SAPIENZA
UNIVERSITÀ DI ROMA

Sapienza, University of Rome

Department of Computer, Control and Management Engineering Antonio Ruberti
PhD program in Automatic Control, Bioengineering and Operations
Research curriculum in Bioengineering

Thesis For The Degree Of Doctor Of Philosophy

Multimodal assessment of emotional responses by physiological monitoring: novel auditory and visual elicitation strategies in traditional and virtual reality environments

Thesis Advisor
Prof. Riccardo Barbieri

External Co-Advisor
Dr. Alessia Paglialonga

Candidate
Edoardo Maria Polo
1889181

Academic Year MMXIX-MMXXII (XXXV cycle)

Abstract

This doctoral thesis explores novel strategies to quantify emotions and listening effort through monitoring of physiological signals. Emotions are a complex aspect of the human experience, playing a crucial role in our survival and adaptation to the environment. The study of emotions fosters important applications, such as Human-Computer and Human-Robot interaction or clinical assessment and treatment of mental health conditions such as depression, anxiety, stress, chronic anger, and mood disorders. Listening effort is also an important area of study, as it provides insight into the listeners' challenges that are usually not identified by traditional audiometric measures. The research is divided into three lines of work, each with a unique emphasis on the methods of emotion elicitation and the stimuli that are most effective in producing emotional responses, with a specific focus on auditory stimuli.

The research fostered the creation of three experimental protocols, as well as the use of an available online protocol for studying emotional responses including monitoring of both peripheral and central physiological signals, such as skin conductance, respiration, pupil dilation, electrocardiogram, blood volume pulse, and electroencephalography. An emotional protocol was created for the study of listening effort using a speech-in-noise test designed to be short and not induce fatigue. The results revealed that the listening effort is a complex problem that cannot be studied with a univariate approach, thus necessitating the use of multiple physiological markers to study different physiological dimensions. Specifically, the findings demonstrate a strong association between the level of auditory exertion, the amount of attention and involvement directed towards stimuli that are readily comprehensible compared to those that demand greater exertion. Continuing with the auditory domain, peripheral physiological signals were studied in order to discriminate four emotions elicited in a subject who listened to music for 21 days, using a previously designed and publicly available protocol. Surprisingly, the processed physiological signals were able to clearly separate the four emotions at the physiological level, demonstrating that music, which is not typically studied extensively in the literature, can be an effective stimulus for eliciting emotions. Following these results, a flat-screen protocol was created to compare physiological responses to purely visual, purely auditory, and combined audiovisual emotional stimuli. The results show that auditory stimuli are more effective in separating emotions at the physiological level. The subjects were found to be much more attentive during the audio-only phase. In order to overcome the limitations of emotional protocols carried out in a laboratory environment, which may elicit fewer emotions due to being an unnatural setting for the subjects under study, a final emotional elicitation protocol was created using virtual reality. Scenes similar to reality were created to elicit four distinct emotions. At the physiological level, it was noted that this environment is more effective in eliciting emotions. To our knowledge, this is the first protocol specifically designed for virtual reality that elicits diverse emotions. Furthermore, even in terms of classification, the use of virtual reality has been shown to be superior to traditional flat-screen protocols, opening the doors to virtual reality for the study of conditions related to emotional control.

Keywords: emotions, listening effort, physiological signals, virtual reality.

Contents

List of Figures

List of Tables

Chapter 1

Summary

The main focus of this thesis is to explore emotional and stress responses through the analysis of various physiological signals, both peripheral and central. The research is divided into three primary projects, each with a unique emphasis on the methods of emotion elicitation and the stimuli that are most effective in producing emotional responses, with a specific focus on auditory stimuli. The findings of this research will provide valuable insights into the development of stimuli and protocols that can be applied in different fields of psychophysiology.

The thesis includes the development of different protocols, utilizing both traditional and innovative stimuli, including immersive stimuli in virtual reality. The advanced signal processing techniques have enabled the identification of physiological profiles in response to both emotional and stress-inducing auditory stimuli. The ultimate goal of this research is to expand our knowledge of the complex relationship between emotional states and physiological responses, and to contribute to the development of innovative techniques and tools for measuring and monitoring emotional and stress-related states.

Below are the three main aims of the thesis along with their descriptions:

1.1 AIM 1: Assessment of listening effort and emotional responses to auditory stimuli by physiological markers

In this project, two experiments were conducted to evaluate physiological responses to auditory stimuli. The first experiment aimed at creating a protocol for studying listening effort, specifically by creating a novel speech-in-noise test that was validated on a sample of normal hearing subjects in order to obtain two-time windows of high and low auditory effort. Physiological markers were extracted from electrocardiogram, skin conductance, respiration, blood volume, electroencephalogram, and pupil dilation data in order to create a framework for studying listening effort. The second experiment was aimed at studying the emotional content of music at the physiological level, using an online dataset that presents physiological recordings of a subject listening to songs that evoke four emotions over a period of 21 days. The results of these experiments will contribute to a better understanding of the physiological responses to auditory stimuli and their potential applications in the field of auditory research. Assessing physiological markers can be useful for

investigating listening effort and emotional responses to auditory stimuli because these markers can provide valuable insights into an individual's psychological and physiological states, which can inform our understanding of how individuals process and respond to auditory stimuli, which is still unclear.

Results relative to AIM 1 can be found in the following studies:

1. **E. M. Polo**, M. Mollura, A. Paglialonga, & R. Barbieri. (2021, September). Listening Effort: Cardiovascular Investigation Through the Point Process. In 2022 Computing in Cardiology (CinC). IN PRESS. DOI: 10.22489/CinC.2022.211

2. **E. M. Polo**, M. Mollura, A. Paglialonga & R. Barbieri. (2022) Decoding Emotions through Music: A Physiological Analysis of Emotion Recognition. Proceedings of the VIII Congress of the National Association of Bioengineering(GNB 2023), Jun 18-20 2023, Padova, Italy. ACCEPTED.

3. **E. M. Polo**, M. Lenatti, M. Zanet, R. Barbieri, A. Paglialonga. 'Preliminary evaluation of the Speech Reception Threshold measured using a new language-independent screening test as a predictor of hearing loss'. Abstract presented at 1st Virtual Conference on Computational Audiology (VCCA). (June 19 2020).

4. **E. M. Polo**, M. Zanet, M. Lenatti, T. van Waterschoot, R.Barbieri, A.Paglialonga, "Development and Evaluation of a Novel Method for Adult Hearing Screening: Towards a Dedicated Smartphone App", Proceedings of the 7th EAI International Conference on IoT Technologies for HealthCare, 2020 [1]. DOI: 10.1007/978-3-030-69963-5_1

5. **E. M. Polo**, M. Zanet, A.Paglialonga, R.Barbieri. "Preliminary Evaluation of a Novel Language Independent Speech-in-Noise Test for Adult Hearing Screening." European Medical and Biological Engineering Conference. Springer, Cham, 2020 [2]. DOI: 10.1007/978-3-030-64610-3_109

6. A. Paglialonga, **E. M. Polo**, M. Zanet, G. Rocco, T. van Waterschoot, R. Barbieri, "An Automated Speech-in-Noise Test for Remote Testing: Development and Preliminary Evaluation", American Journal of Audiology, vol. 29, no. 3S, pp. 564-576, 2020 [3]. DOI: 10.1044/2020_AJA-19-00071

7. M. Zanet*, **E. M. Polo***, M. Lenatti, T. van Waterschoot, M. Mongelli, R. Barbieri, A. Paglialonga. "Evaluation of a Novel Speech-in-Noise Test for Hearing Screening: Classification Performance and Transducers Characteristics." IEEE Journal of Biomedical and Health Informatics (2021) [4]. DOI: 10.1109/JBHI.2021.3100368

8. M. Lenatti, **E. M. Polo**, M. Paolini, M. Mollura, M. Zanet, R. Barbieri, A. Paglialonga. (2021) 'Evaluation of multivariate classification algorithms for hearing loss detection through a speech-in-noise test'. Abstract presented at 2st Virtual Conference on Computational Audiology (VCCA). (June 25 2021).

9. **E. M. Polo**, M. Mollura, R. Barbieri, & A. Paglialonga. (2023, March). Multivariate Classification of Mild and Moderate Hearing Loss Using a Speech-in-Noise Test for Hearing Screening at a Distance. In IoT Technologies for HealthCare: 9th EAI International Conference, HealthyIoT 2022, Braga, Portugal, November 16-18, 2022, Proceedings (pp. 81-92). Cham: Springer Nature Switzerland [5].

10. M. Lenatti, P. A. Moreno-Sánchez, **E. M. Polo**, M. Mollura, R. Barbieri, & A. Paglialonga (2022). Evaluation of machine learning algorithms and explainability techniques to detect hearing loss from a speech-in-noise screening test. American Journal of Audiology, 31(3S), 961-979 [6]. DOI: 10.1044/2022_AJA-21-00194

    * co-first authors

1.2 AIM 2: Assessment of physiological responses to auditory, visual, and combined auditory and visual emotional stimuli

Physiological responses to emotional stimuli can be used to assess the impact of emotional stimuli on behavior and performance. For example, research has shown that negative emotional states can impair cognitive performance, while positive emotional states can enhance performance. By measuring physiological responses to emotional stimuli, researchers can better understand how emotional states influence behavior and performance. This project involved the creation of a protocol divided into three phases using visual, auditory, and combined visual and auditory stimuli. The visual stimuli consisted of images from the International Affective Picture System database, while the auditory stimuli were taken from the International Affective Digitized Sounds database. The protocol was designed to increase arousal levels, with stimuli starting at low levels and gradually increasing to high levels with alternating low and high valence. During each phase of the protocol, electrocardiogram, skin conductance, respiration, blood volume, electroencephalogram, and pupil dilation signals were acquired in order to compare the physiological responses to different types of stimuli and understand which type was the most effective. The results of this protocol will contribute to a better understanding of the physiological responses to visual, auditory, and combined stimuli, and could have applications in various fields of psychophysiology. Overall, assessing physiological responses to emotional stimuli can provide valuable insights into emotional processing and the impact of emotions on behavior and performance, and can be useful for a wide range of research and applied purposes.

Results relative to AIM 2 can be found in the following studies:

1. **E. M. Polo**, Farabbi, A., M. Mollura, R. Barbieri, A. Paglialonga & L. Mainardi. (2022, July). Analysis of the skin conductance and pupil signals for evaluation of emotional elicitation by images and sounds. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 1968-1971) [7].

2. **E. M. Polo**, M. Mollura, A. Paglialonga & R. Barbieri. (2022) 'Quantitative assessment of the influence of sound in affective audio-visual elicitations'. Abstract presented at HEAL 2022

- HEaring Across the Lifespan (16-18 June 2022).

3. **E. M. Polo**, A. Farabbi, M. Mollura, A. Paglialonga, L. Mainardi, R. Barbieri. (2022). Comparative assessment of physiological responses to emotional elicitation by auditory and visual stimuli. IEEE Journal of Translational Engineering in Health and Medicine. SUBMITTED.

4. **E. M. Polo**, A. Farabbi, L. Mainardi, R.Barbieri. Unlocking the Power of Emotion in Marketing: Using Machine Learning to Analyze Neurophysiological Responses to Visual, Auditory, and Combined Stimulation. Abstract accepted for further publication on Frontiers in Human Neuroscience.

5. A. Farabbi, **E. M. Polo**, R. Barbieri, & L. Mainardi. (2022, July). Comparison of different emotion stimulation modalities: an EEG signal analysis. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 3710-3713). IEEE [8].

1.3 AIM 3: Emotional elicitation beyond standard stimulation strategies: auditory and visual stimulation using virtual reality

Virtual reality technology allows for the creation of immersive and highly realistic auditory and visual stimuli that can elicit strong emotional responses. Standard stimulation strategies, such as presenting images or audio recordings on a computer screen or through headphones, may not fully capture the complex and dynamic nature of real-world emotional experiences. By using virtual reality technology, researchers can create highly realistic and immersive stimuli that more closely mimic real-world emotional experiences, allowing for the study of more complex and nuanced emotional responses. This project aimed to study the physiological response to immersive emotional stimuli in virtual reality. A protocol was developed that included four rooms representing emotions with increasing arousal levels (sadness, relaxation, happiness, and fear), interspersed with neutral rooms. During the protocol, electrocardiogram, skin conductance, respiration, and blood volume signals were acquired. To validate the emotional experiences, a survey was administered to each subject after completion of the protocol. The most notable and original design ideas were to provide stimuli that were composed in a similar manner to a realistic scenario and to combine visual and audio levels to enhance the sense of presence. To our knowledge, this is the first study that integrates advanced signal processing and virtual reality. The results of this study will contribute to a better understanding of the physiological responses to immersive emotional stimuli in virtual reality and their potential applications in various fields. Overall, investigating emotion elicitation using virtual reality technology can be useful for creating highly realistic and immersive stimuli, precisely manipulating and controlling auditory and visual stimuli, and improving the reliability and validity of research findings.

These results have just been achieved in the last year and they are in the process of being submitted in the following formats:

1. **E. M. Polo**, M. Mollura, A. Paglialonga, & R. Barbieri. Exploring Emotional Responses

in Virtual Reality Through Skin Conductance Signal. Proceedings of the 2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE 2023), Milan XXX 2023. Accepted abstract.

2. **E. M. Polo**, M. Mollura, A. Paglialonga, & R. Barbieri. Exploring Emotions in Virtual Reality: Enhancing Recognition through Physiological Signals Acquisition. IEEE Transactions on Affective Computing. To be sumbitted soon.

Below are the methodological studies that served for all three AIMS:

1. **E. M. Polo**, M. Mollura, Lenatti, M. Lenatti, A. Paglialonga, & R. Barbieri. (2021, November). Emotion recognition from multimodal physiological measurements based on an interpretable feature selection method. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 989-992). IEEE [9]. DOI: 10.1109/EMBC46164.2021.9631019

2. **E. M. Polo**, M. Mollura, M. Zanet, M. Lenatti, A. Paglialonga, & R. Barbieri. (2021, September). Analysis of the Effect of Emotion Elicitation on the Cardiovascular System. In 2021 Computing in Cardiology (CinC) (Vol. 48, pp. 1-4). IEEE [10].
DOI: 10.23919/CinC53138.2021.9662859

1.4 Thesis outline

Chapter 2 of this thesis provides a comprehensive overview of the background concepts and the overall significance of the study. The chapter begins by focusing on the general context of listening effort and emotions, and analyzing various stimulation protocols used in the literature, as well as different physiological and qualitative measures to delineate different emotional states. This contextualization is important to provide a solid foundation for the subsequent analyses and conclusions in the thesis.

Chapter 3 describes the methodological frameworks proposed in this work. Section 3.1 focuses on the auditory part, presenting a new speech-in-noise test for the study of listening effort and a protocol for emotional elicitation through musical stimuli. Section 3.2 describes the creation of a protocol for emotional elicitation in flat screen through the use of images, sounds, and a combination of both stimuli. Section 3.3 focuses on the creation of a protocol for emotional elicitation in virtual reality. Finally, Section 3.4 focuses on the elaboration of physiological signals monitored during the acquisition of the different protocols created and used in this thesis, as well as the extraction of features for subsequent statistical analysis and classification.

Chapter 4 presents the results of the study, divided by protocol. This chapter provides an in-depth analysis of the results obtained from each protocol, highlighting the differences and similarities in emotional and physiological responses across the different stimuli used in the study.

Chapter 5 provides a discussion on the different protocols presented in the thesis, along with innovations, limitations and future directions. This chapter critically evaluates the effectiveness

of the various methods proposed, highlighting the strengths and weaknesses of each protocol and discussing possible improvements and future research directions.

Finally, Chapter 6 presents the conclusions of this thesis, summarizing the main findings and contributions of the study. The conclusions emphasize the importance of emotional and stress-related states in various fields of psychophysiology and highlight the potential for future research in this area. Overall, this thesis provides a comprehensive analysis of the relationship between emotional states and physiological responses, providing valuable insights into the development of innovative techniques and tools for measuring and monitoring emotional and stress-related states.

Chapter 2

Introduction

2.1 Physiology of emotions

2.1.1 The structure of emotions

Emotions play a fundamental role in the lives of human beings as they represent an evolutionary factor that ensures survival and reproduction through adaptation to the environment [11]. From a biological perspective, emotions are a complex network of neural and hormonal interactions that generate cognitive processes that influence decision-making [12].

They have a key role in human behavior as they control the muscles of the face and other parts of the body, resulting in characteristic bodily structures and attitudes that reflect the individual's emotional state at the moment. Facial expressions, posture, gestures, and mimicry, which often accompany speech, make human communication more marked.

While some emotions may lead to action, others may result in inaction or even paralysis, such as the primary defense mechanism of freezing in highly threatening situations. Additionally, people may choose to regulate their emotional responses based on previous experiences or knowledge, such as choosing to remain still when encountering a venomous snake. In some cases, people may also display apathy, characterized by a lack of motivation or willingness to react, due to various factors including pathological or non-pathological causes. These examples demonstrate that emotions are not always manifested through observable behaviors ([13][14][15]).

According to the Component Process Model proposed by Scherer [16], emotions are composed of five synchronized processes: cognitive appraisal, bodily symptoms, action tendencies, expression, and feelings. Cognitive appraisal refers to the evaluation of events and objects, which helps to determine the emotional significance of a particular stimulus. Bodily symptoms refer to the physiological component of emotional experience, which includes physiological reactions in the central and autonomic nervous systems (CNS and ANS). Action tendencies represent the motivational component of emotions, which drive the preparation and direction of motor responses. Expression refers to the facial and vocal expression that often accompanies an emotional state and serves to communicate reaction and intention. Finally, feelings refer to the subjective experience of an emotional state once it has occurred. Together, these processes work in synchrony to produce the complex and multifaceted experience of emotions. By understanding the various components that make up emo-

tions, researchers can gain insights into the ways in which emotions shape our thoughts, actions, and behaviors.

Emotions play indeed a central role in determining individual's thoughts, actions, subjective perceptions of the world, and behavioral responses. Given the wide-ranging influence that emotions have on the quality of life and on how people interact with one another, this topic has garnered increasing attention in recent years. Researchers have sought to understand the various factors that contribute to the experience and expression of emotions, as well as the ways in which emotions impact cognitive and behavioral processes. Some of the key areas of focus in the study of emotions include the neural and hormonal underpinnings of emotional processing, the role of cultural and social factors in shaping emotional responses, and the influence of emotions on decision-making and behavior. Through a better understanding of these and other aspects of emotions, researchers hope to gain insights into the ways in which emotions can be effectively managed and regulated in order to promote well-being and improve social interactions.

2.1.2 Emotion Classification Models

In order to classify emotions, usually literature refers to two main approaches. The first, the theory of basic emotions [17], posits that emotions are a limited set of distinct states, typically including anger, fear, disgust, sadness, and happiness [18]. These emotions are thought to be independent of one another in terms of their behavioral, psychological, and physiological expressions, and each activates specific neural pathways in the ANS and CNS [19]. It is believed that each emotional state is associated with a unique physiological pattern.

For these models which define emotions as categories, such as the Pick-A-Mood (PAM) model [20], emotions are believed to be universal. The PAM model, for example, uses cartoon characters to help people easily and clearly communicate their moods.

It is important to note that emotions are not always manifested through explicit behaviors, and some affective behaviors, such as frowning, may occur without any underlying emotional experience. This means that people may experience certain emotions without exhibiting significant changes in behavior, and animals may display specific behaviors in response to brain stimulation without experiencing corresponding emotions [19].

Research has also shown that different basic emotions can elicit similar physiological responses [21], highlighting the subtlety of distinguishing between distinct basic emotions from both behavioral and physiological perspectives.

This has led some researchers to advocate for a shift from the basic emotion perspective to a dimensional approach, which views emotional states as the result of the combination of two or more independent dimensions [19]. Dimensional models allow for more flexibility in defining emotions by considering them as dimensions rather than discrete categories. One well-known example of a continuous model is the Russell Circumplex Model, which proposes that all emotions arise from two independent neurophysiological systems related to valence and arousal (See Figure 2.1).

