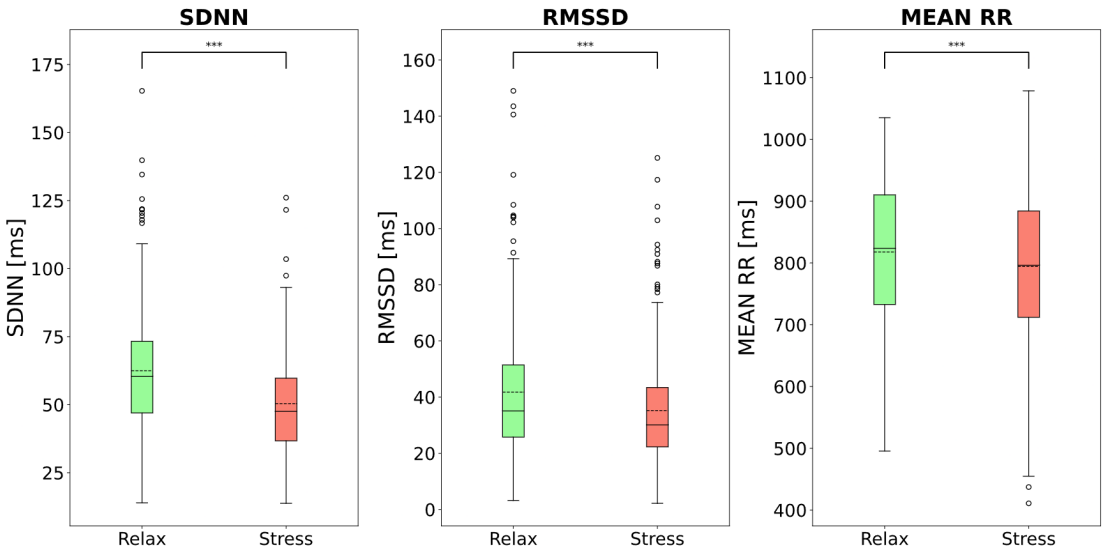

Figure 3.2: Relax vs Stress comparison. Mean value of box plots illustrated with a dashed line. *** $\Rightarrow$ $p < 0.001$.

and with MEAN RR greater than baseline (823.3 [ms] vs 835.9 [ms], $p < 0.01$). In contrast, looking at the comparison with relaxing tasks, there are few statistically significant differences: RMSSD in both relaxing videos is lower than baseline ($p < 0.001$), while MEAN RR is lower than baseline only in the first relaxing video (823.3 [ms] vs 815.3 [ms], $p < 0.05$).

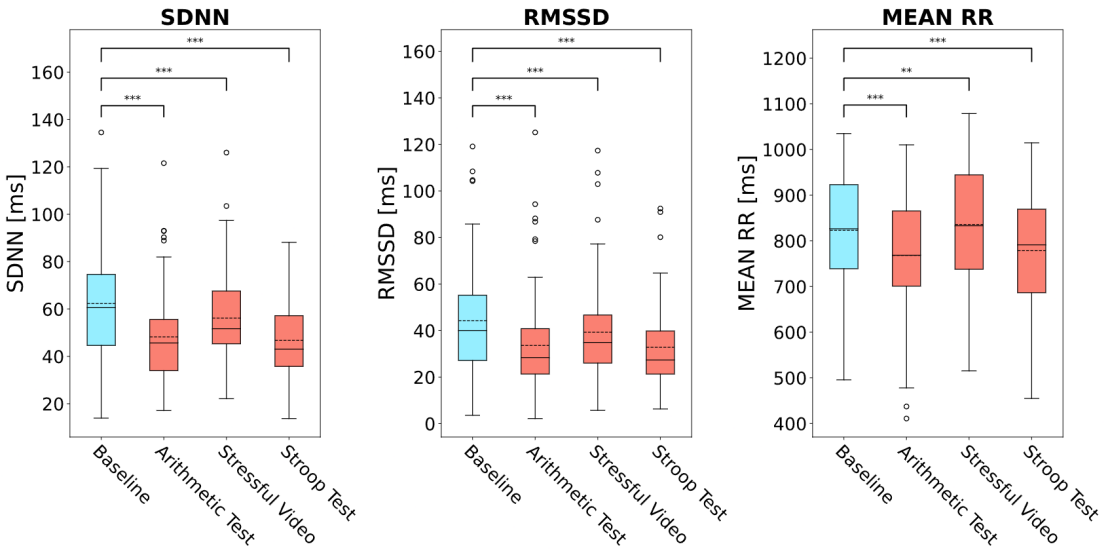

Figure 3.3: Baseline vs stressful tasks comparison. Mean value of box plots illustrated with a dashed line. ** $\Rightarrow$ $p < 0.01$, *** $\Rightarrow$ $p < 0.001$.

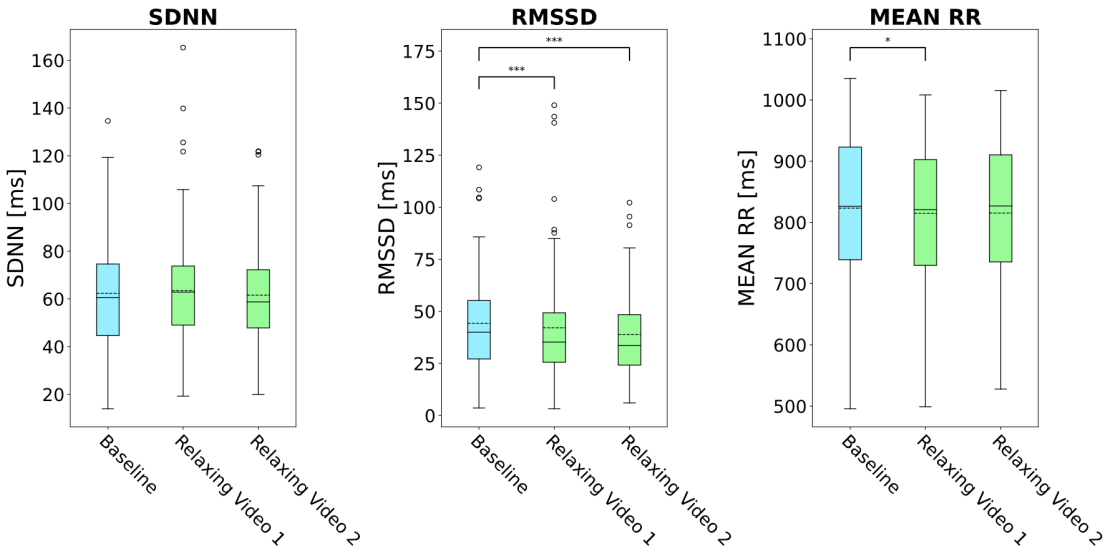

Figure 3.4: Baseline vs relaxing tasks comparison. Mean value of box plots illustrated with a dashed line. * $\Rightarrow$ $p < 0.05$, *** $\Rightarrow$ $p < 0.001$.

3.3 Statistical Comparison of Different Acquisitions

In Fig. 3.5, the trend of SDNN, RMSSD and MEAN RR between acquisition for each task is shown. In order to evaluate the effectiveness of fixed and modulated BB stimulation in reducing stress, a statistical analysis was conducted for each task in the protocol. For each task, the parameter distribution related to each acquisition were evaluated for gaussianity by applying a logarithmic transformation and using the Lilliefors test. In case all distributions were gaussian, the anova RM test was used to identify any statistical differences within the group; otherwise, the Friedman test was employed. If statistically significant differences emerge, a post-hoc analysis is performed using the Wilcoxon signed-rank test. A graphical explanation is shown in Fig. 3.6. The Lilliefors test was implemented with the statsmodel library [48], the anova RM test and the Friedman test with the pingouin library [50], and the Wilcoxon test with the scipy library [49].

No significant statistical differences ($p > 0.05$) emerged from the results of the analysis, except for the arithmetic test where the MEAN RR is significantly lower in the first acquisition compared to the acquisition with fixed BBs (731.4 [ms] vs 783.5 [ms], $p < 0.05$) or the fourth one (731.4 [ms] vs 778.8 [ms], $p < 0.01$). In contrast, the acquisition with modulated BBs presents a MEAN RR value of 779.2 [ms], very close to that with obtained with fixed BBs or in the fourth acquisition. In fact, no statistical differences are present. Interesting, however, is the absence of statistical significance with the first acquisition ($p = 0.079$).

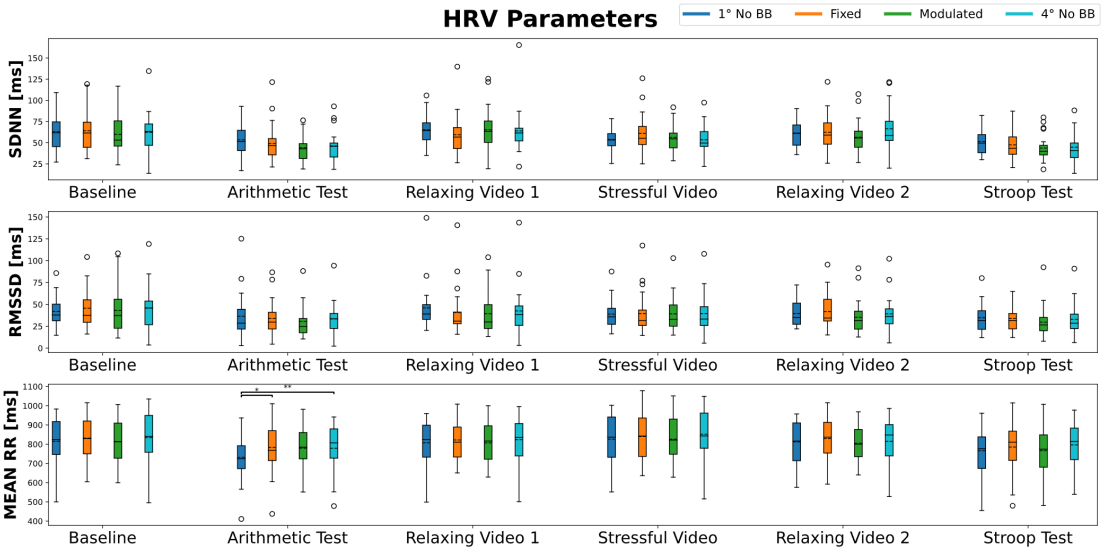

Figure 3.5: Differences of HRV parameters between acquisition for each task. 1°No BB, Fixed, Modulated and 4° No BB are the 4 acquisitions that each participant completed. The x-axis shows the tasks present in each acquisition. Mean value of box plots illustrated with a dashed line. * $\Rightarrow p < 0.05$, ** $\Rightarrow p < 0.01$.