Unlike categorical models, dimensional models allow for a more nuanced understanding of emotions and their complex interrelationships. The use of a circular model, which is created by crossing two main dimensions, reflects the idea that it is difficult to categorize emotions into discrete, sep-

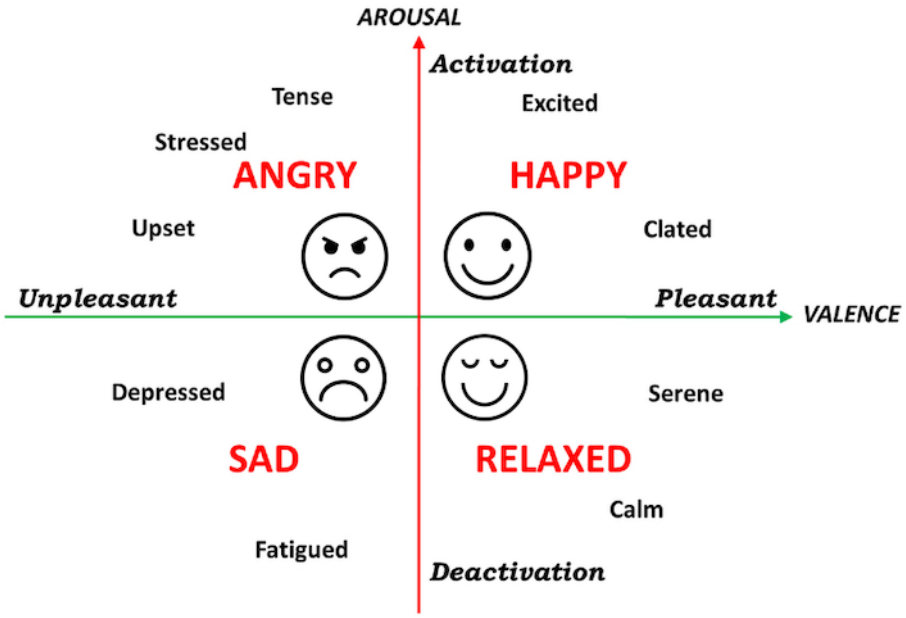

Figure 2.1: The circumplex model of affect is often depicted as a two-dimensional graph with valence (positive or negative emotion) on the x-axis and arousal (level of excitement or activation) on the y-axis.

arate categories. This is because, according to Russell, human beings do not perceive emotions as distinct, separate entities, but rather as ambiguous and overlapping experiences. In order to further refine the model, a third dimension, called Dominance, was added. This dimension ranges from a complete lack of control or influence over events and surroundings to the opposite extreme of feeling influential and in control. There are several other models that have been developed, including the 2D Plutchik's Wheel [22] and the 3D PAD model [23]. These models offer alternative ways of understanding and categorizing emotions, but the Russell Circumplex Model remains a widely recognized and respected framework for studying emotions.

One practical way to evaluate the emotional state of an individual is through the use of self-reports. The Self-Assessment Manikin (SAM) is a well-established model for this purpose [24] (See 2.2). It consists of three scales: valence (which reflects the pleasantness or unpleasantness of an emotion), arousal (which reflects the level of tension or relaxation associated with an emotion), and dominance (which reflects the level of inhibition or uninhibition associated with an emotion). Each scale includes five pictograms, and participants can choose the blank space between each pictogram to indicate intermediate states. This means that responses to each scale can range from a minimum of 1 to a maximum of 9, or from 1 to 5 without spaces. The SAM is a useful tool for assessing affect, as it allows individuals to communicate their emotional experiences in a clear and concise way.

2.1.3 Neurophysiological basis of emotions

Emotional situations activate specific neural circuits that were first identified in animals through techniques such as fear conditioning or direct brain stimulation (e.g., [25][26]). It is generally accepted that emotions serve to ensure the survival of individuals [13][27]. From this perspective, the associated neurophysiological reactions serve two main purposes: increasing the sensitivity of

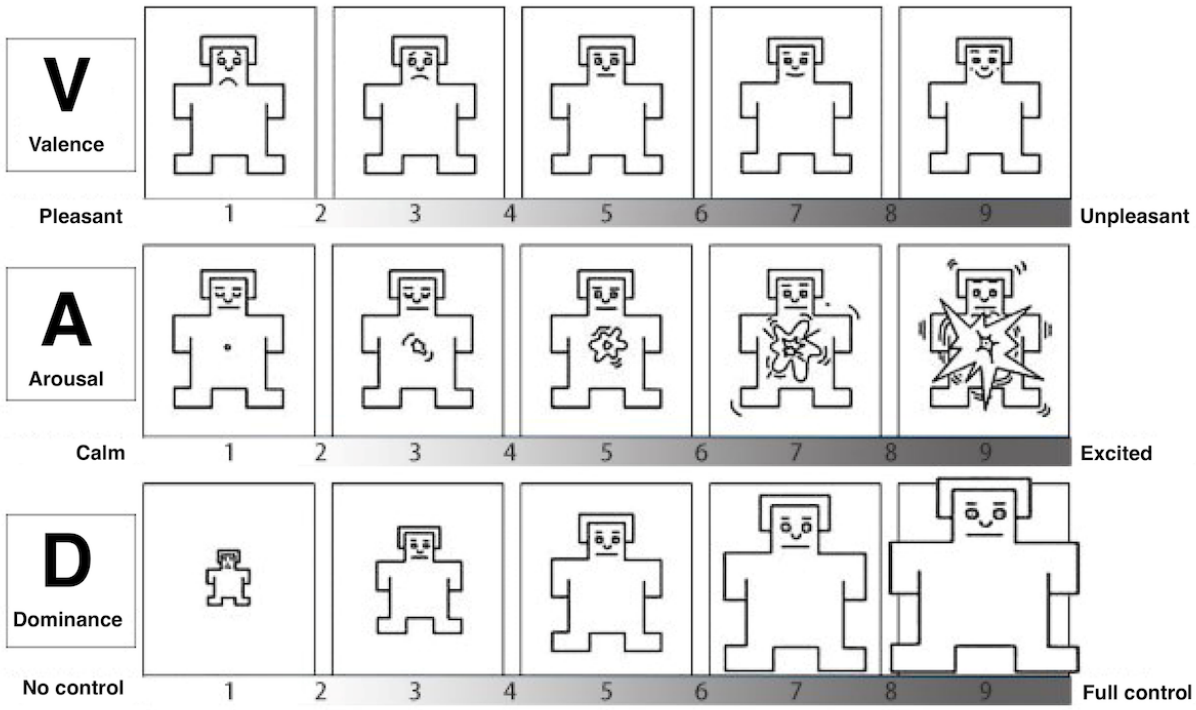

Figure 2.2: The self-assessment manikin (SAM). The first row is used to assess valence (the pleasantness of a stimulus), the second row is used to assess arousal (the intensity of emotion) and the third row represents Dominance (the degree of control exerted by a stimulus).

attention and sensory systems, and preparing the body for a motor response. These reactions are crucial for the survival of the individual, as they help to increase awareness of the environment and prepare the body to respond appropriately to potential threats or opportunities.

The bilateral amygdalae, which are located in the temporal lobe, have a central role in the processing, integration, and modulation of sensory information and the preparation of appetitive or defensive actions. These neural structures receive sensory inputs from the cortex and thalamus and may integrate this information with memories from the hippocampus, outputting to the central nucleus of the amygdala to modulate the accuracy of sensory processing (vigilance) and the autonomic and somatic responses. The connection between the amygdalae and the ANS allows for the coordination of complex patterns of sympathetic and parasympathetic activation in response to different situations [27]. The amygdalae are essential for the regulation of emotions and behavior, as they receive and integrate sensory information and use it to guide the appropriate responses to the environment.

In addition to the neural structures that are involved in the processing of emotions and the regulation of behavior, certain structures are responsible for the processing of reward and pleasure in response to appetitive stimuli. One such structure is the mesolimbic dopamine system, a dopaminergic pathway that connects the ventral tegmental area to the nucleus accumbens (part of the ventral striatum) [28]. This system has been extensively studied in both animals and humans and is known to play a significant role in the processing of reward and pleasure [29]. It has been shown to be activated by pleasurable experiences such as food, sex, and drugs of abuse, and the

activation of this system can lead to the release of dopamine, which is associated with feelings of pleasure and reward. These findings emphasize the importance of the mesolimbic dopamine system in the processing of reward and pleasure and its role in motivation and behavior. Overall, the neural structures and pathways involved in the processing of emotions, behavior regulation, and reward and pleasure are complex and interconnected, and they work together to enable individuals to adapt to and navigate their environment.

There is evidence that other neural structures and neurotransmitter systems may also be involved in the processing of emotions. For instance, the anterior cingulate cortex has been shown to be responsive to negative stimuli, particularly fear and anxiety [30]. It is clear that the neural basis of emotions is a complex and multifaceted process that involves the interaction of various neural structures and neurotransmitter systems. Understanding these interactions is essential for understanding the full range of emotional experiences and their underlying neural mechanisms.

2.1.4 ANS and Emotions

ANS is composed of three distinct anatomical and functional divisions, namely the Sympathetic Nervous System (SNS), its opposing counterpart, the Parasympathetic Nervous System (PNS), and the Enteric Nervous System (ENS), which is not relevant to the topic we are discussing. Both the SNS and PNS contain nerve fibers that transmit sensory information and motor commands to the CNS. Typically, the motor pathways of the SNS and PNS are made up of two neurons: a pre-ganglionic neuron located in the CNS and a post-ganglionic neuron situated in the periphery that connects to target tissues. In contrast, the ENS is a vast and largely independent structure, comprising over 100 million neurons of varying morphologies, that primarily regulates digestive processes [31].

The SNS is tonically active to maintain homeostasis and provides innervation to nearly all tissues in the human body. It facilitates "fight-or-flight" responses through two neurotransmitters, namely acetylcholine (presynaptic) and norepinephrine (postsynaptic). The sympathetic division prepares the body for stressful or emergency situations, inducing a heightened state of physiological arousal. This response entails an increase in heart rate, the force of heart contractions, and the dilation of airways to facilitate breathing. It also causes the release of stored energy and an increase in muscular strength. Additionally, it results in increased sweat production, particularly on the hands, pupil dilation, and hair standing on end. Processes that are deemed less important in emergencies, such as digestion and urination, are slowed down [32].

In contrast, the PNS promotes "rest and digest" responses, which are activities that occur when the body is at rest. This division, which is comparatively smaller than the SNS, is primarily composed of the vagus nerve, providing parasympathetic input to most thoracic and abdominal viscera [31]. By innervating the sinoatrial node, the vagus nerve facilitates cardiac relaxation by reducing stroke volume, heart rate, and blood pressure. The PNS conserves and restores physiological resources by decreasing heart rate and blood pressure and stimulating the digestive tract to process food and eliminate waste. Energy obtained from processed food is utilized to rebuild and repair tissues [32]. ANS schema can be seen in Figure 2.3.

Emotions are complex experiences that affect both our body and mind. They impact our ANS,

which controls many involuntary bodily functions. The two branches of this system, sympathetic and parasympathetic, coordinate our responses to different emotions.

Studies have found that different emotions are associated with distinct patterns of sympathetic and parasympathetic activation. For instance, sadness can result in either increased sympathetic control or sympathetic withdrawal, depending on whether someone is crying or not. Relaxation is linked to sympathetic withdrawal and parasympathetic activation, while happiness involves increased cardiac activity. Fear is a complex emotion that can lead to either vagal deceleration or sympathetic activation, depending on the situation. Anger can lead to reciprocal sympathetic activation and faster breathing, while anxiety is characterized by sympathetic activation and vagal deactivation. Disgust can result in sympathetic-parasympathetic co-activation and faster breathing. Finally, love, tenderness, and sympathy have been found to be associated with decreased heart rate and an unspecific increase in skin conductance level [33].

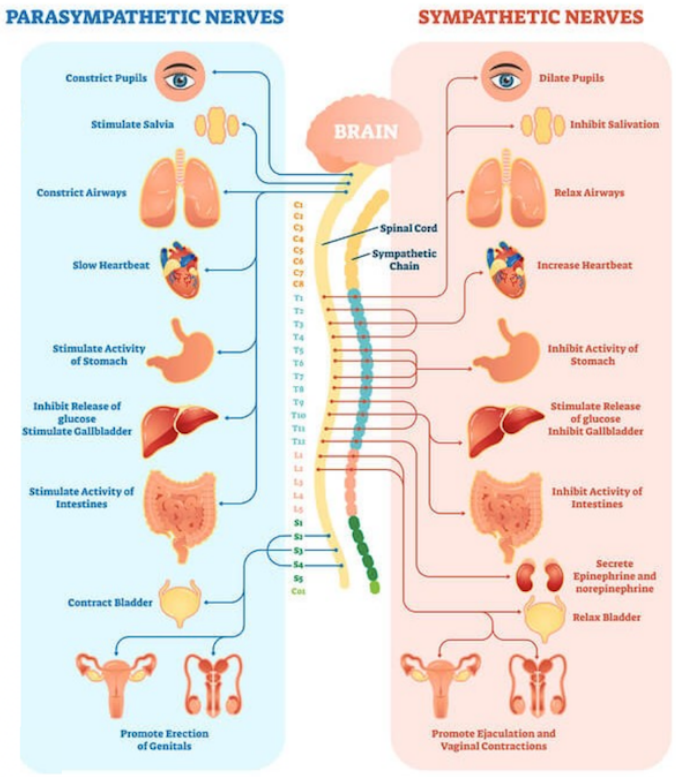

Figure 2.3: Pictorial representation of ANS.

2.1.5 Relevant findings about emotions and physiological responses

Peripheral signals and electroencephalogram (EEG) can be useful in identifying human emotions. The power spectra of EEG signals are commonly examined in various frequency bands to explore their connection to emotional states, and EEG has been frequently used in emotion detection research. In [34] researchers discovered a connection between high valence, high frontal alpha power, and high right parietal beta power. Another investigation [35] analyzed the relationship between parietal-occipital alpha power and arousal ratings using film clips, finding that alpha power

decreased when comparing films with high and low arousal ratings. In addition, the study found that theta power changes were correlated with emotions like interest, joy, and amusement.

Peripheral signals, such as cardiovascular signals, galvanic skin response (GSR), respiration (RESP), and electromyogram (EMG) can also provide insights into the influence of emotions on the autonomic nervous system. These signals can be useful for studying emotional responses in various contexts.

Heart rate variability (HRV), a measure of the variations in the interval between successive heartbeats influenced by both the sympathetic and parasympathetic branches of the ANS, is a highly sensitive indicator of neurobehavioral processes [36]. Heart-related biosignals like HRV have been found to be useful in this regard. The sympathetic branch increases heart rate while the parasympathetic branch decreases it, with these two branches acting in opposition to each other. Emotional valence and arousal level have been found to be correlated with the response of the heart according to research [37][38][39]. Typically, when the parasympathetic nervous system is more active, the heart rate tends to be slower, indicative of a relaxed state or the presence of enjoyable stimuli. Conversely, when the sympathetic nervous system is activated, such as in response to negative stimuli like hunger, fear, pain, or anger, the heart rate tends to increase. This has been demonstrated through experiments where heart rate increases more during fear-related imagery compared to neutral imagery or the repetition of neutral or fearful sentences [40]. Heart rate measurements have also consistently shown that anger, fear, and sadness are more closely linked to heart rate acceleration than disgust, while happiness falls somewhere in the middle [41]). Additionally, research has found that heart rate acceleration is greater during imagery involving disgust, joy, and anger compared to pleasant imagery [42].

GSR, also referred to as electrodermal activity, is a measure of changes in sweat gland activity that is controlled by the sympathetic nervous system. This response can be triggered by emotional stimuli and is often related to an individual's level of arousal. In other words, sweat reactions can be induced by emotional arousal and are closely tied to an individual's level of excitement or intensity. Fear and disgust have been found to be more strongly associated with GSR than happiness [43]. Comparing to other emotions, fear has been shown to produce a higher level of both tonic and phasic components of GSR [40]. These findings suggest that fear and disgust may have a greater impact on the body's sweat gland activity, which can be measured through GSR.

Emotions can impact respiratory activity, and research has demonstrated that combining electrocardiography and respiration volume signals can be effective in identifying both the level of arousal and valence, with an accuracy of 74% [44]. In other words, the combination of these two signals has been found to be quite effective in determining the intensity and positivity or negativity of an emotional experience. Respiratory activity is primarily controlled by the brainstem, but the limbic system, which plays a significant role in emotional regulation, can also have an impact on the final output of respiration. In other words, while the brainstem is the primary regulator of breathing, the emotional centers of the brain can also influence the way we breathe. Different emotional states can be associated with specific respiratory patterns. For example, slow breathing can indicate relaxation, deep and fast breathing can suggest happiness or anger, brief interruptions in respiration can be a sign of tension, and shallow and fast breathing may be indicative of fear [34].

The impact of valence, or positivity or negativity, on respiratory patterns has been studied in the literature. One study [45] found that emotions with high valence, or positivity, were characterized by steady respiration with uniform respiration magnitude, while emotions with low valence, or negativity, differed in their respiratory patterns. These findings suggest that respiration patterns can provide valuable information about both the valence and arousal dimensions of emotional states. However, it is important to note that the connections between respiratory patterns and emotions are not always straightforward, as individuals can express their emotions in different ways through their breathing.

In order to detect emotions, techniques often involve the use of EMG signals recorded from facial muscles to identify expressions that are triggered by emotional stimuli. Additionally, EMG signals recorded from the trapezius muscle may be used to assess any head movements or muscle tension that may be caused by a stressful situation [46]. By analyzing these signals, it is possible to detect emotions and understand how an individual is reacting to different stimuli. Research has shown that EMG signals from the corrugator supercilii muscle, which is responsible for frowning, and the zygomaticus muscle, which is responsible for smiling, can be reliable indicators of negative and positive emotional states, respectively [46][47]. In particular, the corrugator supercilii EMG amplitude has been found to increase in response to negative stimuli and decrease in response to positive stimuli compared to neutral stimuli, while the zygomaticus EMG amplitude has been found to increase in response to positive stimuli but does not discriminate between neutral and negative emotions [47]. This suggests that these facial expressions may convey information about an individual's feelings and attitudes, with frowns generally being perceived as negative and smiles as positive. Other studies have also demonstrated that muscle activity increases during negative valence emotions [48] and that EMG activity is a good indicator of high arousal in general [49]. These findings highlight the important role that facial expressions play in social interactions and communication, and may be useful in understanding how people process and interpret emotional cues from others [50].

Moreover, anger has been found to cause a larger increase in finger temperature than fear [43]. This has been demonstrated through experiments where anger has been shown to increase temperature, while fear has been shown to decrease it. These findings suggest that there may be notable physiological differences between these two emotions, with anger potentially having a more warming effect on the body compared to fear. Further research on this topic could provide additional insight into the ways in which different emotions impact the body.

2.2 Emotions and hearing

2.2.1 Emotions and music

Although the emotional sphere of sounds has been investigated to a much lesser extent than visual stimuli in the field of emotion recognition from physiological signals, there are studies in the literature that have used emotional protocols with sounds and music.

Research has shown that emotional reactions triggered by music are similar to spontaneous emotions in terms of quality, as music can create specific patterns of physiological changes such as heart

rate and blood pressure that are associated with specific emotions. One important study in this area was conducted by Kim and André in 2008 [51], which looked at the key stages of an automatic emotion recognition system using physiological measures. The research focused on male subjects experiencing four emotions: joy, anger, pleasure, and sadness, and achieved an average recognition accuracy of 87%, higher than in previous studies. While the evocative power of music in producing genuine emotional states has not been widely examined, the work of Kim and André adds to the growing evidence that music can indeed elicit genuine emotional responses. Kim and André employed musical induction as a method to evoke emotions in their research, which involved recording subjects' physiological signals while they listened to various musical pieces. They encouraged subjects to choose songs that would help them recall special memories related to the target emotions, as music has a way of accompanying significant events in human social life and is often linked to personal memories. Additionally, music listening often occurs in solitude, reducing the impact of social masking artifacts, and individuals' emotional responses to music can vary greatly based on their past experiences and cultural background. Research has shown that emotional responses to music can be influenced by an individual's musical preferences.

Different studies support the position of music as a good stimulus for eliciting emotional responses accompanied by differential changes in ANS reactions. For example, in [52] researchers conducted a study in which they analyzed differences in ANS responses triggered by live performances. During the Bayreuth Festival in the summers of 1987 and 1988, subjects listened to Wagner's works while two physiological measures, electrodermal response and respiratory activity, were recorded and evaluated for emotional excitement. The researchers found significant differences in the physiological measures during listening and in the musical characteristics such as melody, rhythm, and continuity. The work of Kim and André is further supported by this and other research demonstrating the impact of music on ANS reactions.