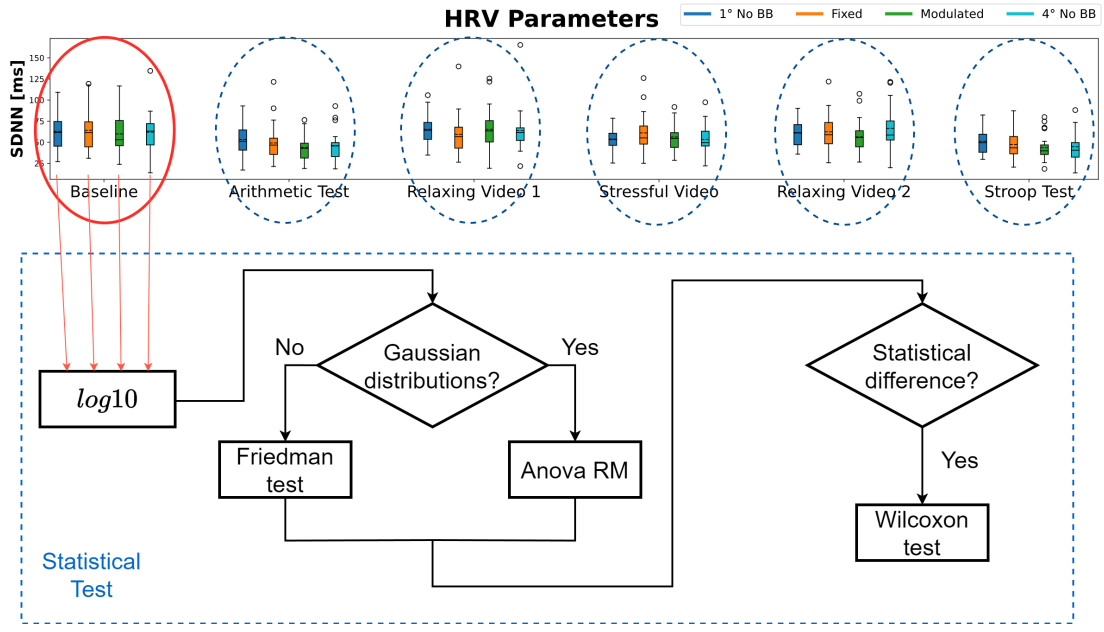

Figure 3.6: Graphical explanation of the statistical analysis performed in each task.

Performance (i.e. the percentage of correct answers) and response times related to the arithmetic test are presented in Fig. 3.7, while those of the Stroop test are shown in Figure 3.8. The same statistical approach described above was employed to find statistical differences in performance and response times between different acquisitions.

Statistically significant differences emerged mainly with the fourth acquisition, suggesting a cognitive task learning effect. Accordingly, a linear correction was performed to compensate for this learning. The corrected graphs of the arithmetic test and Stroop test are shown in Fig. 3.9 and Fig. 3.10, respectively. Linear correction ensures that the first and last acquisitions are equal. Statistically significant differences were found only in arithmetic test performance between the first acquisition and the one with modulated BBs (52.47 % vs 50.17 %, $p < 0.01$) and, consequently, also with the fourth acquisition.

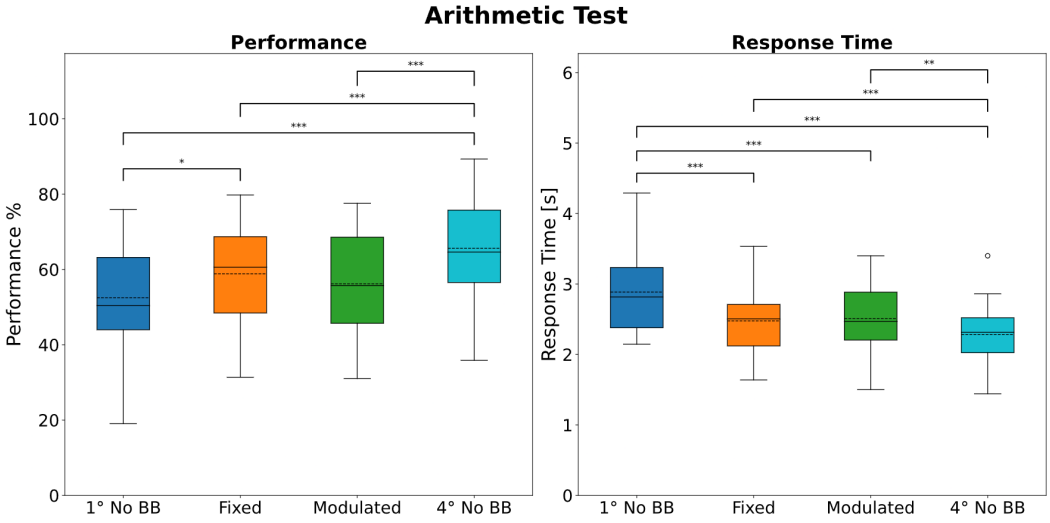

Figure 3.7: Performance and response time during arithmetic test for each acquisition. Mean value of box plots illustrated with a dashed line. * $\Rightarrow p < 0.05$, ** $\Rightarrow p < 0.01$, *** $\Rightarrow p < 0.001$.

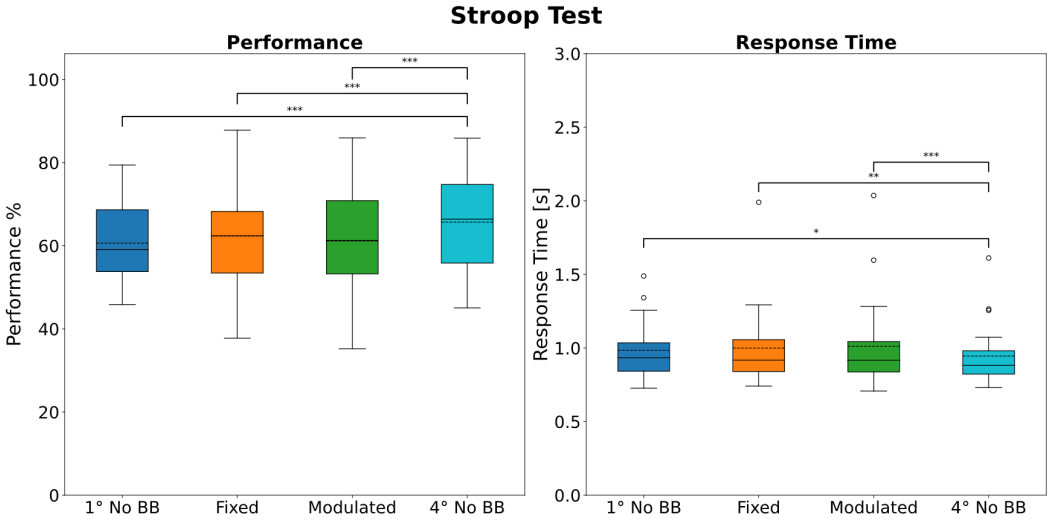

Figure 3.8: Performance and response time during Stroop test for each acquisition. Mean value of box plots illustrated with a dashed line. * $\Rightarrow$ $p < 0.05$, ** $\Rightarrow$ $p < 0.01$, *** $\Rightarrow$ $p < 0.001$.

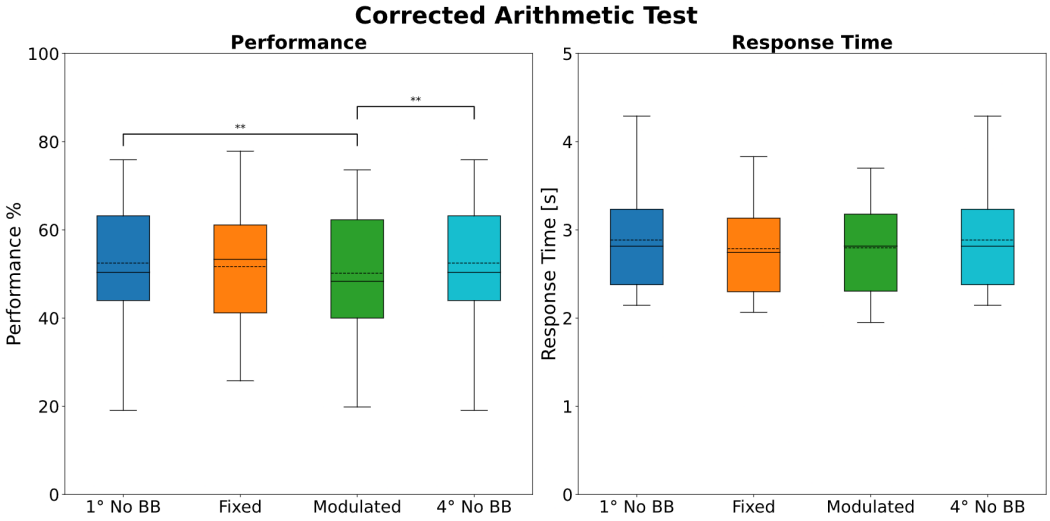

Figure 3.9: Corrected performance and response time during arithmetic test for each acquisition. Mean value of box plots illustrated with a dashed line. ** $\Rightarrow$ $p < 0.01$.

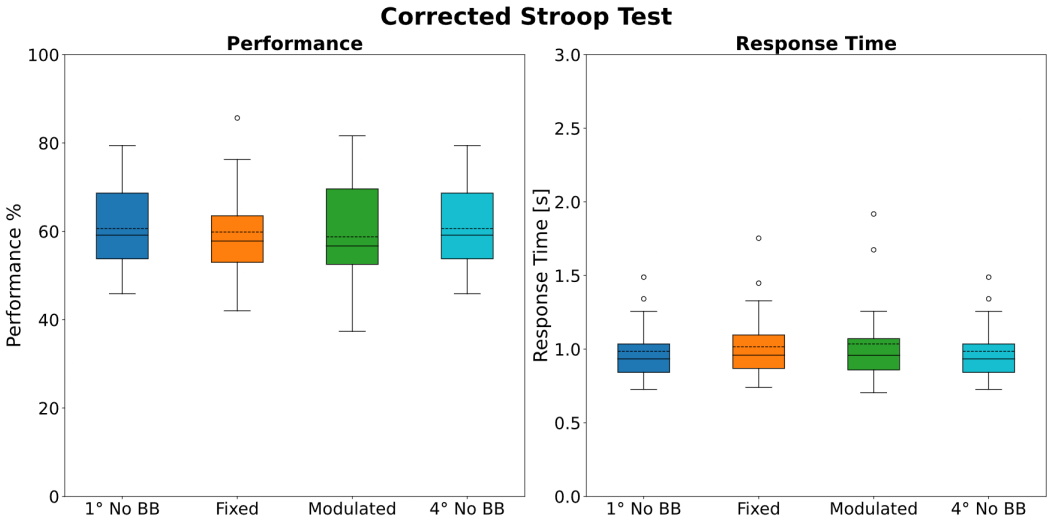

Figure 3.10: Corrected performance and response time during Stroop test for each acquisition. Mean value of box plots illustrated with a dashed line.

Chapter 4

Discussion

4.1 Machine Learning Model Performance

The performance of the model, as shown in the section 3.1.2, declined significantly in acquisitions following the first one, although it maintained an accuracy above 33.33%, which represents the probability of randomly identifying the correct label. This decline is mainly attributable to two factors: the normalization performed through the baseline of the single acquisition and the constraint of assigning only one label every 3 minutes, despite the fact that the model makes a prediction every second.