Some studies have also found results in clinical and therapeutic contexts related to the impact of music on ANS responses. For example, it was discovered that music can excite ANS responses such as vascular constriction, heart rate, muscle tension, and skin temperature (SKT) of the fingers, even if subjects report feeling less anxious and more relaxed [53]. Krumhansl et al. recorded various physiological measures while listeners listened to music that had been judged to be evocative of happiness, sadness, or fear, and analyzed the relationship between the physiological responses and the dynamic ratings of emotions [54]. They found that the directions of the physiological changes were the same for all three emotions, but the changes showed distinct patterns depending on the emotion being examined. For instance, happiness was linked to the greatest changes in respiration, sadness was associated with the greatest changes in heart rate, blood pressure, and SKT, while fear was connected with the maximum change in blood flow velocity. This suggests that music does not simply convey emotions that are easily recognizable, but rather actively generates genuine emotions in the listener. However, it is still unclear if the changes in the ANS and the distinctions in musical emotions align with those observed in emotions that are not related to music.

More research is needed to fully understand the relationship between musical and nonmusical emotions and how they may differ or overlap.

In [55] 45 sounds were used to elicit emotions and were divided into three categories based on

their valence: neutral, pleasant, and unpleasant sounds. The results showed that unpleasant sounds led to a heart rate deceleration, while emotionally arousing sounds (both pleasant and unpleasant) caused an increase in skin conductance response. It was also discovered that the skin conductance response was positively correlated with subjective arousal, which was higher for the unpleasant and pleasant sounds compared to the neutral sounds. These findings suggest that an individual's anxiety trait can influence their psychophysiological reactions to emotionally charged sounds.

Moreover, in [39] during the experiment, participants were seated in a comfortable chair in a controlled environment and listened to sounds while the electrocardiogram (ECG) was continuously acquired using a dedicated hardware module following the Einthoven triangle configuration. Each subject was left alone in the room where the experiment took place for a total of 29 minutes. The results of this study indicated that certain ANS measures, such as the HRV mean value, standard deviation, RMSSD and spectral measures, are the most effective HRV features for identifying emotional states induced by affective sounds.

While some studies have demonstrated the potential of emotional sounds as stimuli for exploring emotions, further research is needed to more fully validate their usefulness in this area. This could open up the possibility of using auditory stimuli in other areas of psychophysiology as well. In order to fully understand the potential and limitations of emotional sounds as stimuli, it is important to continue studying and evaluating their effects on physiological and psychological responses.

2.2.2 Emotional stress: Listening effort

Listening effort, or the cognitive and emotional effort that a person puts into listening to and understanding another person, can significantly impact communication, particularly for certain groups such as those with hearing impairments, nonnative speakers, the elderly, and those working in noisy environments [56]. The feeling that listening is too effortful may lead to tuning out of a conversation, abandoning a hearing aid, or avoiding social events in noisy venues [57][58][59]. The effort required for listening can also increase stress and fatigue, which can have negative consequences for mental and physical health in the long term [60][61].

Listening effort can be considered an emotion in itself, as listeners experience a range of emotions while listening, including frustration, confusion, and enjoyment [62]. Additionally, listeners who perceive a high level of listening effort are more likely to report feeling negative emotions such as boredom or annoyance [63]. The relationship between listening effort and emotions is complex and multifaceted, with both cognitive and emotional factors influencing the listener's experience. For example, the emotional experience of listening effort can be affected by the cognitive demands of the task, such as the clarity of the speaker and the listener's prior knowledge on the topic. At the same time, the listener's emotional response to the communication process can also impact their perception of listening effort [63].

According to [57], listening effort involves the engagement and control of various neural systems for information processing as well as the emotional response to the expenditure of those resources in a given context. Therefore, listening effort encompasses both cognitive and emotional aspects, reflecting the listener's experience and evaluation of the communication process [56].

Overall, the relationship between listening effort and emotions is dynamic and bidirectional,

with both cognitive and emotional factors influencing the listener's experience.

2.2.3 Overview on Listening Effort

Effort is the conscious allocation of mental resources to overcome challenges and achieve goals while performing a task. When it comes to tasks that involve listening, this effort is specifically referred to as listening effort. It involves the use of mental resources to overcome any obstacles or difficulties that may arise while engaging in the act of listening. In other words, it is the effort involved in order to focus and actively listen, rather than just passively hearing the information being presented [61]. In order to effectively understand, remember, and respond to the auditory stimuli that we perceive when the quality of auditory input is compromised, whether due to impaired auditory abilities or unfavorable acoustical conditions, it may be necessary for individuals to allocate a greater amount of cognitive resources. This can involve using mental processes such as comprehension and memory to fully grasp and retain the information being presented through sound. In other words, the act of perceiving auditory objects and events may require more mental effort in order to fully process and comprehend them. This can involve using mental resources such as attention and memory to fully understand and remember what has been heard.

Traditionally, audiological assessment has relied on pure-tone and speech audiometry to evaluate an individual's hearing abilities. Speech audiometry specifically involves measuring the lowest level at which individuals can hear speech and their ability to understand spoken materials at suprathreshold levels. While these measures are important, they may not fully capture the difficulties and challenges that some individuals experience when it comes to listening. Many people report that although they can hear sounds and understand speech, it is tiring or difficult to listen, even when the sounds are loud enough and the speech is accurately recognized [61].

Even when the speech being listened to is highly intelligible and has linguistic context to support it, individuals with normal hearing can still engage in cognitive processes that require effort when faced with acoustically challenging speech, as demonstrated in [64]. Despite having normal hearing, these listeners may still need to put forth extra effort to understand and comprehend the speech they are hearing. This highlights the importance of considering the impact of acoustical challenges on listening comprehension, even for individuals with typical hearing abilities [65].

The most accredited theory of listening effort is the motivational one, introduced in [66], which posits that individuals are motivated to conserve resources that are important for their survival and that effort is the investment of these resources in order to achieve goals. In order to minimize the waste of resources, individuals should only invest the amount of effort that is necessary for task success [59]. According to this theory, the relationship between task demand and effort investment is limited by the resource conservation principle. As the demands of a task increase, the required effort should also increase, up to a certain point. However, the upper limit of this relationship is determined by an individual's ability and the importance of success in the task. If success is deemed impossible, it would be a waste of resources to invest any effort in the task. Therefore, according to motivational intensity theory, effort investment should be a function of task demand, but only if success is possible and the required effort is justified by the importance of success.

However, it is important to keep in mind that there are other factors that relate to listening

effort, including (but not limited to) fatigue [67], and psychosocial considerations [61]. It is important to not confound listening effort with listening fatigue which is simply fatigue resulting from the continued application of effort during difficult listening tasks. When an individual becomes excessively fatigued, they may be less inclined to sustain a high level of effort due to decreased engagement. This reduction in effort can lead to further fatigue, potentially leading the individual to abandon the task in order to avoid experiencing further fatigue [61]. It is important to take steps to prevent or manage fatigue in order to allow individuals to maintain sufficient levels of effort and engagement.

As listening effort continues to be recognized as a crucial element for individuals who utilize hearing aids, it has garnered attention in audiology clinics for both its impact on intelligibility and its potential to influence an individual's decision to adopt and persist in using hearing aids [68]. This connection between listening effort and hearing aid usage has become more prominent in recent years. The effectiveness of hearing aids in enhancing intelligibility is well-established, the extent to which they can alleviate listening effort and mental fatigue is less certain. Despite this, it is important to continue to research and understand the relationship between hearing aid usage and these factors in order to provide the best possible support and care for individuals who use hearing aids [60]. Moreover, prolonged and strenuous listening efforts have the potential to result in negative consequences such as stress, fatigue, and even serious health issues including hypertension and a heightened risk of stroke [69]. These negative effects can occur over an extended period of time as a result of chronic effortful listening. It is important to address and address these potential consequences in order to promote overall health and well-being.

Despite the common experience of listening being effortful and stressful, clinical measures of listening effort have not been widely available in audiology. This has made it difficult for audiologists to accurately assess and address the challenges that some individuals face when it comes to listening, even when their pure-tone and speech audiometry results are within normal limits.

Many audiologists have struggled indeed to find a clear and standardized method for measuring listening effort, leading to frustration and difficulty in accurately assessing it. This has been noted by several researchers, including [70][57]. The lack of a clear and reliable way to measure listening effort has made it difficult for audiologists to accurately assess the mental effort that individuals put into listening tasks, and to understand how it may be affected by factors such as hearing ability and the quality of the auditory input.

2.2.4 Tests used for assessing Listening effort

Listening effort is an important aspect of auditory processing that can be evaluated through a variety of tests. One of the most commonly used tests is the speech in noise test, which measures an individual's ability to understand speech in the presence of background noise. This test is important because it reflects real-world listening situations and can provide valuable information about an individual's communication skills in noisy environments. Other tests include listening to words or sentences embedded in various types of noise that must be recognized by the subjects being examined. In [71], for example, three degraded conditions were examined in order to assess auditory processing abilities. The first condition involved unmasked speech produced by a computer speech

synthesizer, while the second and third conditions both involved speech produced by a natural voice that was masked by either speech-shaped noise or two-talker babble. These masked conditions were presented at a signal−to−noise ratio (SNR) of −8 dB, a level that has been shown in previous research to result in similar levels of performance when using these particular stimuli and maskers.

The reading span test is often used to evaluate working memory, as it is designed to simultaneously assess both memory storage and processing abilities. Subjects were required to comprehend a series of sentences and recall either the first or final words of the presented sequence. This test is commonly used to assess the individual's ability to hold and manipulate information in memory while also processing new information [72].

Regarding speech in noise tests, these are used to find speech reception thresholds (SRTs) which are measures of the minimum level of SNR at which an individual can understand speech with a certain level of accuracy at different intelligibilities. SRTs serves to determine noise thresholds perceived by the subjects which allow tests to be customized and SNRs to be selected to create various levels of difficulty. In [73], for example, three Speech Reception Threshold tests were conducted in stationary noise in order to estimate the SNRs that were required for the correct repetition of a certain percentage of sentences. Specifically, the SNRs required for the correct repetition of 50%, 71%, or 84% of the sentences were measured (SRT50%, SRT71%, and SRT84%, respectively). In this way three different difficulty levels were obtained.

2.2.5 Behavioral markers

Self-report measures of listening effort often come in the form of a closed-set questionnaire [74] or rating scale (e.g. [70][75]). An example of the questionnaire method can be found in the multidimensional speech, spatial, and qualities hearing scale [74] which measures the extent of listening difficulties experienced in various real-world settings. Included in the original 49-item inventory are questions that relate to the effort required in everyday listening situations. Participants give a response on a scale of 0 to 10 with lower numbers indicating more difficulty or effort [57].

In general there are a variety of ways in which individuals can respond to auditory stimuli. For example, one method is by verbal identification of the heard word or sentence and the percentage of words correctly repeated [76]. Another approach is through the use of a response button, as demonstrated in [77]. In addition to simply determining whether an individual is able to correctly identify or respond to auditory stimuli, the speed of their response may also provide further insight into the listening effort involved in speech perception. This idea has been proposed by researchers such as in [76] and in [77]. By measuring response speed in addition to accuracy, it may be possible to gain a more comprehensive understanding of the cognitive demands and effort associated with processing spoken language.

One advantage of using self-report measures to assess listening effort is the simplicity and ease of administration. These measures can provide valuable insight into an individual's perceived level of effort during speech processing, but it is important to bear in mind their subjective nature [57]. For instance, there may be differences in how individuals define and experience effort, and some people may base their self-report ratings on task performance rather than mental exertion. Additionally, research has suggested that older adults may tend to underestimate their perceived listening effort

compared to younger adults, despite potentially experiencing similar difficulties in a listening task [78]. Therefore, relying solely on self-report measures may not always provide an accurate reflection of the mental effort required to comprehend an auditory message.

2.2.6 Physiological markers

Physiological measures involve the recording of changes in the activity of the central and/or autonomic nervous system during a specific task. Researchers have used a variety of methods, including functional magnetic resonance imaging (fMRI) and EEG to examine changes in CNS activity and investigate the impact of listening effort on the brain. Additionally, some studies have also looked at changes in ANS activity, such as GSR and pupil dilation (PUPIL), as potential indicators of listening effort.

According to fMRI when individuals were asked to focus on degraded speech, researchers observed increased activity in the left inferior frontal gyrus, as compared to when they listened to clear speech. This finding suggests that the brain may need to engage additional cognitive resources in order to comprehend spoken language that is less intelligible [79].

EEG measures offer a high level of temporal precision for tracking mental processing, with a temporal resolution at the millisecond level. This makes EEG a valuable tool for researchers studying the dynamics of brain activity, as demonstrated in [80][81][82]. The temporal resolution of EEG allows for a detailed analysis of brain activity over time, providing insight into the sequential steps involved in mental processing. Studies conducted in recent years have indicated that when the quality of an auditory signal is reduced through a process called vocoding, alpha oscillations in the brain increase during listening. This suggests that alpha power is not only influenced by domain-general factors such as the amount of information that needs to be retained, but also by challenges arising from varying levels of sensory degradation. In other words, even mild impairments in the sensory system can affect the alpha oscillation system [61]. Moreover, the frontal midline theta, a brain wave activity largely influenced by sources localized in or near the medial frontal cortex, was found to have increased power as the SNR decreased and was positively correlated with self-reported effort [83].

Regarding GSR, according to [84] digit recall accuracy is generally similar across various task conditions, but there are significant differences in the sweating responses depending on the level of demand of the listening condition. Specifically, GSR responses tend to be higher in medium- and high-demand listening conditions. In one experiment, indeed, a majority of participants showed significant changes in GSR from low- to medium- and high-demand conditions, indicating that the cognitive demands of the task may have an impact on ANS activity. Moreover, in [85] GSR reactivity (i.e., the increase in GSR relative to baseline) was found greater for the fast speaking rate than for the normal speaking rate but it was not found to fluctuate during SNR or repetition scores decrease, indicating a possible link with task demand but not with the variation of noise in vocal signals.

As noted in [86], several studies have found that HRV decreases as the demands of a listening task increase. This trend has been observed for various HRV measures and may therefore serve as a reliable indicator of the effort required for listening. Moreover, decrease in HF-HRV (reflect-

ing parasympathetic nervous system withdrawal) was observed with an increase in speaking rate indicating sensitivity of this measure to increased task demand.

The Pre-ejection period (PEP), which represents the time between the initiation of contraction in the left ventricle of the heart and the opening of the aortic valve, has primarily been utilized in research on the effects of manipulations in the motivation dimension rather than the demand dimension. In a study on PEP reactivity [87], it was found that this measure was higher in a condition of high demand and high importance of success compared to other three conditions of different task demand and motivational level.

Pupillometry, a method for measuring pupil size, has been utilized in numerous research studies to explore the relationship between the demands of a listening task and changes in pupil size. Pupil dilation appears consistent with subjective notions of relative difficulty even in these equally intelligible cases [88]. In [89] authors found that the pupil response systematically increased with decreasing speech intelligibility, emphasizing the importance of this physiological variable for the assessment of listening effort. Moreover, older adults with hearing impairments demonstrated a smaller decrease in pupil size from difficult to easier listening conditions compared to younger adults with normal hearing. This difference may be due to the fact that older adults tend to have smaller absolute pupil sizes, which could limit the ability to detect significant changes in pupil size. Overall, despite the valuable insights gained from these studies, accurately interpreting the meaning of variations in pupil size across different populations still remains a challenge.

2.2.7 Research gaps on Emotions and hearing

There are numerous gaps in our understanding of the emotional responses generated by auditory stimuli, their relationship with physiological signals, and the potential importance of these stimuli in the field of psychophysiology. While research on auditory emotion recognition has increased in recent years, much remains unknown about the specific mechanisms by which auditory stimuli influence emotional responses. For example, little is known about the role of specific frequencies or patterns of sound in evoking particular emotions, or how these stimuli interact with other physiological signals such as heart rate and skin conductance. Further research on the emotional responses generated by auditory stimuli has the potential to provide valuable insights into the nature of emotions and their underlying physiological mechanisms, as well as inform the development of effective treatments for emotional disorders.

There has been also much discussion and debate within the field about the most appropriate and effective way to measure listening effort. The complexity of listening effort means that there is no single measure that can capture all of the different dimensions and factors that contribute to the experience of listening. This is why it is important to consider the specific characteristics of the task and the aspects of increased listening demands that are relevant to the specific context when selecting a method to assess listening effort. It is therefore necessary to consider a range of measures in order to fully understand and evaluate the experience of listening effort, and that these measures should not be treated as interchangeable or equivalent. To the best of our knowledge, there are studies that examine the relationship between physiological signals, which are considered more objective compared to qualitative measures, and listening effort. However, what is lacking are

studies that incorporate both central and peripheral signals in order to provide a more comprehensive understanding of how the body and physiology respond to auditory stimuli that may be challenging to comprehend.

2.2.8 Study objectives (AIM 1)

The primary objective of this study is to examine the relationship between emotional musical stimuli and physiological responses using the AuBT online dataset [90]. By analyzing the data, we aim to better understand how and if music is capable of eliciting emotions and the specific mechanisms underlying these responses. By adding to the existing literature on the relationship between music and physiology, we hope to provide valuable insights into the role of music in emotion regulation. By using this dataset and advanced signal processing techniques, we aim to add to the current body of knowledge and contribute to the field of music and emotion research.

Our second goal was to create a speech-in-noise test that was simple and brief in order to avoid inducing fatigue in subjects that could compromise the measures and be confused with listening effort. In the protocol we developed, both autonomic and central signals were acquired in order to better investigate how our body responds to different levels of difficulty in listening to auditory stimuli.

2.3 Emotional visual/auditory stimuli and emotion-related protocols for emotion recognition

2.3.1 Overview on auditory and visual stimuli in emotion-related protocols

One way to elicit emotional responses is to ask an actor to portray a specific mood or emotion [91]. This technique has been widely used to evaluate emotional reactions through facial expressions and, to some extent, through physiological signals [92]. However, it can be difficult to ensure that the physiological responses of non-actors are consistent and reproducible when using this method. Additionally, the emotions displayed by actors in emotion assessment databases may not accurately reflect the emotions experienced in everyday life [93]. An alternative approach is to present specific stimuli, such as images, sounds, videos, or video games, to a participant in order to induce emotional responses. This approach has the advantage of not requiring the use of a professional actor and of producing responses that are more representative of those experienced in real life situations.

The most commonly used emotional images in literature in emotion recognition experiments come from the International Affective Picture System (IAPS) which is a collection of images commonly used in psychological research to induce emotional states in participants [94]. The collection includes over 2,000 images, each of which has been rated based on its valence (i.e., the degree of positive or negative emotional response it elicits) and arousal (i.e., the intensity of the emotional response it elicits). One advantage of the IAPS is that they provide a standardized way to induce emotional states in participants, which makes it possible to compare the results of different studies. Many studies in the field of emotion recognition have drawn upon the IAPS image dataset in order to evaluate the relationship between physiological signals and emotional content generated

by images. The IAPS dataset contains a wide range of images that are designed to evoke specific emotions in viewers, and researchers have used these images to study how the body responds to different emotional stimuli.

The Open Affective Standardized Image Set (OASIS) has been introduced more recently [95]. It is a freely available online collection of 900 color images depicting a wide range of themes, including humans, animals, objects, and scenes. These images have been rated on two affective dimensions: valence, which refers to the positive or negative affective response elicited by the image, and arousal, which reflects the intensity of the affective response evoked by the image. The OASIS images were sourced from the internet, and valence and arousal ratings were obtained through an online study involving a total of 822 participants.

Regarding sounds, the International Affective Digitized Sounds (IADS) is a collection of sounds commonly used in psychological research to induce emotional states in participants [96]. The collection includes over 1000 sounds, each of which has been rated based on its valence and arousal. Like the IAPS, the IADS provide a standardized way to induce emotional states in participants, allowing for comparison of results across studies. More recent studies focused on the use of more complex stimuli such as video-clips or film scenes which could better evoke specific emotions. In this regard, many datasets were created containing physiological recordings to be use for emotion classification aims. Following some examples of this datasets are reported.

The DEAP dataset contains data on human affective states, obtained by recording the EEG and physiological signals of 32 participants as they watched 40 one-minute long music video excerpts [97]. The stimuli for this dataset were selected from the Last.fm website, a platform that allows users to track their music listening habits and discover new music and events. A list of emotional keywords was compiled and expanded to include synonyms and variations, resulting in a total of 304 keywords. These keywords were used to search the Last.fm database for corresponding tags, and the top music videos most frequently labeled with each tag were selected. Of the 120 initially selected videos, they were each rated by at least 14 volunteers, and the final 40 videos for use in the experiment were chosen based on the strength and consistency of these ratings in order to maximize the intensity of the elicited emotions.