As explained in section 2.6, the first 3 minutes of each acquisition are devoted exclusively to collecting data that will be used by the standard scaler to normalize the data collected in the remaining 15 minutes. This procedure aims to capture the natural variability of each participant and use it as a reference point for subsequent tasks to detect variation and identify the participant's stress level. To further clarify, during the first acquisition, the data collected for each participant is normalized to their personal baseline, and the model is trained on that specific normalization. However, in subsequent acquisitions, the normalization used to predict stress level depends on the baseline of that specific session, which might reflect a different emotional state than the previous one. For example, if the participant, for personal reasons, started out more stressed than during the previous acquisition, an external stressful stimulus might simply maintain his stress level, without significantly increasing it compared to the baseline, thus the normalization of the data would make the physiological response to the stressful stimulus similar to that of the baseline, hindering the model's predictions.

The problem with labeling derives from the impracticality of requiring the subject to frequently assess his stress level, as repeatedly interrupting a task for this assessment could be both annoying and stressful, especially during tasks designed

to induce relaxation. Consequently, the labels assigned to tasks do not reflect the instantaneous stress level, but rather an average of those 3 minutes.

4.2 Protocol effectiveness

As evidenced in Fig. 3.2, the decrease in the three HRV parameters (SDNN, RMSSD and MEAN RR) indicates the effectiveness of the protocol in generating a stressful state in participants. Specifically, three types of stressful tasks were employed: an arithmetic test, the Stroop test and a stressful video.

As can be seen in Fig. 3.3, all three of the proposed stimuli elicited a state of stress relative to the neutral (baseline) condition. It is interesting to note that the stressful video is the least stimulating task. This is probably because, as detailed in appendix A, the selected videos all came from horror movies, which makes the response extremely subjective and, in addition, there is the risk that the subject has already seen those movies, further reducing the effectiveness of the stimulus. Instead, in Fig. 3.4 it is observed that the SDNN do not differ significantly from the neutral baseline video. However, it emerges that the RMSSD parameter, and the MEAN RR parameter but only with the first relaxing video, indicate a higher state of stress compared to the baseline condition, suggesting that the subject may be in recovery phase from the previous stimulus.

4.3 Binaural Beat effects

Katmah et al. [5] and Al-Shargie et al. [24] reported that the use of 16 Hz BB reduced stress and improved concentration during cognitive tasks. However, in Fig. 3.5, the three HRV parameters examined showed no statistical difference between the four acquisitions, regardless of the task, with the only exception for the arithmetic task and the MEAN RR parameter, where the first acquisition has a significantly smaller value than with the fixed BBs or the fourth acquisition. Nevertheless, this one difference is not sufficient to claim that BBs benefited the physiological parameters studied, because the difference is present only in a single task and in a single parameter of the 3 studied, and the difference is also seen with the fourth acquisition, where BBs were not used. This suggests, therefore, that both fixed and modulated stimulation caused no changes in the physiological parameters considered, and the only differences found should be explained in other ways. In addition, the results shown in Fig. 3.9 and Fig. 3.10 seem also to indicate that BB stimulation did not produce the expected benefits: acquisitions with modulated BBs have a significantly lower performance of 2.3 % than those without BBs, but fixed BBs did not bring the improvements predicted in the literature [5, 24], and there is no significant difference between acquisition with fixed and modulated BBs.

There could be three explanations for this results:

- The heterogeneity in terms of age and educational level of the group used in this study, in contrast to the homogeneity of the groups studied in the aforementioned studies, which consisted mainly of university students.
- The cognitive tasks proposed in the protocol have strict time limits, this is to make the task more stressful and challenging, but potentially obscuring the modest effects of BB.
- A 2.3 % reduction in arithmetic test performance could be explained either by a higher complexity of the random generated questions compared with acquisitions without BBs, or by an incorrect approximation of the learning effect, which for simplicity was considered linear, or by the possibility that a volume modulation of a sound wave could distract the participant, reducing their performance.

4.4 Limitations

As also highlighted by the review done by Vos et al. [8], stress is commonly treated as a binary variable, and the proposed machine learning models focus on identifying stressful phases only once the data have been collected, without concern for their reusability in future contexts. Moreover, there is no standardized protocol for inducing stress and many studies in the literature rely on a single type of stressful stimulus, used throughout the experiment (as in [5, 24]).

This study faced precisely these challenges:

- devising a protocol that could repetitively induce stress in the same participants, while minimizing habit effects;
- develop a machine learning model that was also applicable in future acquisitions.

Another limitation of this study is due to the type of data collected: in order to simplify data acquisition and minimize participant discomfort, we opted to use only the polar H10 band, which allows only the acquisition of HRV and chest acceleration. This allowed the experiment to be conducted on an extremely heterogeneous group of people, but it also limited the ability to describe the physiological response to stress, which is not limited to HRV and acceleration alone, but also involves other aspects, such as electrodermal activity.

Chapter 5

Conclusion

This study implemented a closed-loop system that modulates a BB stimulation according to the subject's stress, measured in real-time. To achieve this result, it was necessary to:

- design a protocol that alternately induced stress and relaxation in the participant, in a repeatable manner over multiple acquisitions;
- develop a model that predicted stress level in real time;
- implement a system that adjusted BB stimulation according to the detected stress.

A protocol consisting of 6 tasks was tested: the baseline, required for data normalization, 1 horror video and 2 cognitive tasks (arithmetic test and Stroop test) to induce stress, and 2 relaxing videos, to promote recovery from stressful phases. The protocol was shown to be effective in inducing stress in participants. However, BB stimulation did not affect the studied physiological parameters and did not lead to an increase in cognitive performance, contrary to what has been reported in the scientific literature. This phenomenon could be attributed to the intrinsic difficulty of cognitive tasks, which, while designed to induce stress even in the presence of habit or learning, could obscure the positive effects of BBs.

Future studies could focus on refining the proposed protocol, maintaining the ability to induce stress in successive acquisitions while still allowing the effects of BBs to be observed and studied. Moreover, in this thesis, BB modulation seems to have slightly reduced the performance in the arithmetic task compared with no BBs: this may have been caused by the fact that volume modulation distracted the participant. To overcome this problem, it may be interesting to combine BBs with music, as already explored in [25]. This option was not investigated in this study because of the use of videos, particularly horror videos, which require audio for full effectiveness.

The ability to estimate stress using a simple wearable sensor such as Polar H10 chest strap is of significant importance, as it facilitates further research aimed at reducing the amount of instrumentation required or integrating other sensory technologies, while keeping research costs low and allowing for a larger study sample, as minimal preparation for the experiment is required, and individuals less familiar with experimental research tend to feel less intimidated by such a simple device. This allows for the acquisition of more heterogeneous data and a deeper understanding of the physiological response to stress.

Appendix A

Protocol Videos

This appendix lists, within tables, all the videos used in the protocol. The tables have 5 columns:

- **Name**: the name given to the video in this thesis, useful for reference.
- **# Acq**: the acquisition number where the video was used. It is important to point out that in each acquisition there are 2 relaxing video that alternate stressful tasks. In that case it will be specified whether it is the first or the second video of that acquisition.
- **Source**: the link or reference from which the video was taken.
- **Start**: the minute at which the video starts.
- **End**: the minute at which the video ends.

It is important to point out that the order in which the videos were displayed was purely random (but the same for everyone), and the order actually used was reported in the tables.

Baseline Videos				
Name	**# Acq**	**Source**	**Start**	**End**
Abstract Liquid	1	https://www.youtube.com/watch?v=jgm58cbu0kw	0:00:15	0:03:15
Fractal Zoom	2	https://www.youtube.com/watch?v=pCpLWbHVNhk	0:35:30	0:38:30
Kaleidoscope	3	https://www.youtube.com/watch?v=q2fIWB8o-bs	1:04:13	1:07:13
Bubbles	4	https://www.youtube.com/watch?v=14XxolEJloE	2:53:20	2:56:20

Relaxing Videos				
Name	**# Acq**	**Source**	**Start**	**End**
Fire and rain	1-1°	https://www.youtube.com/watch?v=uRWacmxGIIc	0:01:15	0:04:15
River bird	1-2°	https://www.youtube.com/watch?v=F1gK85IEeDI	0:00:37	0:03:43
Rain	2-1°	https://www.youtube.com/watch?v=mPZkdNFkNps	1:51:15	1:54:15
Forest 1	2-2°	https://www.youtube.com/watch?v=XJZ-ENsgWyI	0:01:17	0:04:17
Fire popping	3-1°	https://www.youtube.com/watch?v=6j3hPg0t5fo	0:00:00	0:03:00
River city	3-2°	https://www.youtube.com/watch?v=E8fiyj1hNwk	0:00:00	0:03:00
Forest 2	4-1°	https://www.youtube.com/watch?v=XJZ-ENsgWyI	0:04:22	0:07:20
Waterfall	4-2°	https://www.youtube.com/watch?v=sa2JCQ1Dils	0:00:00	0:03:00

Stressful Videos				
Name	**# Acq**	**Source**	**Start**	**End**
Witch	1	Film "The Blair Witch project – Il mistero della strega di Blair (1999)"	1:14:32	1:17:29
Host	2	Film "Host - Chiamata mortale (2020)"	0:53:43	0:56:52
Pulse	3	Film "Pulse – Kairo (2001)"	0:25:58	0:29:04
Rec	4	Film "Rec (2007)"	1:06:55	1:10:04

Bibliography

[1] J D Quick. «Health consequences of stress». In: *Journal of Organizational Behavior Management* 8 (1987), pp. 19–36 (cit. on p. 1).

[2] Lawrence R Murphy. «Stress management in work settings: A critical review of the health effects». en. In: *Am. J. Health Promot.* 11.2 (Nov. 1996), pp. 112–135 (cit. on p. 1).

[3] Peggy A Thoits. «Stress and health: major findings and policy implications». en. In: *J. Health Soc. Behav.* 51 Suppl.1_suppl (2010), S41–53 (cit. on p. 1).

[4] Stavroula Leka, Aditya Jain, Sergio Iavicoli, and Cristina Di Tecco. «An evaluation of the policy context on psychosocial risks and mental health in the workplace in the European Union: Achievements, challenges, and the future». en. In: *Biomed Res. Int.* 2015 (Oct. 2015), p. 213089 (cit. on p. 1).

[5] Rateb Katmah, Fares Al-Shargie, Usman Tariq, Fabio Babiloni, Fadwa Al-Mughairbi, and Hasan Al-Nashash. «Mental stress management using fNIRS directed connectivity and audio stimulation». en. In: *IEEE Trans. Neural Syst. Rehabil. Eng.* PP (Jan. 2023), pp. 1–1 (cit. on pp. 1, 19, 30, 32, 50, 51).

[6] Patrice Boucher and Pierrich Plusquellec. «Acute stress assessment from excess cortisol secretion: Fundamentals and perspectives». en. In: *Front. Endocrinol. (Lausanne)* 10 (Nov. 2019), p. 749 (cit. on p. 1).

[7] Santtu M Seipäjärvi, Anniina Tuomola, Joona Juurakko, Mirva Rottensteiner, Antti-Pekka E Rissanen, Jari L O Kurkela, Urho M Kujala, Jari A Laukkanen, and Jan Wikgren. «Measuring psychosocial stress with heart rate variability-based methods in different health and age groups». en. In: *Physiol. Meas.* 43.5 (May 2022), p. 055002 (cit. on p. 1).