The CASE dataset offers a solution for analyzing emotions in real-time as participants view various videos from Youtube [98]. To do this, a joystick-based annotation interface was created to enable the simultaneous reporting of valence and arousal levels, which are usually recorded independently. Alongside this, eight physiological recordings were obtained such as the ECG, blood volume pulse (BVP), EMG, GSR, RESP, and SKT. These recordings were synchronized to provide insight into emotional states.

The MAHNOB-HCI database is a collection of multimodal data recorded in response to affective stimuli, with the goal of emotion recognition and implicit tagging research. To collect this data, a multimodal setup was used that included synchronized recordings of face videos, audio signals, eye gaze data, and peripheral and central nervous system physiological signals [99]. A preliminary study was conducted using an online affective annotation system, in which participants reported their emotions in response to the videos played by a web-based video player. Based on the results of this study, different video clips were selected that received the highest number of tags in different

emotion classes (e.g., the clip with the highest number of sad tags was chosen to induce sadness). Overall, in contrast to datasets such as IAPS and IADS, which are widely validated emotional stimulus datasets in the literature, there is no true equivalent video stimulus dataset that can be used as a gold standard. Instead, various studies have used ad hoc selected videos that are somewhat manually labeled and somewhat sourced from online platforms, which lack the reliability and robustness of the previously mentioned datasets.

2.3.2 Relevant studies on emotion recognition through physiological signals

Here will be presented some relevant studies on the relationship between physiological signals and the classification of emotions which were able to reach high accuracies. The ability to accurately classify emotions based on physiological signals has the potential to facilitate the development of new technologies for emotion recognition and to improve our understanding of the underlying processes that drive emotional experience.

The purpose of [100] was to use wearable computers to collect physiological signals from the ANS, such as GSR, heart rate, and temperature, from 14 subjects during an emotion elicitation experiment. These signals were then mapped to various emotions, including sadness, anger, fear, surprise, frustration, and amusement, which were elicited by movie scenes or math problems. The results showed that the Marquardt backpropagation algorithm was able to accurately categorize emotions with 84.1% accuracy.

To classify emotions, researchers in [101] selected four physiological signals and collected data from 60 undergraduate participants experiencing a target emotion (fear, neutral, joy) induced by film clips. The raw signals were processed to obtain 22 features, which were then analyzed using canonical correlation analysis, a method that combines dimension-reducing techniques and predictor variables. This approach allowed the researchers to overcome the challenge of high-dimensional classification and achieve a recognition accuracy of 85.3%.

In [39] an automatic classification system was tested on a group of 27 healthy volunteers who received standardized stimuli while their ECG signals were recorded. HRV features that exhibited significant changes in arousal and valence were used as input for the system. Results from the Leave-One-Subject-Out procedure showed that a quadratic discriminant classifier was able to accurately recognize valence at a rate of 84.72%, and arousal at 84.26%.

[102] aimed to determine the best algorithm for distinguishing negative emotions like sadness, fear, surprise, and stress using physiological data from 12 subjects. Physiological signals, including GSR, ECG, SKT, and BVP, were recorded and analyzed. The emotional stimuli used in the study were audio-visual film clips that had previously been tested for appropriateness and effectiveness in a preliminary experiment. The results of the emotion classification showed that the SVM algorithm had an accuracy of 100.0%.

There are other studies that have found higher levels of accuracy than the studies mentioned above, but they have not been extensively reported because they were subject-dependent, meaning that the experiment was conducted primarily on a single subject [51][103][49].

While the use of physiological measures has become increasingly popular for emotion recognition, there are still many limitations and challenges that need to be addressed. One major challenge is

the lack of consensus on the best protocols for collecting and analyzing physiological data. Different studies have used a wide range of methods and equipment, making it difficult to compare and integrate the results. Moreover, despite the high levels of accuracy achieved in machine learning, very little is said about how the models work. Even less is said about training and test sets, and whether entire subjects or just certain acquisitions are maintained on the test set. Additionally, little is said about which features are the most important and which signals may be the most effective for classifying emotions.

2.3.3 Beyond standard emotion stimulation strategies

The ability to evoke emotional states in a reliable and ethical manner is known as emotion elicitation. This is a critical step in the development of a system that can detect, interpret, and adapt to human affective states [104]. Emotion elicitation can be either active or passive. Active methods involve the use of tools such as behavioral manipulation, social interaction, and dyadic interaction to directly influence subjects. Passive methods, on the other hand, present stimuli such as images, sounds, or videos to evoke emotional responses. Passive methods are less realistic than active methods due to their lack of a "sense of presence", as they rely on non-immersive devices such as screens. Additionally, passive stimuli are non-interactive, which means that subjects cannot interact with the scene, limiting the ability to recognize emotional states during interactive tasks [105]. There is ongoing debate about the most effective way to elicit genuine emotions in laboratory settings, despite the fact that many studies have successfully used active methods to manipulate emotions. Active methods, which involve direct interaction with subjects, are generally more ecologically valid [104] and immersive [106] than passive methods, which present stimuli such as images or videos. However, despite the potential advantages of active methods, the majority of studies in the field have focused on passive methods. One possible explanation for this is the availability of a large body of literature with established standards, norms, and practices for conducting experiments using readily available passive stimuli to induce emotions [105]. In recent years, there has been a significant increase in the use of virtual reality (VR) as an active method for eliciting emotions in psychological studies as well as well as being a potent tool in numerous therapies.

There are many different definitions of VR, which generally overlap in certain key areas. When we use the term "VR", we specifically mean computer-generated immersive imagery. Many definitions also state that VR must be interactive, which distinguishes it from passive media such as 3D movies, 360-degree video, and others. VR can be divided into three categories [107]:

• Non-Immersive Virtual Reality: This type of VR technology involves a computer-generated virtual environment in which the user remains aware of and influenced by their physical surroundings. Video games are a common example of non-immersive VR.

• Semi-Immersive Virtual Reality: This category of VR provides an experience that is partially based in a virtual environment. Flight simulators for pilot training are an example of semi-immersive VR.

• Fully Immersive Virtual Reality: This type of VR generates the most realistic simulated experience, including sight and, less commonly, other senses such as smell, taste, and touch.

Thanks to all these aspects, VR technology is emerging as a promising solution for providing

medical assistance, both general and specialized. Furthermore, it has the potential to become a widespread technique for psychotherapy [108]. Several studies have revealed that VR exposure therapy is as effective as traditional in-person therapy in treating acrophobia [109], spider phobia [110], and fear of flying [111]. VR exposure therapy has also been utilized as a replacement for traditional imaginal exposure therapy for Vietnamese combat veterans dealing with posttraumatic stress disorder [112]. VR sessions have also been implemented in treating eating disorders and obesity, with the goal of transforming individuals' perceptions of their bodies [113].

It has been demonstrated that VR can create more realistic scenarios when compared to laboratory experiments, where the subjects may be less stimulated during different tasks. For example, the virtual version of the Multiple Errands Test (VMET) was developed and tested in different clinical populations. The VMET is a VR-based tool in which in a virtual supermarket, participants are requested to buy various products presented on shelves and to comply with different rules. In [114], VMET was used for studying markers for detecting Parkinson's subjects at risk for developing dementia. This VR tool could integrate the traditional neuropsychological evaluation of executive functions in Parkinson's subjects with a more ecologically valid evaluation. The VMET has shown to be more sensitive in the early detection of executive deficits compared to traditional measures.

In another study [115], a virtual class was developed for rehabilitating children with dyslexia. Tasks were performed by using a virtual blackboard while patients were sitting at a desk and looking at the blackboard. Significant improvements were observed during different weeks of the study.

Regarding the emotion field, a number of studies have utilized VR to stimulate emotions in controlled experiments [116][117][118] even if results are based mainly on qualitative assessment related to subjective ratings. Researchers in this field [119] for example successfully developed a computer-based emotion predictive model by eliciting active emotions via Immersive Virtual Environment. Their findings validated the use of VR in eliciting and recognizing emotions. Another similar study based on VR games was able to achieve satisfactory accuracy in classifying emotion for valence and arousal dimensions [120].

One of the major advantages of VR over traditional emotion elicitation methods is the concept of "presence", as noted by the authors in [121]. Presence, or the sense of being in a virtual environment, can be described as the perceptual illusion of not being mediated by technology. This is a key characteristic of a virtual experience and helps to differentiate it from traditional methods. According to [122], presence levels are correlated with the strength of experienced emotions, meaning that high levels of presence often result in strong emotions. VR technology is a simple and intuitive way to deliver consistent sensory information and interact with the virtual environment. There are numerous potential benefits to using VR for emotion recognition. One advantage of VR is that it allows researchers to carefully control the stimuli presented to participants, which can lead to more accurate and reliable results. Another benefit is that VR can provide a more immersive and engaging experience for participants, potentially resulting in more natural and spontaneous emotional responses.

To the best of our knowledge, there is a lack of research on the use of VR equipment in emotion studies, including the type of media content, screening processes, and the availability of public datasets. This highlights the need for a literature review and guide on VR-based emotion research.

2.3.4 Research gaps on emotion recognition

Affective computing, the recognition of emotions from physiological signals, is a growing field with many potential applications. However, there are still significant gaps in our understanding of how emotions are reflected in physiological signals, and how best to measure and analyze these signals. Many studies focus on achieving high classification performance, using methods such as feature extraction and principal component analysis (PCA), without fully considering the underlying physiological relationships. This approach may not be optimal for generalization and interpretability of results. Furthermore, most stimuli used in affective computing studies are video or film clips, which can make it challenging to disentangle the emotional contributions of the audio and visual elements. There have been relatively few studies examining the emotional content of sounds alone. To fully understand the complexities of affective computing, it is necessary to consider the diverse range of physiological and sensory factors that contribute to our emotional experiences.

2.3.5 Study objectives (AIM 2 & AIM 3)

One goal of the study could be to clarify the different physiological responses to visual, auditory, or combined stimuli in order to advance the literature on the relationship between different types of stimulation and physiological response. This could also help to understand which stimuli may be more effective and interesting for subjects, in order to extend their use to different fields of psychophysiology. By understanding these relationships, the study could potentially contribute to a greater understanding of the factors that influence physiological responses to stimuli. This information could be useful in a variety of contexts, such as in the design of psychological experiments or in the development of new techniques for measuring physiological responses. (AIM 2)

Another aim of the study is to create a new dataset for an experiment in emotion recognition based on physiological signals, conducted in virtual reality. Using a more immersive environment compared to passive stimuli could help to better evaluate the relationship between physiological signals and emotions, and understand if this type of environment is more effective in stimulating emotions in individuals. Additionally, the study could try to verify if it is possible to achieve higher classification performance compared to existing literature, using a more immersive virtual reality environment. The ultimate goal would be to advance the literature in this field, providing new data and methods for emotion recognition based on physiological signals. (AIM 3)

Chapter 3

Materials and Methods

3.1 Physiological responses to auditory stimuli

3.1.1 Development of speech-in-noise test

The purpose of the present experiment was to develop an adaptive speech-in-noise test that was both brief and straightforward, in order to minimize fatigue in subjects while also being reliable and effective in detecting hearing loss, as part of the characterization of biomarkers for listening effort.

Test stimuli

To facilitate the processing of vocal stimuli, this study employed non-semantic vowel-consonant-vowel (VCV) stimuli, such as "ama", "ata", and "asa". The use of VCV stimuli, which do not convey any meaning, is critical in effectively assessing adults' consonant recognition abilities because it minimizes the involvement of higher-level processing centers and reduces the impact of factors such as subjects' education, literacy, or native language on test results (as noted in [123] [124]). Additionally, the use of meaningless stimuli can also minimize the effort required to understand them. The speech stimuli used in this study included 12 consonants (/b, d, f, g, k, l, m, n, p, r, s, t/) pronounced with the vowel /a/ (e.g. "aba", "ada"), recorded by a professional native English speaker (as described in [125] [126] [127]). English was selected as the language for these materials due to its widespread usage as a primary and second language, as well as its frequent use on the internet (as outlined in [128]). To ensure that the average level of the recordings was equal, they were modified to meet the equal speech level requirement outlined in the ISO 8253-3:2012 standard (as established by the International Organization for Standardization in 2012). In addition, speech-shaped noise was added to the VCVs by filtering a Gaussian white noise with the international long-term average speech spectrum (as described in [129]) and a low pass filter with a cut-off frequency of 1.4 kHz and a roll-off slope of 100 dB/octave. This noise was then attenuated by 15 dB to create a noise floor, as suggested by [130].

Characterization of speech stimuli

To minimize anxiety and perceived difficulty, as well as other high-level cognitive efforts such as short-term memory and reading speed that may impact speech recognition performance, par-

ticularly in older adults, the stimuli were presented in a multiple-choice format. Specifically, a three-alternative forced-choice (3AFC) task was employed, with the three options displayed on the screen in a manner that maximized their opposition in terms of manner, voicing, and place of articulation (as discussed in [125] [126] [127]). The speech waveforms used in this study were modified by the addition of filtered speech-shaped noise ([131] [132]), and the SNR was varied between -50 and +20 dB at 2 dB intervals. Specifically the noise signal was obtained by filtering a Gaussian noise using the international long term average speech spectrum [129]. The intelligibility of these speech stimuli was then measured using the Short-Time Objective Intelligibility (STOI) measure [133], which calculates the correlation coefficient between the temporal envelopes of the reference (speech stimulus) and the processed speech signal (speech plus noise) in short segments of 386 ms. In order to account for the random nature of noise, 100 STOI simulations were conducted for each stimulus using different noise realizations. Previous research has shown a strong correlation between STOI-estimated psychometric functions and fixed-level experimental estimates in various languages [134] [135]. The STOI-based psychometric functions of the speech stimuli were estimated by fitting the average of the STOI values from the 100 simulations to a cumulative normal function, a commonly used model for describing the dynamics of speech audiometry scores. The mean and standard deviation of the cumulative normal function were then estimated using probit transformation [136] and weighted linear regression [132], without making assumptions about the estimated performance at extreme SNR values. In order to estimate the STOI-based psychometric functions for a 3AFC task, the psychometric functions were sampled at 0.25 dB SNR intervals and mapped into the range of 33% to 100%, given that there are three options, the probability of selecting the correct answer is not less than 33%. Figure 3.1 shows the main steps from speech stimuli to create the corresponding psychometric functions. In detail, each VCV stimulus and the same stimulus with added noise at different SNR were divided into temporal windows of 386 ms (a). Subsequently, the correlation was calculated for each synchronized temporal window between the two signals, and all of these correlation coefficients were averaged in order to find a value that represents the intelligibility of the vocal stimulus at a given SNR (b). The curve obtained from the correlation coefficients was then fitted with a psychometric curve that represents the probability of obtaining a correct response as a function of difficulty (c).

Staircase procedure

Due to their simplicity and flexibility, staircase procedures are frequently used in psychophysical research [137] [132] [138]. These procedures utilize responses to previous trials to determine the presentation level of the next trial, with the aim of presenting stimuli at a level where the probability of an upward step is equal to the probability of a downward step. For example, a one-up/one-down staircase targets the 50% performance level, while a one-up/two-down staircase targets the 70.7% performance level, and a one-up/three-down (1U3D) staircase targets the 79.4% performance level [139].

In staircase procedures, one of the fundamental assumption is that there is a positive correlation between the intensity of the stimulus and the level of performance so as if the stimulus increases in intensity, performance increases as well. It is commonly observed that the assumption of monotonic-

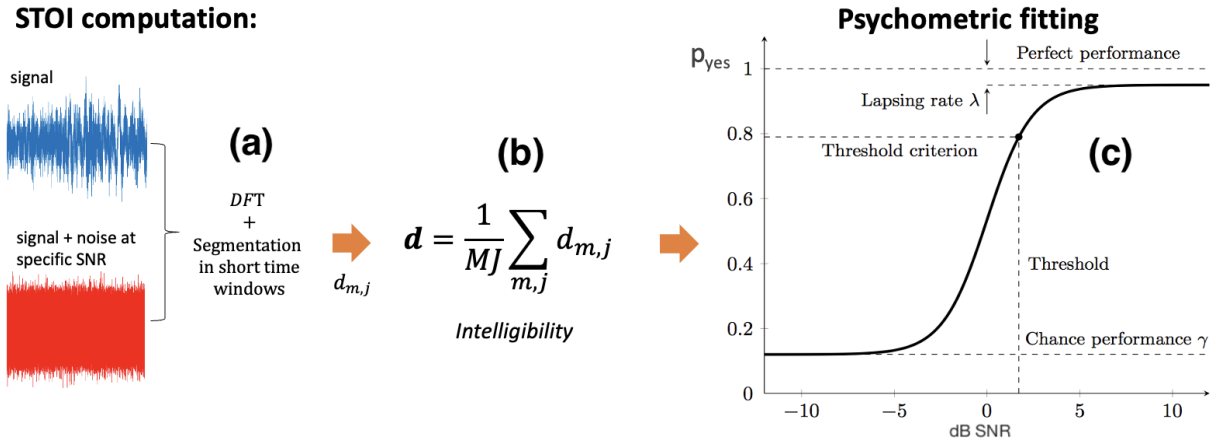

Figure 3.1: Main steps involved to pass from STOI computation of the pseech stimulus to the psychometric function

ity holds true for simple stimuli, such as pure tones, but may not apply to more complex stimuli, like speech, due to their greater diversity and multiple dimensions [140]. When dealing with complex stimuli, it is possible that different stimuli presented in subsequent trials may exhibit varying levels of performance even when the stimulus level is increased. This is because the relationship between stimulus and performance may not be consistent across all stimuli, leading to non-homogeneity in the psychometric function. This has been documented in previous research [139]. During typical speech-in-noise test, the non- homogeneity of the speech stimuli is often addressed through the use of level correction: the presentation level of the stimulus is increased by a certain amount, equal to the difference between the individual SRT of the stimulus and the average SRT of all stimuli in the set [131][141][142] even if the level correction may not fully compensate for inhomogeneity of speech stimuli [143]. In our case, we use a 3AFC task in a 1U3D configuration so that the SRT points at the target probability of 79.4% [139]. 1U3D, indeed, has been established to enhance test efficiency, precision, and convergence when utilized in conjunction with 3AFC [144][145]. This rule dictates that following one erroneous response (1U), the stimulus presentation level shall be augmented, whereas, following three correct responses (3D), it shall be reduced. Additionally, it should be noted that, at each trial, a random stimulus is selected from the set, and the same stimulus may be presented in consecutive trials.

Figure 3.2 shows the 12 curves related to the VCV stimuli in the top panel and the same curves after the level correction in the bottom panel. Specifically, in the bottom panel, each original curve was shifted on the SNR axis so that the target probability (i.e. 79.4%) corresponds to SNR equal to the average SRT of the VCV stimuli.

Conventional staircase procedures (CSP) are based on the assumption that the intelligibility of test stimuli is uniform across the set. As a result, the intelligibility is adjusted using fixed SNR steps: +2 dB increase (Δup) following one incorrect response and -2 dB decrease (Δdown) after three correct responses. After twenty reversals, the staircase procedure was terminated and the SRT was determined by calculating the average of the SNR midpoints of the final eight ascending runs, at a target probability of 79.4% [137].

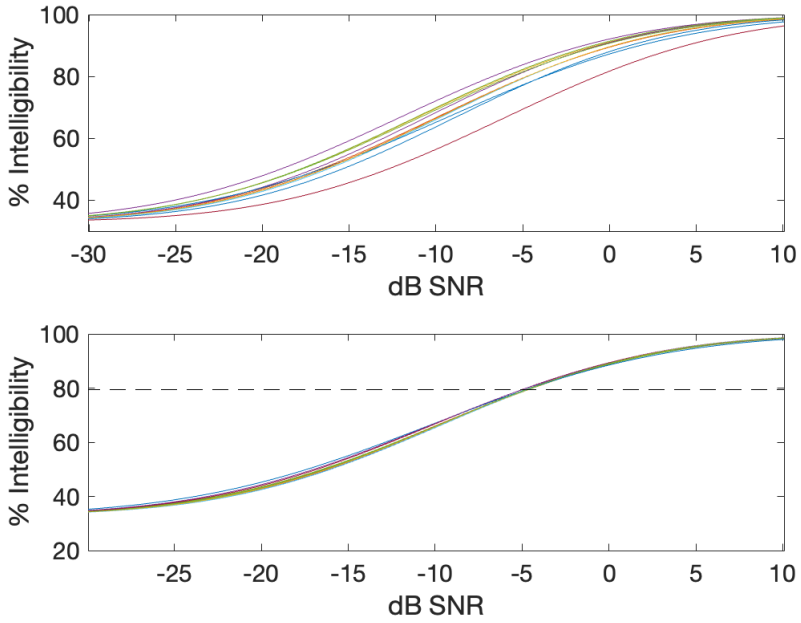

Figure 3.2: The top panel shows the 12 psychometric curves obtained by processing the VCV speech stimuli. The bottom panel shows the same 12 curves after the level correction. The black dotted line represent the percentage of intelligibility equal to the target probability of 79.4% for the 1U3D algorithm

The revised staircase procedure (RSP) instead takes into account the nonuniform intelligibility of the test stimuli within the set. Therefore, it is possible that a change in the SNR may not consistently produce the AFC adjustment in intelligibility. In fact, it is possible that an increase or decrease in SNR may result in a corresponding decrease or increase in intelligibility, respectively, due to inherent variations in intelligibility among the VCVs in the set.