[8] Gideon Vos, Kelly Trinh, Zoltan Sarnyai, and Mostafa Rahimi Azghadi. «Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review». en. In: *Int. J. Med. Inform.* 173.105026 (May 2023), p. 105026 (cit. on pp. 1, 19, 34, 51).

[9] Susan Standring. *Gray's anatomy : the anatomical basis of clinical practice.* 41th. Philadelphia: Elsevier Limited, 2016 (cit. on p. 2).

[10] Cables and Sensor. *12-Lead ECG Placement Guide with Illustrations*. URL: https://www.cablesandsensors.eu/pages/12-lead-ecg-placement-guide-with-illustrations (cit. on pp. 3, 5).

[11] T N Guidelines and Guidelines American. «Guidelines Heart rate variability». In: *European Heart Journal* 17 (1996), pp. 354–381 (cit. on pp. 6, 7, 9).

[12] J Pan and W J Tompkins. «A real-time QRS detection algorithm». en. In: *IEEE Trans. Biomed. Eng.* 32.3 (Mar. 1985), pp. 230–236 (cit. on p. 6).

[13] Gautam, Ruchita, and Anil Sharma. «Detection of QRS complexs of ECG recording based on wavelet transform using Matlab». In: *International Journal of Engineering Science and Technology* 2 (July 2010) (cit. on p. 6).

[14] Raritan Costin, Cristian Rotariu, and Alexandru Pasarica. «Mental stress detection using heart rate variability and morphologic variability of EeG signals». In: *2012 International Conference and Exposition on Electrical and Power Engineering*. 2012, pp. 591–596. DOI: 10.1109/ICEPE.2012.6463870 (cit. on p. 9).

[15] Nida Ali, Jonas P Nitschke, Cory Cooperman, Mark W Baldwin, and Jens C Pruessner. «Systematic manipulations of the biological stress systems result in sex-specific compensatory stress responses and negative mood outcomes». en. In: *Neuropsychopharmacology* 45.10 (Sept. 2020), pp. 1672–1680 (cit. on p. 9).

[16] Melisa A Gantt, Stephanie Dadds, Debra S Burns, Dale Glaser, and Angelo D Moore. «The effect of binaural beat technology on the cardiovascular stress response in military service members with postdeployment stress». en. In: *J. Nurs. Scholarsh.* 49.4 (July 2017), pp. 411–420 (cit. on pp. 9, 12).

[17] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. «Stress and heart rate variability: A meta-analysis and review of the literature». In: *Psychiatry Investig.* 15.3 (Mar. 2018), pp. 235–245 (cit. on p. 9).

[18] Fred Shaffer and J P Ginsberg. «An overview of heart rate variability metrics and norms». In: *Front. Public Health* 5 (Sept. 2017) (cit. on p. 9).

[19] Hendrik Bonnemeier, Uwe K H Wiegand, Axel Brandes, Nina Kluge, Hugo A Katus, Gert Richardt, and Jürgen Potratz. «Circadian profile of cardiac autonomic nervous modulation in healthy subjects:» en. In: *J. Cardiovasc. Electrophysiol.* 14.8 (Aug. 2003), pp. 791–799 (cit. on p. 9).

[20] Guan-Zheng Liu, Yan-Wei Guo, Qing-Song Zhu, Bang-Yu Huang, and Lei Wang. «Estimation of respiration rate from three-dimensional acceleration data based on body sensor network». en. In: *Telemed. J. E. Health.* 17.9 (Nov. 2011), pp. 705–711 (cit. on p. 11).

[21] Hillel Pratt, Arnold Starr, Henry J Michalewski, Andrew Dimitrijevic, Naomi Bleich, and Nomi Mittelman. «A comparison of auditory evoked potentials to acoustic beats and to binaural beats». en. In: *Hear. Res.* 262.1-2 (Apr. 2010), pp. 34–44 (cit. on p. 12).

[22] Brian Moore. *An introduction to the psychology of hearing*. Leiden, Netherlands: Brill, Apr. 2013 (cit. on p. 12).

[23] J C R Licklider, J C Webster, and J M Hedlun. «On the frequency limits of binaural beats». en. In: *J. Acoust. Soc. Am.* 22.4 (July 1950), pp. 468–473 (cit. on p. 12).

[24] Fares Al-Shargie, Rateb Katmah, Usman Tariq, Fabio Babiloni, Fadwa Al-Mughairbi, and Hasan Al-Nashash. «Stress management using fNIRS and binaural beats stimulation». en. In: *Biomed. Opt. Express* 13.6 (June 2022), pp. 3552–3575 (cit. on pp. 12, 19, 30, 32, 50, 51).

[25] Patrick A McConnell, Brett Froeliger, Eric L Garland, Jeffrey C Ives, and Gary A Sforzo. «Auditory driving of the autonomic nervous system: Listening to theta-frequency binaural beats post-exercise increases parasympathetic activation and sympathetic withdrawal». en. In: *Front. Psychol.* 5 (Nov. 2014), p. 1248 (cit. on pp. 12, 52).

[26] B Mahesh. «Machine Learning Algorithms-A Review». In: *International Journal of Science and Research* (2018), pp. 381–386 (cit. on p. 13).

[27] Corinna Cortes and Vladimir Vapnik. «Support-vector networks». en. In: *Mach. Learn.* 20.3 (Sept. 1995), pp. 273–297 (cit. on p. 14).

[28] A Patle and D S Chouhan. «SVM kernel functions for classification». In: *2013 International Conference on Advances in Technology and Engineering (ICATE)*. Mumbai: IEEE, Jan. 2013 (cit. on p. 14).