To optimize the procedure, we determined an optimal ratio of Δdown to Δup. Specifically, the magnitude of the upward and downward changes in intelligibility, designated as Δup and Δdown, respectively, were based on the estimated intelligibility of the stimuli presented at each trial. Following an incorrect response, the next stimulus and SNR were chosen in such a way that the change in intelligibility as estimated using the STOI measure was approximately 7.7% for each upward step and approximately -5.7% for each downward step. These changes were determined through a preliminary analysis of the average slope of the STOI-based psychometric functions around the point of inflection, where most of the staircase trials are sampled. The changes were set such that the ratio of Δdown to Δup in the linear portion of the functions was approximately 0.7393, the recommended value for transformed and weighted 1U3D procedures [146]. Despite the nonlinearity of psychometric functions, an increase or decrease in percent performance may result in larger variations of SNR in the nonlinear areas of the functions and smaller variations in the linear region, which is desirable for higher accuracy. In order to avoid sudden changes in SNR in subsequent trials and decrease the potential for bias, a maximum step size was established for Δdown and Δup at 4 dB and 5.4 dB SNR, respectively. This guarantees that the step size stays within a suitable range regardless of the nonlinearity of the functions. The SRT was determined by evaluating the mean intelligibility of the central segments of the final four ascending sequences.

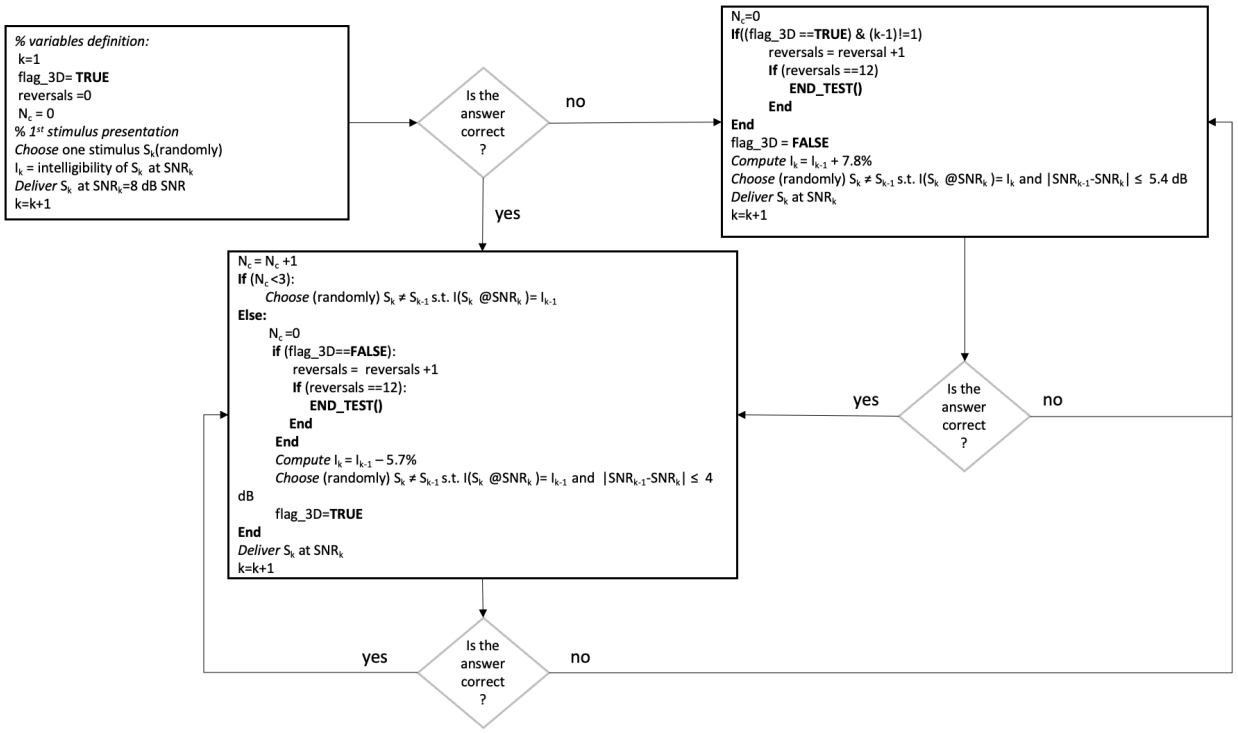

Figure 3.3: Work flow of the RSP procedure. N_c =*#correct* , S = *Stimulus* and I = *Intelligibility*.

The complete functioning of the RSP can be seen in Figure 3.3. Figure 3.4 shows the CSP after the correction level has been applied to psychometric functions.

Computational evaluation of RSP and CSP

In order to evaluate the performance of RSP compared to CSP, computational simulations of the test were conducted for both procedures, simulating fictitious responses based on the intelligibility percentage at each trial, taken as the probability of giving a correct response. These simulations were performed 500 times for both staircase procedures, and the mean, standard deviation, range of the SRT, total number of stimuli presented, and percentage of correct responses were all assessed. The distributions of the measured parameters were tested using the Shapiro-Wilk test. Figure 3.5 illustrates, in the top panel, an example of the simulated test for RSP in red and for CSP in blue, respectively, where the SRT obtained are very similar. In particular, the top panel shows all the trials of the two tests in relation to the relative SNR, while the bottom panel shows the same but relatively to intelligibility. It can be seen, in fact, how the steps are regular in the top panel for CSP as the difficulty variable is the SNR which varies by + or – 2 dB, while the regularity of the steps is visible for RSP in the bottom panel which in fact moves in steps of intelligibility. We see that in the two simulations, the trend is similar for the two variables SNR and intelligibility, and also the performance are similar in terms of the percentage of correct responses and SRT but the number of trials for RSP is clearly lower. The outcomes of computational simulations involving the new and the traditional methods, applied to VCV set of speech material, are presented in Table 3.1.

A significantly lower number of trials, indeed, was required for the RSP method (mean value 58) compared to the CSP method (mean value 95) (Wilcoxon rank-sum test: p « 0.001). The

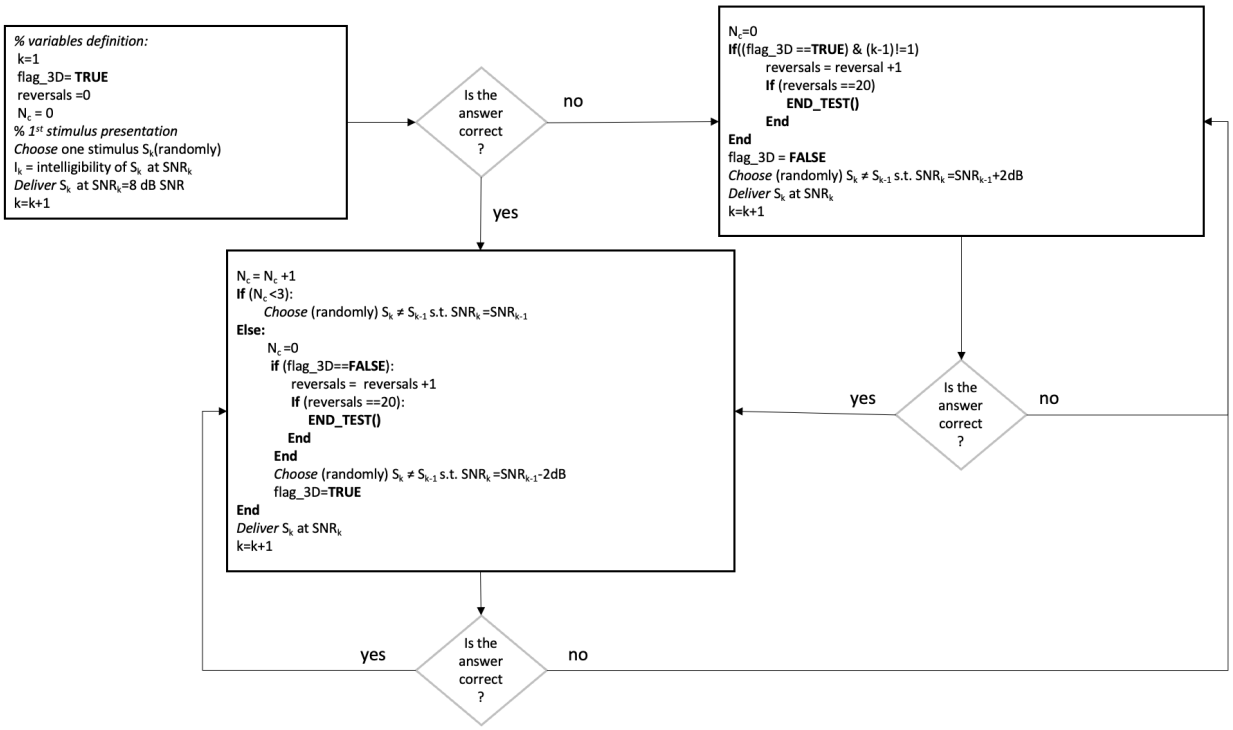

Figure 3.4: Work flow of the CSP procedure. N_c =*#correct* , S = *Stimulus* and I = *Intelligibility*.

difference in the number of trials between the two methods can be attributed to the difference in the termination rules, with the RSP method requiring 12 reversals and the CSP method requiring 20 reversals. The percentage of correct responses was similar in the two procedures. The RSP method generally yielded higher average values compared to the CSP method (t-test: p $\sim$ 0).

The average difference in the mean estimates of the SRT between the RSP and CSP methods was less than 3 dB SNR. On average, the RSP method resulted in higher (i.e., worse) values for the SRT compared to the CSP method (t-test: p « 0.001). This finding is consistent with the literature. According to a study by McShefferty et al. [40], indeed, the average just noticeable difference for changes in SNR and intelligibility, as assessed through individual judgments of speech clarity for sentences in noise, was 3 dB across different testing procedures (adaptive vs fixed-level) and hearing abilities (normal vs impaired).

Table 3.1: Mean (μ), standard deviation (s.d.) and range of total number of trials (#trials), percentage of correct responses (%corr), and SRTs obtained by using computational simulations of the CSP (top) and RSP (bottom) on the VCV set of speech materials (N = 500 simulations).

	#trials	%corr	SRT range (dB SNR)
		CSP	
μ (s.d.)	95 (10.9)	83.5 (1.24)	-4.5 (2.23)
range	64 - 135	79.8 - 86.9	-10.3 - 2.97
		RSP	
μ (s.d.)	58 (8.8)	85.8 (1.49)	-1.8 (3.44)
range	35 - 96	80 - 90.62	-11.2 - 8.8

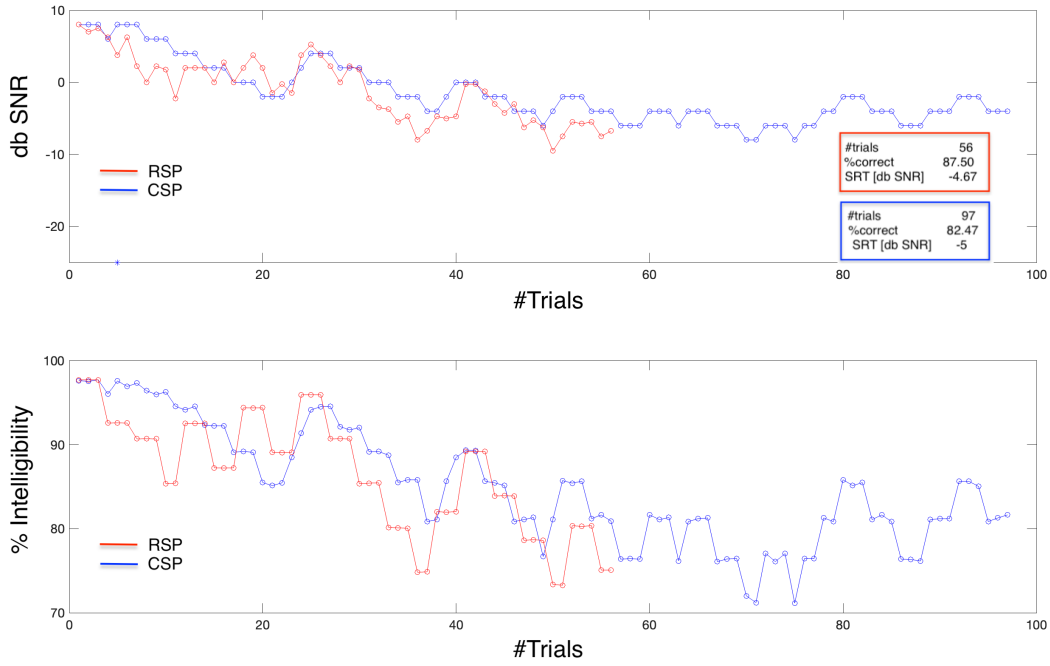

Figure 3.5: The top panel shows the SNR levels for two simulation runs of RSP (in red) and CSP (in blue) which obtained similar SRTs. The bottom panel shows the two simulation runs in relation to the intelligibility variable

Experimental evaluation of RSP and CSP

In order to reduce the fluctuations in signal-to-noise ratio during the RSP procedure, we employed a k-means clustering methodology [147] to partition the VCV stimuli into sub-groups, utilizing k values ranging from 2 to 12 and conducting 10000 iterations. The most appropriate clustering was determined by taking into account two criteria: (i) The proximity of the psychometric functions of the clusters at SNRs ranging from 1 to 4 dB, in order to guarantee that the SNR changes were minimal and perceptible as the staircases shifted between clusters. (ii) The proximity of the psychometric functions of the stimuli within each cluster at SNRs less than 1 dB, to minimize the dissimilarity among stimuli within a cluster [148]. Figure 3.6 shows the 4 clusters obtained resulting in: Cluster1 (asa), Cluster2 (afa, aga, aka, ata), Cluster3 (aba, ada, ala, ana, apa, ara), and Cluster4 (ama).

For this study, a sample of 26 young adults (11 males, 15 females) were selected, with an average age of 24.2 years (ranging from 23 to 26 years) and all were native speakers of Italian. The participants were confirmed to possess normal hearing based on the ISO 7029:2017 Standard. The examination was conducted using a Macbook Air in combination with a clinical audiometer Amplaid 177+, AmplifonTM). The TDH49 headphones were used and the output levels of the audiometer were calibrated as per the guidelines of ISO 8253-3:2012 standard. The participants underwent testing with both the CSP and RSP procedures during a single session in a randomized order, with breaks in between. To evaluate the variations among individuals in relation to the number of stimuli presented, the percentage of correct responses and SRTs, the appropriate statistical tests

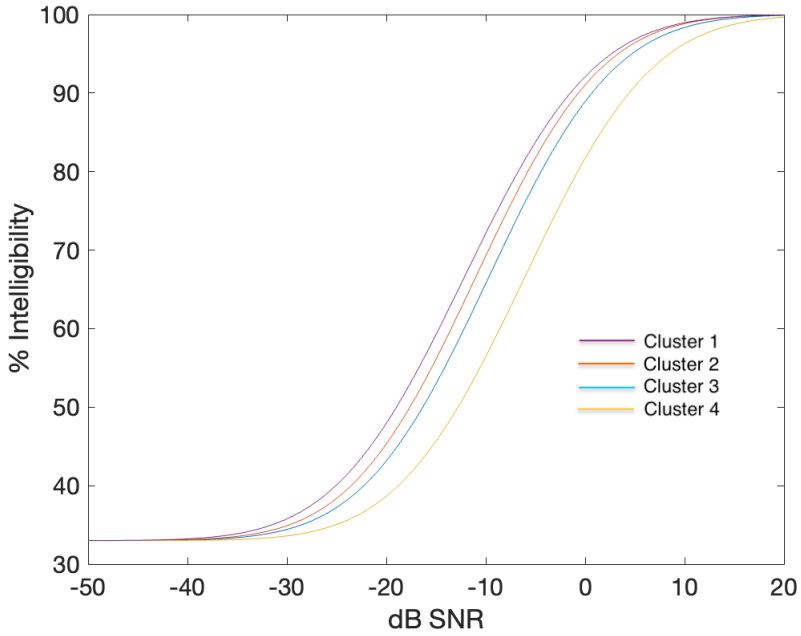

Figure 3.6: The 4 clusters obtained from the 12 VCV psychometric curves. Cluster1 (asa), Cluster2 (afa, aga, aka, ata), Cluster3 (aba, ada, ala, ana, apa, ara), and Cluster4 (ama)

were applied, paired samples t-tests or Wilcoxon signed-rank tests were used depending on the conformity of the data to normality.

Additionally, the efficiency of the test was assessed as the ratio of the ideal sweat factor "Kmin" to the actual sweat factor "K." The ideal sweat factor, "Kmin," was determined using the formula: Kmin = (target probability (p) * (1 - target probability (p)))/slope, where the target probability (p) in this study was set as 0.794 and the slope was computed from the average VCV recognition in normal hearing individuals, as 4.41%/dB, as in [149]. A truncated variation of the CSP methodology was also implemented, where the individual outcomes were evaluated after 12 reversals (CSP(trunc)) similar to the RSP. The normality of the measured variables was evaluated utilizing the Shapiro-Wilk test. To examine any potential inter-individual variations in SRT, number of trials, and percentage of correct responses, the paired samples t-test was utilized when data was found to be normally distributed as determined by the Shapiro-Wilk test. However, in instances where data was not normally distributed, the Wilcoxon rank sum test was employed. Furthermore, to adjust for multiple comparisons, the Bonferroni correction method was implemented. Figure 3.7 presents an illustration of the application of the RSP and CSP procedures on one of the participants. Both procedures initiate at a SNR of +8 dB, and the upward and downward steps employed are determined by the 1U3D algorithm. As a result, the estimated SRT obtained from both procedures is similar, estimated to be around -15.6 dB SNR. The CSP procedure utilizes a fixed step size of ±2 dB SNR and terminates after 20 reversals, resulting in a total of 130 trials and a percentage of accurate responses of 82.8%. Conversely, the RSP procedure employs a variable step size, with a ratio of Δdown/Δup less than 1 and concludes after 12 reversals, leading to a total of 80 trials and a percentage of accurate responses of 92.5%, which is higher than that obtained by the CSP

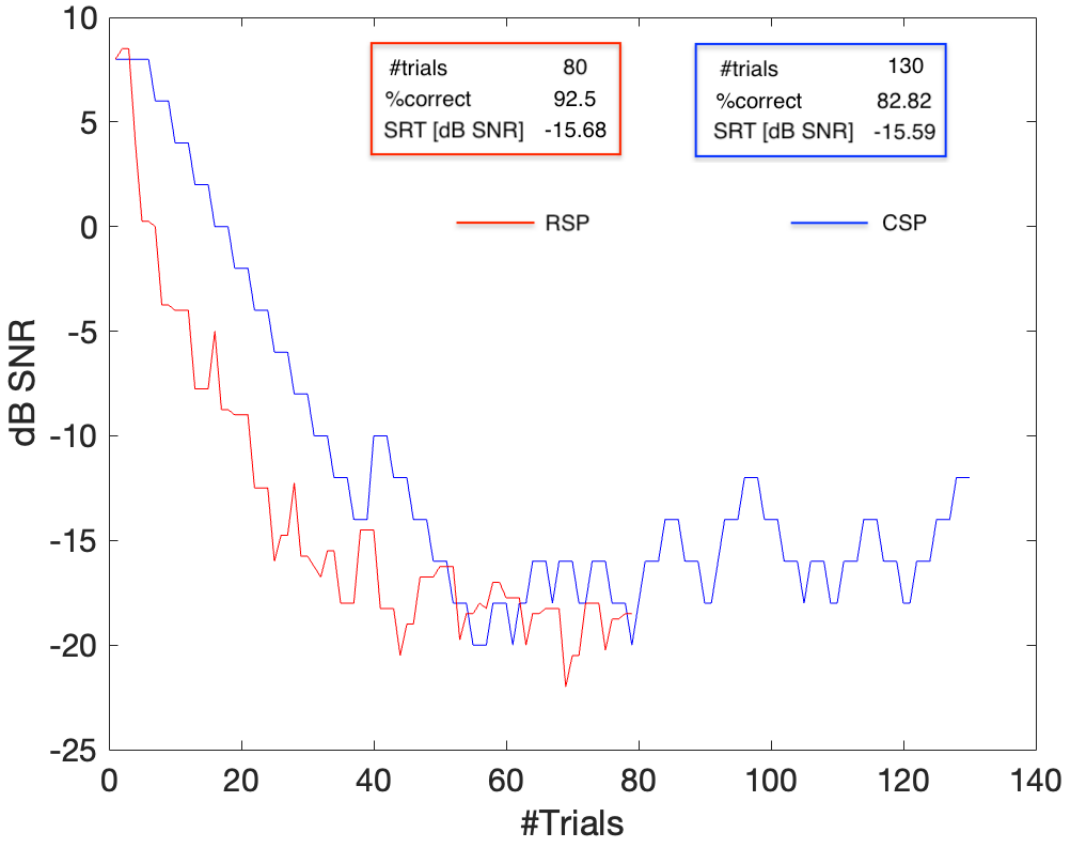

Figure 3.7: An example of the level of SNR as a function of trial number, observed using the CSP (blue) and RSP (red) procedures in a single participant, is presented. The text boxes accompanying the graph display the total number of trials (#trials), the percentage of correct responses (%correct), and the SRTs for each of the aforementioned procedures.

procedure.