[29] F. Pedregosa et al. «Scikit-learn: Machine Learning in Python». In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on pp. 14, 17, 18, 28, 35, 39).

[30] Lars Buitinck et al. «API design for machine learning software: experiences from the scikit-learn project». In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*. 2013, pp. 108–122 (cit. on pp. 14, 17, 18, 28, 35, 39).

[31] B Jijo and A Abdulazeez. «Classification Based on Decision Tree Algorithm for Machine Learning». In: *Journal of Applied Science and Technology Trends* 2 (2021), pp. 20–28 (cit. on p. 16).

[32] A Parmar, R Katariya, and V Patel. *Lecture Notes on Data Engineering and Communications Technologies*. Vol. 26. Springer Science and Business Media Deutschland GmbH, 2019, pp. 758–763 (cit. on pp. 16, 17).

[33] M W Gardner and S R Dorling. «Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences». In: *Atmospheric Environment* 32 (1998), pp. 2627–2636 (cit. on p. 17).

[34] Jongyoon Choi, Beena Ahmed, and Ricardo Gutierrez-Osuna. «Development and evaluation of an ambulatory stress monitor based on wearable sensors». en. In: *IEEE Trans. Inf. Technol. Biomed.* 16.2 (Mar. 2012), pp. 279–286 (cit. on p. 19).

[35] *Polar H10 Heart Rate Sensor User Manual*. URL: `https://support.polar.com/e_manuals/h10-heart-rate-sensor/polar-h10-user-manual-english/manual.pdf` (cit. on p. 22).

[36] *Polar H10 Heart Rate Sensor*. URL: `https://www.polar.com/en/sensors/h10-heart-rate-sensor` (cit. on p. 22).

[37] Polar Research and Technology. *Polar H10 Heart Rate Sensor System White Paper*. 2019. URL: `https://www.polar.com/sites/default/files/static/science/white-papers/polar-h10-heart-rate-sensor-white-paper.pdf` (cit. on p. 22).

[38] Marcelle Schaffarczyk, Bruce Rogers, Rüdiger Reer, and Thomas Gronwald. «Validity of the Polar H10 sensor for heart rate variability analysis during resting state and incremental exercise in recreational men and women». en. In: *Sensors (Basel)* 22.17 (Aug. 2022), p. 6536 (cit. on p. 22).

[39] Mouna Benchekroun, Pedro Elkind Velmovitsky, Dan Istrate, Vincent Zalc, Plinio Pelegrini Morita, and Dominique Lenne. «Cross dataset analysis for generalizability of HRV-based stress detection models». en. In: *Sensors (Basel)* 23.4 (Feb. 2023), p. 1807 (cit. on p. 25).

[40] Delaram Jarchi, Dario Salvi, Lionel Tarassenko, and David A Clifton. «Validation of instantaneous respiratory rate using reflectance PPG from different body positions». en. In: *Sensors (Basel)* 18.11 (Oct. 2018), p. 3705 (cit. on p. 26).

[41] Andre Kuijsters, Judith Redi, Boris de Ruyter, and Ingrid Heynderickx. «Inducing sadness and anxiousness through visual media: Measurement techniques and persistence». en. In: *Front. Psychol.* 7 (Aug. 2016), p. 1141 (cit. on p. 28).

[42] Fares Al-Shargie, Masashi Kiguchi, Nasreen Badruddin, Sarat C Dass, Ahmad Fadzil Mohammad Hani, and Tong Boon Tang. «Mental stress assessment using simultaneous measurement of EEG and fNIRS». en. In: *Biomed. Opt. Express* 7.10 (Oct. 2016), pp. 3882–3898 (cit. on p. 29).

[43] Matthias Norden, Amin Gerard Hofmann, Martin Meier, Felix Balzer, Oliver T Wolf, Erwin Böttinger, and Hanna Drimalla. «Inducing and recording acute stress responses on a large scale with the Digital Stress Test (DST): Development and evaluation study». en. In: *J. Med. Internet Res.* 24.7 (July 2022), e32280 (cit. on p. 29).

[44] K Ushiyama, T Ogawa, M Ishii, R Ajisaka, Y Sugishita, and I Ito. «Physiologic neuroendocrine arousal by mental arithmetic stress test in healthy subjects». en. In: *Am. J. Cardiol.* 67.1 (Jan. 1991), pp. 101–103 (cit. on p. 29).

[45] P Karthikeyan, M Murugappan, and S Yaacob. «A study on Mental Arithmetic Task based human stress level classification using Discrete Wavelet Transform». In: *2012 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT)*. Kuala Lumpur, Malaysia: IEEE, Oct. 2012 (cit. on p. 29).

[46] Ahmad Rauf Subhani, Wajid Mumtaz, Mohamed Naufal Bin Mohamed Saad, Nidal Kamel, and Aamir Saeed Malik. «Machine learning framework for the detection of mental stress at multiple levels». In: *IEEE Access* 5 (2017), pp. 13545–13556 (cit. on p. 34).

[47] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. «Imbalanced‡ learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning». In: *Journal of Machine Learning Research* 18.17 (2017), pp. 1–5. URL: http://jmlr.org/papers/v18/16-365 (cit. on p. 35).

[48] Skipper Seabold and Josef Perktold. «Statsmodels: Econometric and statistical modeling with python». In: *9th Python in Science Conference*. 2010 (cit. on pp. 41, 44).

[49] Pauli Virtanen et al. «SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python». In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2 (cit. on pp. 41, 44).

[50] Raphael Vallat. «Pingouin: statistics in Python». In: *The Journal of Open Source Software* 3.31 (Nov. 2018), p. 1026 (cit. on p. 44).