Table 3.2 presents the results obtained from the group of 26 normal hearing young adults, who were subjected to the CSP, CSP(trunc), and RSP procedures with VCV stimuli. The results demonstrate that, on average, the number of trials in the RSP procedure was lower compared to the CSP and CSP(trunc) procedures. These differences were found to be statistically significant through the application of the paired samples t-test, with p-values less than 0.001 for the comparisons of CSP vs CSP(trunc) and CSP vs RSP. Additionally, the comparison between CSP(trunc) and RSP procedures also yielded a statistically significant result, with a p-value of 0.006. Furthermore, the percentage of correct responses obtained in the RSP procedure was found to be slightly higher than that of the CSP and CSP(trunc) procedures, and the observed differences were statistically significant as established through the paired samples t-test, with p-values less than 0.001 for the comparison of the three procedures. Additionally, the mean values of SRT estimated by the three procedures were comparable, with a mean difference of less than 0.7 dB.

The results of statistical analysis of the data showed that there was no statistically significant difference in SRT estimates among the three procedures, CSP, CSP(trunc) and RSP, as established by the paired samples t-test with p-values of 0.25, 0.24 and 0.12 respectively for the comparison of

Table 3.2: The mean (μ) and standard deviation (s.d.) of the total number of trials (#trials), percentage of correct responses (%corr), SRT, and efficiency are presented in the table, with measurements taken using the CSP (first column), the CSP(trunc) (second column), and the RSP (third column) procedures. These results were obtained from a sample of 26 participants.

	CSP	**CSP(trunc)**	**RSP**
#trials μ (s.d.)	124 (11.5)	87 (8.7)	82 (8.7)
%correct μ (s.d.)	86.2 (1.03)	89.0 (1.23)	91.0 (1.18)
SRT [dB SNR] μ (s.d.)	-15.4 (1.45)	-15.1 (2.27)	-15.8 (1.86)
Efficiency	1.42	0.83	1.31

the three procedures. Additionally, the observed variability in the SRT estimates was found to be higher with the CSP(trunc) and similar between RSP and CSP as determined by the F-test with the p-values of 0.03, 0.22 and 0.33 respectively for the comparison of the three procedures. Table 3.2 also shows that the efficiency of the RSP and CSP procedures were similar and higher than that of the CSP(trunc) procedure. The results indicate that the mean difference in SRT between the RSP method and the traditional CSP method was small, with a difference of less than 0.5 dB. Furthermore, the variability of SRTs observed in the study was consistent with the measurement errors reported in the literature for normal hearing individuals. For example, a measurement error between 0.7 and 1.0 dB has been reported for individuals with good SRT (i.e., SRT less than 4.1 dB SNR) in the digit triplet test [150]. Prior research also revealed similar test-retest differences for the RSP and CSP methods [148].

In a related study, RSP procedure had also a shorter duration of testing, with an average of 35% shorter than the CSP procedure [148]. The shortened duration of the test may present benefits in a clinical setting, as it decreases the chance of bias due to perceptual learning effects and reduces instances of fatigue during evaluation. This is a crucial aspect to accurately assess the listening effort.

Furthermore, it is noteworthy that the CSP (truncated) method utilizes a slightly higher number of trials compared to the RSP, despite utilizing the same quantity of reversals. Our findings, as presented in Table 3.2, indicate that the RSP method exhibits superior efficiency in comparison to the CSP (truncated) method, owing to a combination of lower variability in SRT estimates (1.86 versus 2.27) and a slightly lower number of trials (82 versus 87). This suggests that the RSP method is able to generate more accurate SRT estimates through more efficient trial placement when compared to the CSP (truncated) method.

Our research has yielded significant results regarding the effectiveness of the RSP method. In both simulated and experimental settings, the utilization of the RSP method resulted in a substantial increase in the proportion of correct responses when compared to the CSP method. Normal hearing individuals displayed an average correct response rate of 91% with the use of the RSP method and 86% with the CSP method, as evidenced by the data in Table 3.2. This increase in performance can be attributed to the fact that individuals underwent fewer incorrect trials and thus encountered fewer difficult stimuli when utilizing the RSP method as opposed to the CSP method. It should be noted that the percentage of stimuli presented below the individual SRT was equivalent for both

the CSP and RSP methods (33.4±9.49% vs 35.4±10.44%, respectively), as confirmed through a paired samples t-test with a p-value of 0.51. However, the SNRs utilized in the RSP method were closer to the SRT in comparison to those used in the CSP method, as exemplified in Figure 3.7, which rendered the stimuli less challenging for the participants and thus, led to an improvement in performance. The reduction of stimuli presented at very low SNR/low intelligibility levels may aid in decreasing the perceived difficulty of the procedure and the corresponding fatigue, particularly among older adults.

It appears that the RSP methodology has the potential to be a viable substitute for conventional staircase procedures. It may yield equivalent outcomes while being more expedient. However, further investigations are essential to validate these conclusions and to evaluate the full capabilities of the RSP technique.

The above-described revised staircase procedure is part of a larger ongoing project for the development of an online platform called WHISPER - Widespread Hearing Impairment Screening and Prevention of Risk which is a novel, artificial intelligence (AI)-powered, web-based system aimed at supporting widespread screening and prevention of hearing impairment. In addition to the speech-in-noise test, the platform includes a risk-factors questionnaire related to hearing loss and visual memory testing: digit and spatial span tests together with explainable AI models to predict potential hearing loss by comparing the result of the pure tone audiometry examination with the outcome from the speech-in-noise test [1] [2] [6] [5].

3.1.2 Listening effort experiment

In this subsection, we will be discussing the protocol that was implemented for the assessment of listening effort in individuals with normal hearing. The protocol involved the acquisition of physiological signals and the utilization of the speech-in-noise test described in the previous section. This combination of techniques allowed for a comprehensive evaluation of the listening effort experienced by the subjects, providing valuable insights into the cognitive and physiological processes involved in auditory perception. The protocol was carefully designed to minimize any potential sources of bias and to ensure the reliability of the results. The data collected through this protocol will be instrumental in the development of new strategies and technologies to enhance the listening experience for individuals with normal hearing and to assist those with hearing impairments.

Protocol

The assessment of listening effort by using speech-in-noise tests is a crucial aspect in the field of audiology and speech-language pathology. It is well established that speech recognition in the presence of background noise can be a challenging task for individuals with normal hearing, as well as for those with hearing impairments. The SRT is a widely used measure for evaluating the ability of an individual to recognize speech in noise, and it is defined as the minimum SNR at which an individual can accurately recognize a specified percentage of speech materials.

In this study, we employed our recently developed speech-in-noise test (See section 3.1.1). In order to evaluate different levels of difficulty, we utilized the individual SRT as a benchmark to differentiate between low and high difficulty trials. As research has shown that the just noticeable

difference in speech for changes in SNR and intelligibility is around +3 dB SNR [151], we defined low difficulty trials as those with a SNR higher than the individual SRT+2 dB SNR, and high difficulty trials as those with an SNR lower or equal to the individual SRT+2 dB SNR. We added +2 dB to the SRT to divide the trials into easy and difficult categories, particularly because finding many consecutive trials at high difficulty level (i.e., SNRs<SRT) was difficult and thus the 2 dB increase served both to have a greater number of trials on which to evaluate and compare physiological signals, and to have a balance between trials that are challenging but still understandable and trials with too low SNR where the subject made errors due to lack of comprehension. This approach enabled us to obtain a more comprehensive understanding of the listening effort experienced by the subjects, and provided valuable insights into the cognitive and physiological processes involved in auditory perception. By utilizing the individual SRT as a benchmark, we were able to ensure that the level of difficulty of the trials was tailored to the specific abilities of each subject, resulting in a more accurate assessment of their listening effort. Typically, in studies that evaluate physiological measures for listening effort, a speech-in-noise test is first conducted to obtain the individual SRT and then a separate test is conducted to evaluate physiological measures, introducing likely fatigue due to the subject completing two tests in succession. In contrast, in this study, the design of the speech-in-noise test was optimized to be short and easy, in order to reduce the occurrence of fatigue and provide a more objective assessment of the physiological measures. This was done by carefully selecting the speech materials and the level of difficulty of the trials, as well as by controlling the duration of the test. By minimizing the occurrence of fatigue, we were able to obtain a more accurate representation of the listening effort experienced by the subjects and to identify the physiological signals that were most closely associated with listening effort.

During the study, we specifically focused on analyzing equal duration segments of consecutive trials at the two fixed levels of difficulty, namely low (L) and high (H) difficulty, for each subject. Our experimental protocol involved the participation of 13 female and 8 male subjects with an average age of 26.18 $\pm$ 1.47 years. Prior to conducting the test, all participants underwent pure-tone audiometry on both ears to ensure that their hearing thresholds were within the normal range (pure-tone average thresholds among 500, 1000, 2000 and 4000 Hz $<$ 20 dB HL). The pure-tone audiometry was conducted to exclude participants with hearing impairments and ensure that the results obtained were representative of individuals with normal hearing.

To minimize behavioral heterogeneity among the participants, they were instructed to avoid coffee and smoking for two hours prior to the experiment. This was done to control for the effects of caffeine and nicotine on physiological signals and to ensure that the results obtained were not influenced by these factors. Additionally, in order to obtain a baseline measurement, 2 minutes of data was recorded while the subjects were looking at a grey screen just before the start of the speech-in-noise test. This baseline measurement served as a reference point for the analysis of the physiological signals recorded during the speech-in-noise test and helped to identify any changes in the physiological signals that were related to the listening effort. Moreover, subjects who wore headphones (UXD CT887) were able to adjust the volume of stimuli through a slider during the test through an initial training phase before the start of the test in order to have a volume that was comfortable for them.

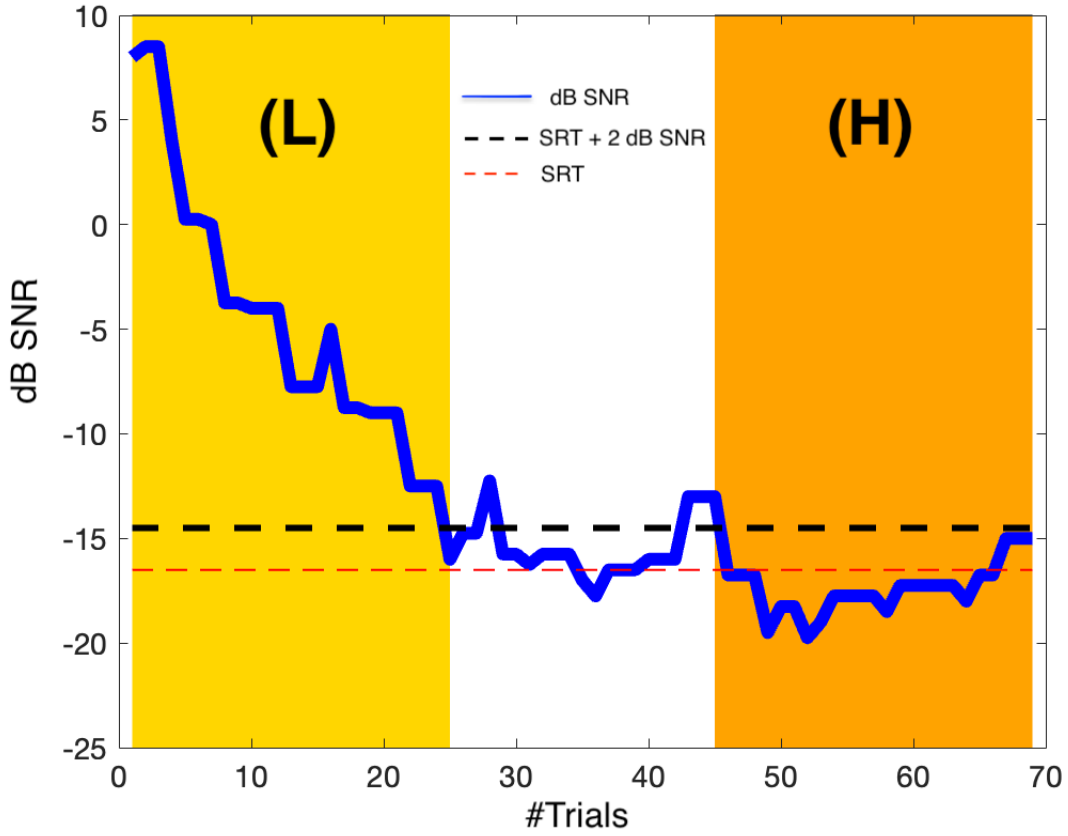

Figure 3.8: Test trials for one of the subjects are shown. In particular, the two time windows at the two levels of effort are highlighted by the colored boxes L = low effort and H = high effort. The solid red line shows the subject's individual SRT while the dashed black line represents the SRT increased by 2 dB SNR.

Furthermore, by collecting physiological signals during the speech-in-noise test, such as ECG, BVP, GSR, PUPIL, RESP and EEG, we were able to gain additional information on the neural and autonomic responses associated with listening effort. This allowed us to investigate the relationship between physiological signals and listening effort both at the central and peripheral level, and to identify potential biomarkers that can be used to predict and monitor the listening effort of individuals with normal hearing and hearing impairments.

The Figure 3.8 shows the test performance for a subject where the y-axis represents the SNR level for each stimulus and the x-axis represents the number of stimuli. The two colored areas identify low and high effort trials, while the dotted line indicates the subject's individual SRT.

Statistical analysis

A thorough statistical analysis was conducted to evaluate and compare the various features extracted from the physiological signals among the two effort levels (L and H) and the baseline (B) within the same subjects. To accomplish this, a variety of statistical tests were employed, specifically, the Friedman's test was utilized if at least one of the variables was found to be non-normal and the one-way ANOVA test with repeated measures was applied if all variables were determined to be

normal. Additionally, to ensure the accuracy and integrity of the results, a Bonferroni correction was applied to all comparisons.

3.1.3 Emotions and music (AuBT protocol)

Despite the increasing interest in utilizing music for clinically-oriented purposes such as monitoring depression, anxiety, stress, chronic anger, and mood disorders [152] [153], the emotional aspect of sound has received comparatively less attention compared to the extensive research on visual stimuli in the field of emotion recognition from physiological signals [101] [154]. In light of this, we have conducted an analysis on an online dataset of emotion recognition through music, which has received relatively little attention in literature, particularly with regards to the physiological changes induced by emotional music. In the field of emotion recognition, most studies tend to overlook the interpretation of features and instead focus on achieving high accuracy rates through dimensionality reduction methods and calculating a high number of features without specifying which signals and features may be most relevant for creating a framework that can be applied in similar studies. Our objective is to create a simple framework for emotion recognition by investigating the meaning of features in relation to specific emotions and music.

Protocol

The data employed in our study was procured from the University of Augusta, Germany [90]. Our research entailed a systematic assessment of a single individual over a period of 25 consecutive days. The participant was exposed to a diverse array of musical selections, chosen on a daily basis, in order to evaluate various emotional states: joy, anger, sadness, and pleasure. The purpose of this musical exposure was to guide the participant through a progression of emotional states, beginning with joy, then progressing to anger, sadness, and ultimately ending with pleasure. To elicit the intended emotional state, the participant was prompted to select songs that evoked personal memories. The effectiveness of this method is founded on the principle that most individuals have a tendency to associate certain musical melodies with specific emotional states as a result of their frequent exposure to music throughout their daily lives. Furthermore, the study aimed to evaluate the effectiveness of this approach to elicit specific emotional states in an individual over an extended period of time, by comparing the physiological and behavioral responses before and after exposure to music.

During the course of the study, the participant was closely monitored utilizing various biosensors to acquire four key physiological signals: GSR, ECG, EMG, and RESP. The purpose of this monitoring was to provide a comprehensive understanding of the physiological changes that occur within the participant as a result of exposure to different musical stimuli and emotional states.

After each round of signal collection, the participant was given a break of varying duration to allow the physiological parameters to return to their baseline, prior to proceeding with the next recording. This was done to ensure that any variations in the physiological signals were a direct result of the musical stimulus and not residual effects from previous recordings. In total, 25 recordings were obtained for each of the four emotions and for each of the physiological signals, providing a robust dataset for analysis.

The recordings were acquired throughout the duration of the musical pieces, but were subsequently standardized to 120 seconds from the initiation of each piece. This was done to ensure that the data collected was consistent and comparable across all musical pieces and emotions. By standardizing the recordings in this way, it was possible to accurately compare the physiological responses to different musical stimuli and emotions.

The ECG was recorded at a sampling rate of 256 Hz, while GSR, EMG, and RESP were recorded at a sampling rate of 32 Hz.

Furthermore, in light of the advancements in wearable sensor technology currently available, it was decided to exclude the facial EMG sensor from the study. This decision was made due to the increased invasiveness of the sensor and its lower level of integration with other available sensors. Instead, the focus of the study was on analyzing only three signals for which integrated sensors are already available on the market, which are ECG, GSR, and RESP. This approach allowed us to acquire a high-quality data set that was non-invasive and easily comparable to other similar studies.

Statistical analysis

A comprehensive statistical analysis was carried out to determine if there were any differences in physiological responses to varying emotions. Specifically, in order to evaluate the responses of the same subject within different days, the non-parametric Friedman's test was employed with Bonferroni correction applied to adjust for the number of comparisons conducted, taking into account the same feature.

Classification of emotions

After the signal processing of all the signals involved in the study and feature extraction step which are going to deeply described in the next section (3.4), simple emotion recognition models were developed using the k-nearest-neighborn (KNN), support vector machine (SVM), decision trees (DT) and linear discriminant analysis (LDA) classifiers described in details in section 3.2.2. These models have been widely used in various fields, including emotion recognition, making them suitable choices for this study. In order to provide a simple and interpretable model, a feature selection method based on a graphical visualization was applied [9]. This method, named Square Method (SM), allows to identify the most important features of the data that contribute to the classification of the emotions. The graphical visualization method chosen for this study, is a 2D boxplot, where for each pair of features, the center of the boxes represents the average of the two features and the sides of the boxes represents the 95% confidence intervals for the average estimation. By analyzing these boxplots, it was possible to identify the features that appeared most frequently in the pairs where no or little intersection between the boxes was detected. These features were considered the most important for the classification of emotions and were selected for the creation of the model. Figure 3.9 illustrates the procedure for constructing 2D boxplots as a graphical method for feature selection. This approach allowed to achieve a simple and interpretable model which can be easily understood and replicated in future studies. It also allowed to identify the most relevant features for the classification of emotions, which can be further studied and used in other research. Below is the algorithm shown in more detail:

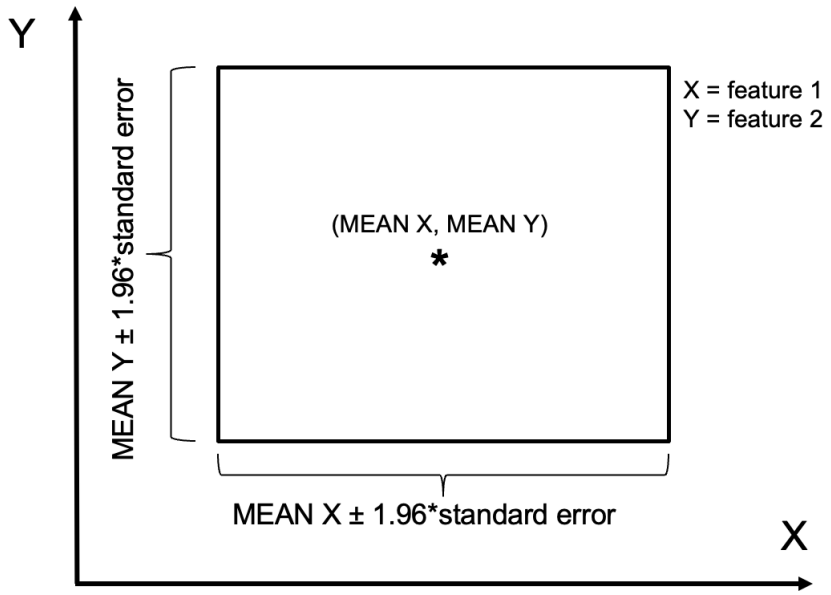

Figure 3.9: The method for constructing 2D boxplots to be used for the selection of the most relevant features for discriminating the four emotions (i.e., joy, anger, sadness, pleasure) is shown. Each 2D box is centered at the mean of two features for the specific emotion and the sides of the box are composed of the 95% confidence intervals for the estimate of the mean.

- For each emotion considered, 2D boxplots are created for all possible pairs of features.
- The area of intersection between each pair of rectangles is computed, and the ratio between the area of intersection and the area of each rectangle is calculated and noted as R.
- For each pair the sum of R is obtained, and if the sum exceeds a predefined threshold, the corresponding feature pairs are discarded.
- The correlation between not discarded features is calculated, and among the correlated features, only that which appear in more pairs are retained.
- The feature importance is then determined by ranking the selected features based on the number of pairs in which they appear.

Given that the dataset comprises repeated measurements of the same individuals on distinct days, and to avoid data contamination from the possibility of utilizing measures from the same day for both training and testing the model, as well as the limited size of the dataset, a modified leave-one-out cross-validation method (LOOCV) was applied to the entire dataset. LOOCV is a resampling method used for model evaluation and selection. In this method, the dataset is divided into subsets, with one sample being left out as the validation set, while the remaining samples are used as the training set. This process is repeated for each sample in the dataset, so that each sample acts as the validation set once. Figure 3.10 shows an example of LOOCV with five iterations.

Once the three best features were chosen from the SM method, the following analysis was carried out using these three features. In order to thoroughly evaluate the performance of the model on

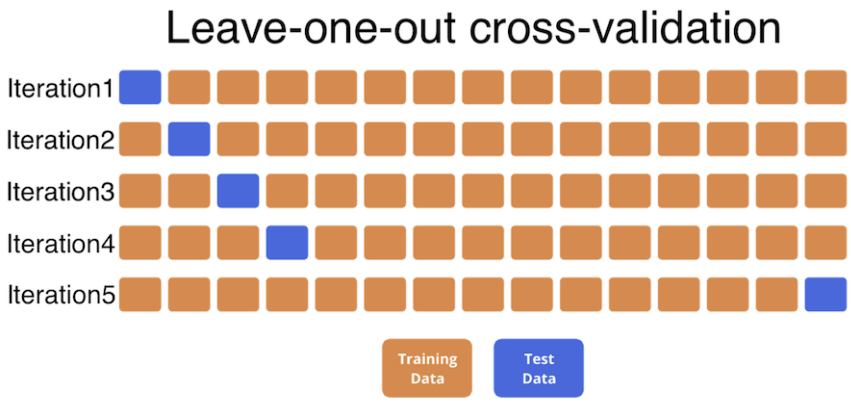

Figure 3.10: The functioning of the leave-one-out cross-validation method is shown. In this particular case, it is a 5-fold leave-one-out cross-validation. The data is randomly divided into five subsets of equal size. In each iteration of five, each subset of data serves as the test set and the other four subsets serve as the training set.

the dataset, which consists of repeated measurements of the same individuals on different days, it was necessary to take measures to prevent data leakage and contamination. Given the limited size of the dataset, a modified leave-one-out cross-validation method was applied to the entire dataset. In particular, to evaluate the influence of different days on the classification, a leave-5days-out cross validation was applied. In each iteration, five entire days of measurements were left as the validation set. This approach allowed us to see if the performance of the model persisted even when a larger number of observations were left out as the validation set. Subsequently, in order to see if the performance persisted even with fewer observations, a leave-1day-out approach was applied, where in each iteration, only one day of measurements were left as the validation set. This approach enabled us to evaluate the robustness of the model with a smaller validation set. Finally, a leave-one-out approach was implemented, leaving only one observation in the validation set. This approach enabled us to evaluate the performance of the model with the smallest possible validation set.

The same analysis was repeated with the same three features for binary classification of arousal and valence. By conducting these evaluations, we were able to determine the robustness and reliability of the models in different scenarios and also the efficiency of the music as a potential stimulus to evoke emotions.

The performance of each model was evaluated using the mean validation accuracy. Accuracy was used as the performance metric as the dataset is balanced both in terms of the four emotions and in terms of valence and arousal.

3.2 Emotional protocol with visual, auditory, and combined stimuli

In this section, we will present another experiment that was conducted to compare the effects of different types of stimulation on physiological and central responses. The experiment specifically focused on comparing audio, visual, and audio-visual stimulation. The objective of this experiment

was to investigate how the body responds physiologically and centrally to these different types of stimulation and to evaluate the changes in physiological activation in the same subject when the type of stimulus is altered.

The study design involved recruiting a sample of participants and exposing them to different types of stimulation. Physiological signals such as ECG, BVP, GSR, RESP, PUPIL were recorded during the stimulation, and central measures such as brain activity were also recorded using EEG. Participants were also asked to self-report their emotional state.

3.2.1 Design of the Protocol

In the present study, we recruited a total of 21 participants, comprising of 13 females and 8 males, with an average age of 26.18 $\pm$ 1.47 years, to take part in our experimental protocol. These experiments were carried out at the SpinLab of Politecnico di Milano, and all participants provided written informed consent, which was obtained in accordance with the guidelines set forth by the Politecnico di Milano Research Ethical Committee. These procedures were implemented to ensure that the rights and welfare of the participants were protected throughout the study.

During the course of the experiment, various physiological signals were obtained from the participants, ECG, BVP, GSR, PUPIL and EEG.

Specifically, the ECG, BVP, RESP and GSR were measured using the Procomp Infiniti device, which had a sampling frequency of 256 Hz for GSR and RESP and 2048 Hz for ECG and BVP. The PUPIL signal was measured using the Tobii Pro X2 Compact eye-tracker, which had a sampling frequency of 60 Hz. The EEG signal was obtained using a DSI 24 headset, which was equipped with 19 dry electrodes placed at specific locations on the scalp according to the international 10-20 system. The headset had a sampling rate of 300 Hz and used a 16-bit A/D converter for accurate data acquisition.

These measurements were conducted in order to obtain a comprehensive understanding of the physiological responses of the participants to the experimental conditions. The use of multiple physiological signals also allowed for the examination of various aspects of the participants' physiological responses.

During the course of the experiment, the participants were seated in a comfortable chair and were closely monitored by an experimenter to ensure that there were no pronounced movements or sensor detachment. To minimize behavioral variability, the participants were instructed to avoid consuming coffee and smoking for two hours prior to the experiment. Additionally, all subjects underwent a pure tone audiometry examination using the Amplaid 177+ clinical audiometer with TDH49 headphones to confirm that their hearing thresholds were within the normal range (pure tone average thresholds at 500, 1000, 2000, and 4000 Hz $<$ 20 dB HL).

This approach was adopted to control for any potential sources of variation and to ensure that the results obtained were a reflection of the experimental manipulation rather than other extraneous factors. By conducting the pure tone audiometry examination and controlling for the use of coffee and smoking prior to the experiment, we aimed to ensure that any differences observed in the physiological and central responses were a result of the experimental manipulation rather than other factors such as hearing loss or caffeine consumption.

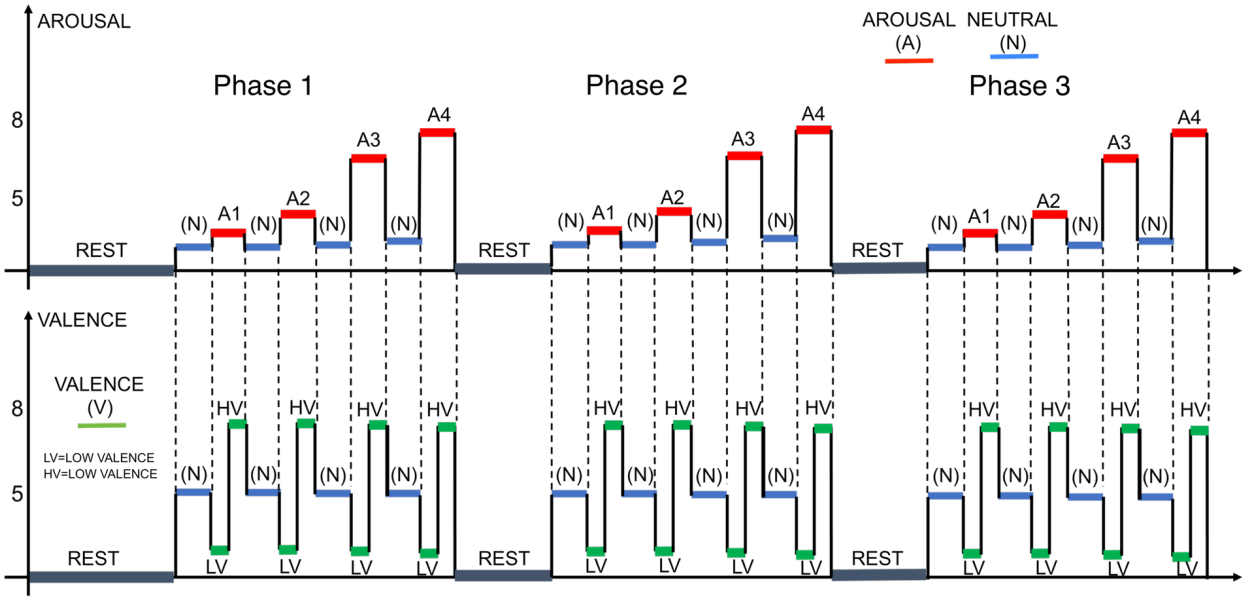

Figure 3.11: The three randomized phases of the protocol (i.e., IAPS-only, IADS-only, IAPS+IADS) are shown. Each phase is characterized by four increasing arousal stages highlighted in red, which consist of the first half with low valence stimuli, and the second half with high valence stimuli highlighted in green, interspersed with neutral stages highlighted in blue.

The acoustic stimulation was delivered through UXD CT887 headphones, and participants who had vision issues were allowed to wear glasses during the experiment. The protocol consisted of three randomized phases, with 2-minute breaks in between: IAPS-only, IADS-only, and IAPS + IADS. Specifically, in a given phase, only images are presented to the subject, in another phase the subject will only listen to sounds, and in a final phase, the stimuli are provided by a combination of images and sounds that have a semantic match with each other.

This protocol is a modified version of an earlier protocol that employed only IAPS images as stimuli [155]. The updated protocol is composed of four distinct phases that include a preliminary resting period of five minutes during which participants are shown a gray screen. Following this initial stage, four sessions of increasing arousal levels are alternated with 90-second neutral sessions. Each session of heightened arousal includes six visual and/or auditory stimuli, each lasting 15 seconds, that are chosen to have a low valence in the initial portion and a high valence in the latter half. The entire experiment takes approximately 47 minutes to complete. Figure 3.11 shows all the phases of the protocol. Figure 3.12 shows the set up used for the experiment.

The primary objective of this design is to systematically elevate the level of arousal in a gradual manner while maintaining a neutral emotional valence throughout each session. To achieve this, the design employs a series of neutral sessions that present low arousal stimuli with a neutral emotional valence. These neutral sessions serve as a reference point for rest between two emotionally charged sessions. Under the IAPS+IADS condition, visual stimuli in the form of pictures are presented in conjunction with auditory stimuli that have a semantic match. For example, a visual stimulus of a growling dog would be presented along with an auditory stimulus of a barking dog. Additionally, in the IAPS+IADS condition, pictures and sounds from the neutral sessions of the IAPS-only and IADS-only conditions are presented in combination, in order to create a more comprehensive

Figure 3.12: The set up used for the experiment. On the left hand of the subject is possible to see the BVP and GSR sensors of the Procomp device. On the box in front of the subject is possible to see the Tobii Pro X2 Compact eye-tracker and on the head of the subjects the DSI 24 headset is shown.

emotional experience. Both arousal and valence levels are set according to IAPS and IADS scores as shown in Table 3.3. To compare IAPS and IADS labeling with subjective ratings, participants were asked to perform a self-assessment of the stimuli seen and heard during the experiment immediately after the protocol was completed. Each stimulus was quickly reproduced, and subjects were asked to select arousal and valence levels using self-assessment manikins based on their own emotional reactions. The subjective ratings did not deviate significantly from the valence and arousal values shown in Table 3.3, indicating that the chosen stimuli were effective in eliciting both negative and positive valence and increasing arousal at each phase of the protocol, ensuring that the protocol is well-designed for measuring emotional responses.

3.2.2 Statistical analysis and Classification

A comprehensive examination was conducted to evaluate the physiological response of the body to various stimulation modalities, such as the visual presentation of images, the auditory presentation of sounds, and the combination of both. The study focused on two emotional dimensions, namely valence and arousal, and sought to determine how these dimensions affected the body's reaction. The examination was conducted across three distinct phases, and comparisons were made between each phase in relation to the two emotional dimensions.

To analyze the arousal dimension, the features of each arousal session in one phase were compared

Table 3.3: Valence and Arousal medians and ranges for all sessions (Neutral, Arousal1, Arousal2, Arousal3 and Arousal4) of each phase (IAPS-only, IADS-only and IAPS+IADS) are shown. Matched IAPS and matched IADS refer to the stimuli used in IAPS+IADS.

Session	Valence Rating	Arousal Rating
IAPS only		
Neutral	5.02(4.39-5.55)	3.36(3.07-3.84)
Arousal1	4.96(3.92-7.24)	4.68(4.42-4.85)
Arousal2	4.81(2.52-7.62)	5.38(5.08-5.48)
Arousal3	5.10(2.14-7.2)	6.55(6.09-6.80)
Arousal4	4.75(1.45-7.57)	7.12(6.90-7.35)
IADS only		
Neutral	5.47(4.34-5.99)	3.74(2.88-4.15)
Arousal1	5.16(3.54-7.12)	4.77(4.47-4.94)
Arousal2	4.93(2.46-7.78)	5.40(5.05-5.87)
Arousal3	4.88(2.44-7.38)	6.35(6-6.84)
Arousal4	4.86(1.68-7.67)	7.14(7.03-7.88)
matched IAPS		
Arousal1	5.67(3.65-7.13)	4.46(4.13-4.75)
Arousal2	4.44(2.49-6.83)	5.43(5.18-5.53)
Arousal3	4.76(2.16-6.83)	6.21(6.06-6.79)
Arousal4	4.19(1.48-7.61)	7.18(7.13-7.31)
matched IADS		
Arousal1	5.86(4.52-7.05)	4.62(4.38-4.87)
Arousal2	4.82(3.02-6.11)	5.50(5.34-5.74)
Arousal3	4.54(2.06-6.77)	6.42(6.07-6.64)
Arousal4	4.54(1.99-6.94)	7.35(7.28-8.16)

to the corresponding session in the other two phases. Similarly, for the valence dimension, the features computed during low and high valence of the same arousal session were compared across the three phases. The normality of the data was determined using the Shapiro-Wilk test for each feature value in low and high valence, as well as for each entire arousal session. As at least one variable was found to be non-normally distributed, a non-parametric and pair-wise Friedman's test was applied. If a comparison was found to be significant, a multcompare comparison was then conducted to determine which phases exhibited significant differences. In all cases, the Tukey correction method was applied.

Although the primary objective of this study was to compare the efficacy of different types of stimulation, such as visual, auditory, and combined audio-visual, a classification task was also conducted. Specifically, the aim of this classification was not to develop high-performance machine learning models, but rather to assess the generalization capabilities of machine learning models across different types of stimulation and within each type of stimulation. For this reason, simple machine learning models are used together with the simple and interpretable feature selection method SM described in section 3.1.3. The ultimate goal was to evaluate how well these models were able to differentiate between emotions based on the type of stimulation used, and to identify

which type of stimulation resulted in a clearer separation of emotions from physiological signals, and which was less effective, leading to more ambiguous physiological experiences.

Firstly, for each type of stimulation, the two middle phases of increasing arousal (A2 and A3) were excluded, thus considering only the extremes of Russell's circumplex model. Specifically, phases A1 and A4 were selected as they represent states of very low and very high arousal respectively, with the aim of facilitating the classification task. Phases A2 and A3 are, in fact, situated between low-to-neutral and neutral-to-high arousal respectively. Figure 3.13 depicts the four different phases of arousal within Russell's circumplex model, with each phase being characterized by stimuli of low valence in the first half and high valence in the second half.

The dataset used in this study consisted of 264 observations, with 12 for each subject (representing each of the 4 quadrants for each of the 3 phases of stimulation). To avoid altering the extracted signal features, a dimension reduction method was not applied for classification. Instead, the dataset was split into a training set (70%) and a test set (30%), with 7 subjects randomly excluded from the latter. The customized SM was applied on the training data to select the most important and least correlated features for separating the 4 emotions. Then, a stratified 5-fold cross validation was conducted on the training set using different machine learning models, with the best performing model selected and applied to the test set. As the dataset was balanced, the average accuracy of validation and test accuracy were computed as performance metrics.

Moreover, to determine the most effective type of stimulation for classification, the feature selection method was applied to three different datasets: images only, sounds only, and combined stimuli, each with 88 observations. Due to the limited number of observations, only a stratified 5-fold cross validation was performed, followed by the calculation of the average accuracy of train and validation.

As different models have different characteristics and can be more or less effective depending on the problem, several models were used, including KNN, LDA, logistic regression (LR), SVM, random forest (RF) and Adaboost (ADB). Below are brief descriptions of the algorithms associated with the different models.

- KNN is an algorithm that identifies the K nearest training samples in the feature space to a given input and predicts the class or value of the input based on the labels of its K nearest neighbors.

- LDA is a simple classification algorithm that looks for a way to separate data into different groups based on certain features. It does this by finding a straight line that best separates the data into these groups. Once this line is found, it can be used to classify new data points based on where they fall in relation to the line.

- LR for multiclass is a classification algorithm that extends binary logistic regression to handle more than two classes. It works by fitting separate logistic regression models for each class and then choosing the class with the highest predicted probability.

- SVM is a classification algorithm that seeks to find the optimal hyperplane that best separates different classes in a high-dimensional feature space. It works by transforming the data into

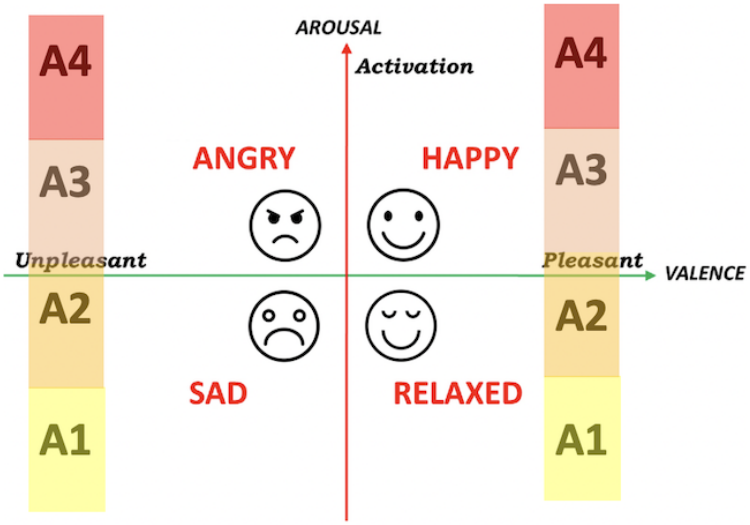

Figure 3.13: The circumplex model of affect. The arousal ranges related to the four arousal sessions (A1, A2, A3 and A4) are shown in the figure.

a higher-dimensional space where a clear linear boundary can be drawn between the classes. The hyperplane is chosen to maximize the margin between the classes, making the classifier more robust to noise and outliers.

- RF is a classification algorithm that constructs multiple decision trees and aggregates their predictions to improve accuracy and prevent overfitting. It randomly selects a subset of features for each tree and bootstrap samples of the training data to increase diversity among the trees.
- ADB is an algorithm used for classification and regression tasks. It works by iteratively training weak models on weighted training data and then adjusting the weights to emphasize misclassified samples in subsequent iterations. The final model is a weighted sum of the weak models.

3.3 Emotions and virtual reality

The design and implementation of a suitable protocol for evoking a range of emotions through fully-immersive VR was the central focus of this section. In particular, the protocol aimed to elicit feelings of sadness, relaxation, happiness, and fear by utilizing an escalating arousal strategy that alternates between periods of neutral stimuli and periods of emotionally-charged stimuli.

It provides a systematic and comprehensive approach for evoking a range of emotions through fully-immersive VR, which has the potential to be applied in a wide range of applications, such as therapy, education, and entertainment.

3.3.1 Design and Implementation of the Protocol

To inform the development of this protocol, a series of studies were conducted using VR as a means of eliciting emotional responses. These studies were carefully conducted and evaluated in order to gain a better understanding of how different types of VR experiences can influence emotional responses. The findings from these studies were then used as a blueprint for the design of the protocol.

One specific example of the protocol that was developed in this project can be found in [156], in which the authors conducted an investigation into the psycho-physiological patterns evoked during the free exploration of both an actual art museum and a virtualized version of the same museum through a 3D immersive virtual environment. This study was designed to explore how the experience of visiting an art museum can be replicated in VR and how this replication can influence emotional responses.

In the initial phase of this project, a comprehensive elicitation experience was developed and validated through the utilization of a variety of sources including prior knowledge obtained from video-game design, literature, datasets such as IAPS, and multimedia platforms such as Netflix. Additionally, feedback was solicited from a first group of subjects to further refine and improve the elicitation experience.

In the second phase of this project, an emphasis was placed on both the acquisition of physiological and circumstantial data through the use of two forms and a ProComp device. This was achieved by carefully collecting data on a variety of physiological signals such as ECG, BVP, GSR and RESP, as well as circumstantial data such as self-reported emotions and subjective ratings of the elicitation experience. This data collection process was crucial in providing a comprehensive understanding of the emotional responses elicited by the experience developed in the initial phase. Unfortunately, it was not possible to acquire the EEG signal in this phase as the experiment involved subject's head movement within the immersive scenes, which introduced movement artifacts. Additionally, the headset used consisted of headbands that were placed on the head, which did not allow for proper adhesion of the electromyographic headset to the scalp. Efforts were made to standardize the experimental station in order to reduce the variability in responses among subjects. The methodology for achieving this standardization is detailed in Figure 3.14. To ensure the hygiene of the study, the hands of each subject were disinfected, the skin conductance electrodes were thoroughly cleaned with alcohol, and the virtual reality headset was sterilized and cleaned prior to each use.

To further refine the emotional settings to some of the subjects were asked to complete a pre-protocol survey, which aimed to assess the coherence of the emotional settings. Additionally, a post-protocol survey was administered to other subjects in order to estimate the degree of adherence of their subjective experiences to the expected ones. The results of these surveys were used to make any necessary adjustments to the protocol and ensure that it effectively evoked the intended emotions. Two survey methods were used to gather data on the emotional responses elicited by the protocol. The first survey method was the SAM which required participants to select a value on a scale of 1-5 for both arousal and valence for each emotional room and the PAM which required participants to choose an emotion from a list of 9 possible options. Additionally, some personal information was also collected. In the pre-protocol survey, participants were asked to rate the images and sounds

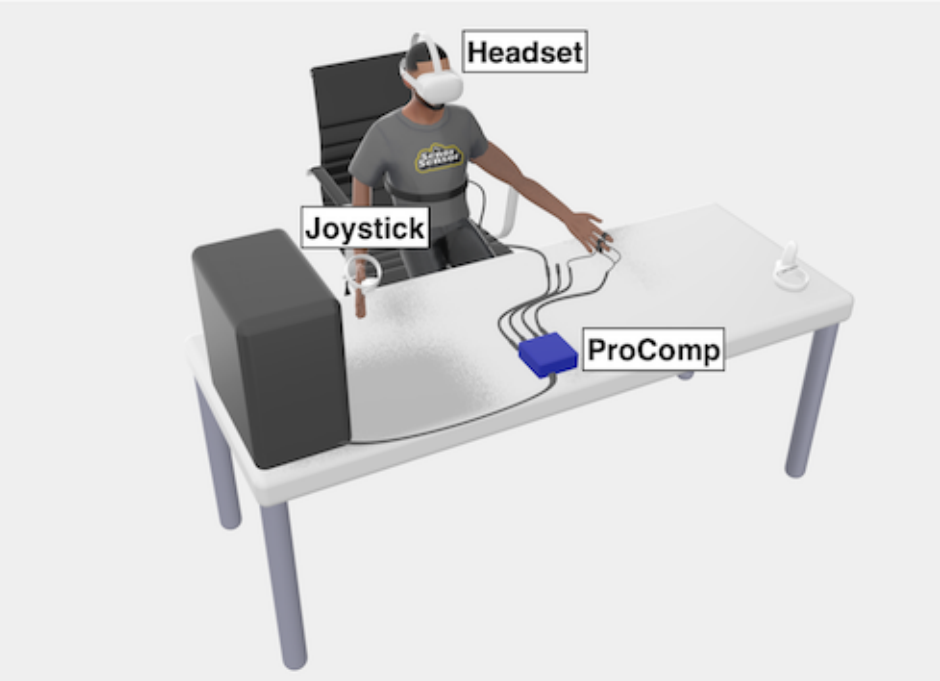

Figure 3.14: Virtual reproduction of the experiment setup.

presented in each emotional room. In the post-protocol survey, participants were invited to rate the same images and sounds based on the sensations they felt during the specific experiences. This allowed for the comparison of the subjective experience of the images and sounds with the intended emotional settings. The results of these surveys were used to evaluate the effectiveness of the protocol in evoking the intended emotions and to make any necessary adjustments to the protocol.

The characterization of the scenes in the study is determined by the quantity, intensity, and type of stimuli present, including visual, acoustic, and tactile elements. Throughout the course of the protocol, participants are able to interact with their virtual surroundings to a certain extent, utilizing virtual hands as a means of interaction. The experiences were created using Unity, providing comprehensive control over the visual and auditory components of the scenes. The methodology for selecting audio-visual cues, as outlined in [157] which is an extensive study of the cues associated with affective states, was utilized as the foundation to select the type of stimuli for the protocol. A detailed description of each scene can be found in the subsequent section.

The protocol was constructed in a manner similar to that of the preceding section, with a gradual increase in arousal, so as to avoid scenes of excessively high arousal impacting the physiological responses to subsequent scenes. The goal was to represent four fundamental emotions within the VR scenes, namely sadness, pleasure, happiness, and fear, interspersed with neutral scenes. The total duration of the protocol is 42 minutes, comprising a total of 9 distinct experiences (i.e., initial scene, four emotion al scenes and 4 neutral scenes). The initial 2 minutes scene was used as adaptation period, while the remaining scenes each last 5 minutes.

3.3.2 Emotionally-Inducing Virtual Reality stimuli

Initial scene of adaptation

The introductory experience serves a two-fold purpose. Firstly, it aims to mitigate the potential for a positive bias in both valence and arousal that may be caused by the use of virtual reality, particularly in subjects who have not previously had experience with it. Research has shown that VR can elicit significantly higher levels of physiological arousal when compared to an imaginative approach, as outlined in reference [158]. Secondly, this initial experience serves as a tutorial, teaching the mechanics of interaction within the virtual environment (See Figure 3.15).

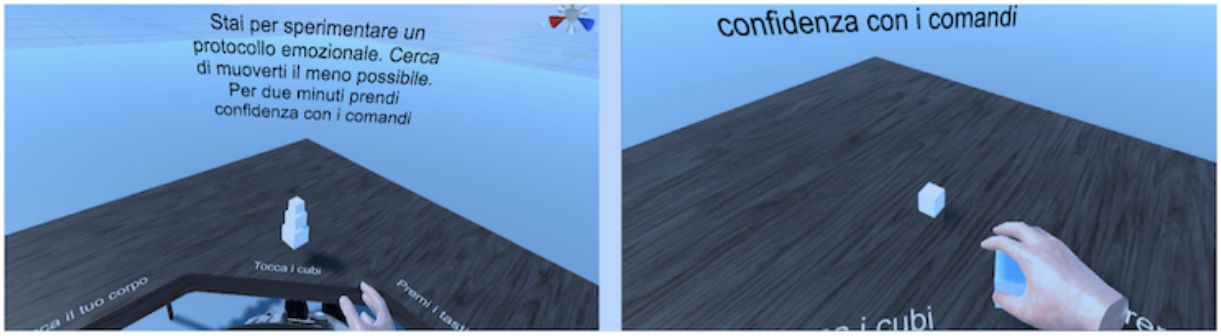

Figure 3.15: Two screenshot from the initial scene of adaptation of the protocol

Neutral scenes

The aim of these experiences is to reset the emotional state at each stage of the protocol. Instead of using a traditional monochromatic baseline as is typical in 2D protocols, the approach taken for the rooms was to minimize the stimuli delivered to the subjects while incorporating a 3D immersive environment. The settings are designed to simulate real-world neutral scenarios that provide a range of stimuli with the objective of evoking a neutral response from the participants. The four scenes can be observed in Figure 3.16 and depict two living rooms and two offices. The static visual elements tend to feature angular shapes with straight lines. The color scheme was set to a gray hue with a slightly desaturated tone.

Sadness

The virtual reality experience presents a scenario in which the participant is traveling in a vehicle during a rainy day, ultimately coming to a halt in traffic resulting from a car accident. The visual aesthetic features a dominant purple-blue hue with a low level of brightness. The auditory elements of the scene primarily consist of the sound of rain, the engine of the car, and the crying of a woman near the location of the accident. Some of the audio sources are not within the field of view of the participant. To augment the emotional response, a constant, royalty-free sad song, is played throughout the experience (See Figure 3.17).

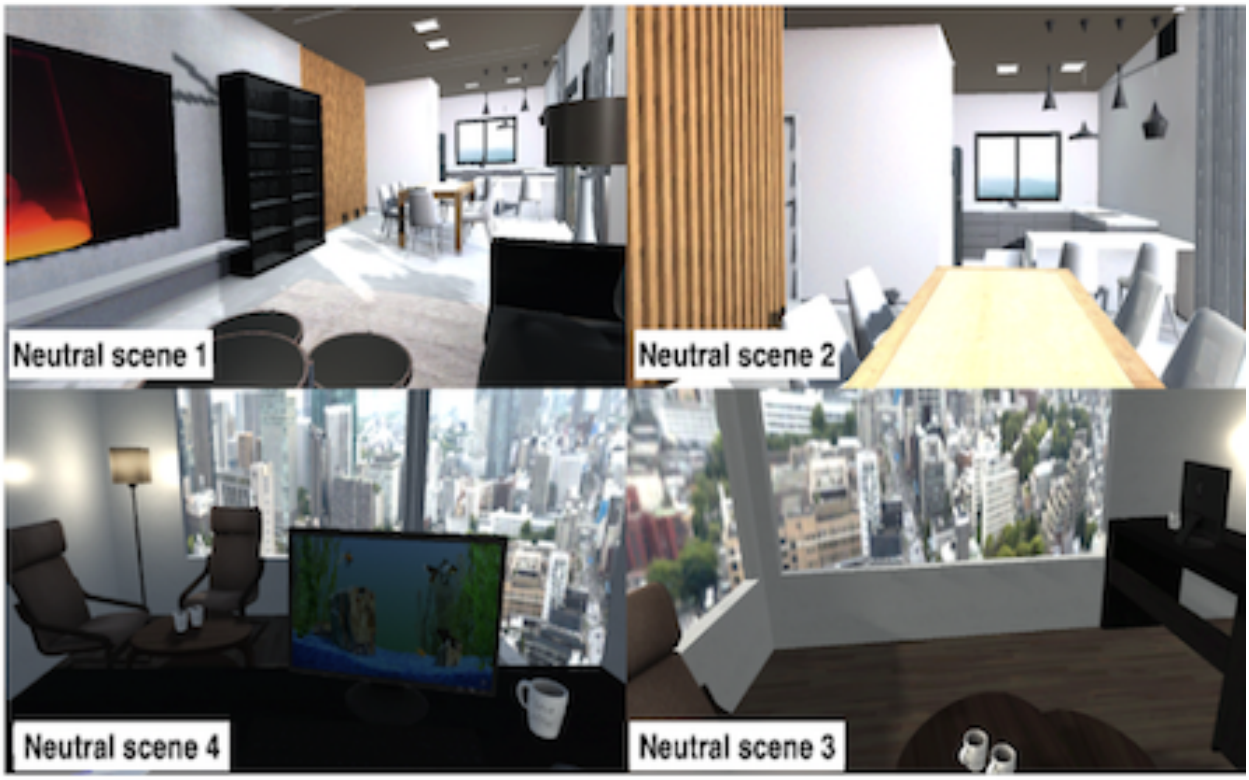

Figure 3.16: The neutral scenes that separate the emotionally charged scenes are shown.

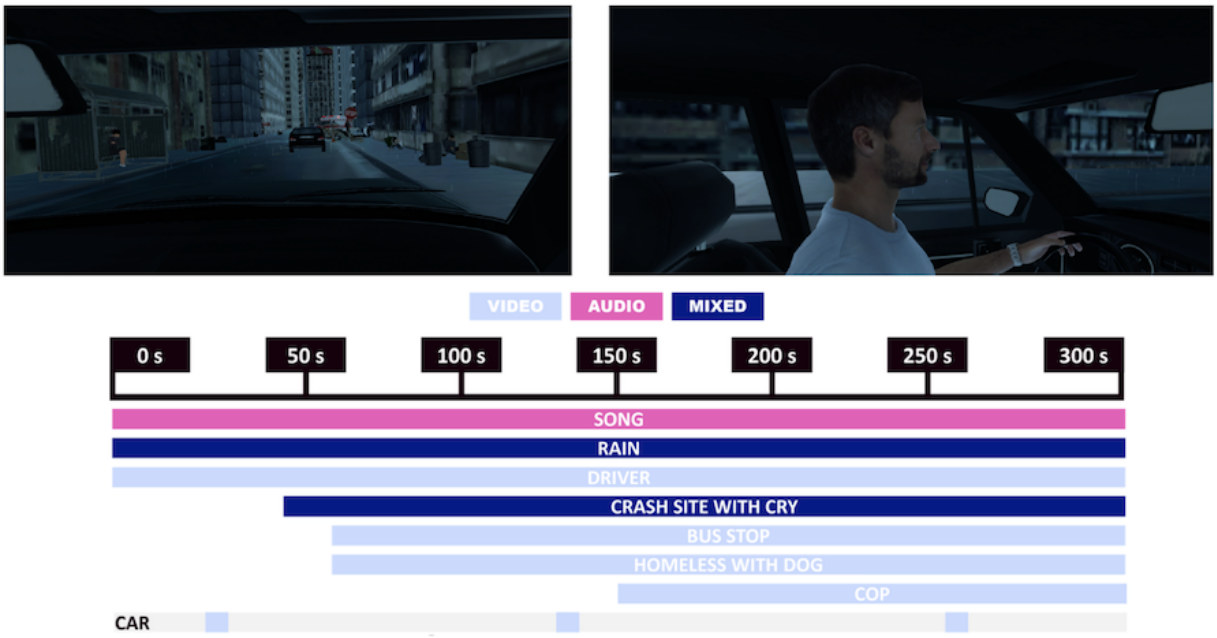

Figure 3.17: Two screenshots from the sadness-inducing scene along with all the stimuli used in the scene presented in chronological order.

Relaxation

The setting for this stage of the protocol was designed to resemble a desert island in the ocean. The static visual elements primarily feature rounded shapes. Non-static visual elements, such as

the ship on the horizon, clouds, and slowly moving seagulls, are also present. The auditory elements include the sound of a gentle breeze, birds chirping on a distant lighthouse, and a background song (See Figure 3.18).

Figure 3.18: Two screenshots from the relaxation-inducing scene along with all the stimuli used in the scene presented in chronological order.

Happiness

The elicitation of happiness in the protocol was primarily based on the guidelines provided by the IAPS dataset. The images selected tend to be those associated with appetitive motivation, such as those depicting sexual content, food, and the instinctual drive to protect one's offspring. In order to evoke the desired emotional response, the subject's avatar is placed in an everyday setting, such as a kitchen, surrounded by realistic depictions of food and beverages during a celebratory party. The color scheme of the scene is primarily in the green-yellow range, with high saturation and moderate brightness. The sound of a glass bottle being dropped twice is included, along with a happy background music and low chatter, all of which are visible to the subject (See Figure 3.19).

Fear

The individual in question is seated in a bathtub within an aging bathroom located within a structure purported to be haunted, during a tumultuous thunderstorm. Of particular note are the alleged paranormal occurrences, such as the manifestation of a dark ghost and the presence of a levitating jar. Additionally, two specific sound effects (designated as IADS A3-110 and A4-296) have been included in the scene. The visual cues utilized in this scene are primarily characterized by angular shapes and a red-yellow color palette, with the overall ambiance being dim and lacking in saturation. The motion and transitions within the scene are rapid, and the sounds audible in the scene are pronounced and located just outside the field of view, yet still in close proximity to the

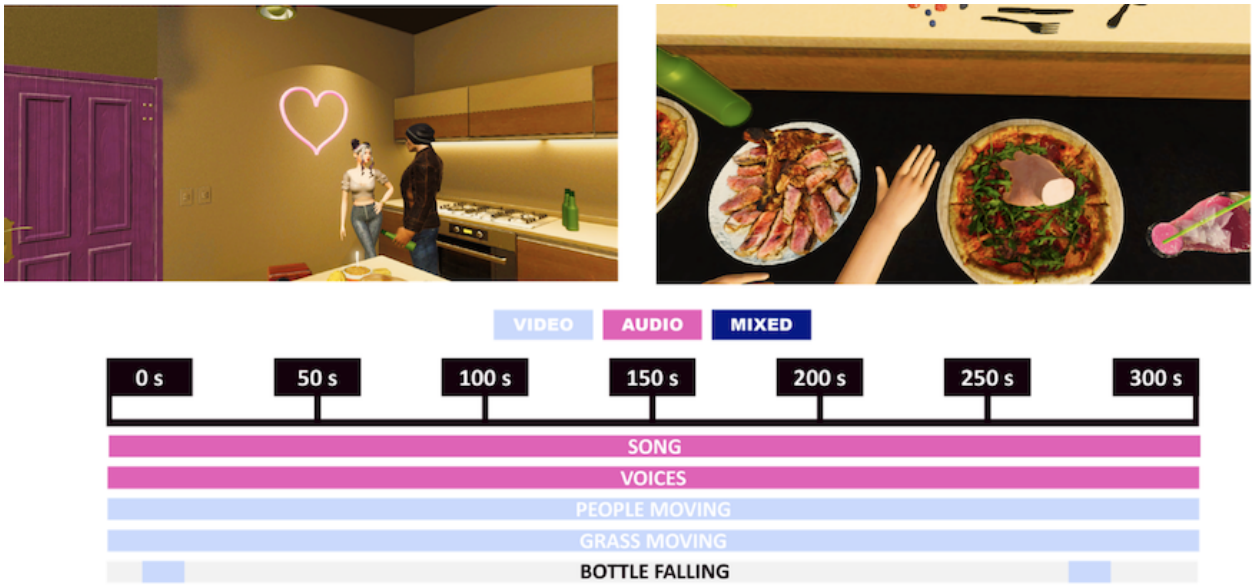

Figure 3.19: Two screenshots from the happiness-inducing scene along with all the stimuli used in the scene presented in chronological order.

subject (See Figure 3.20).

Figure 3.20: Two screenshots from the fear-inducing scene along with all the stimuli used in the scene presented in chronological order.

Figure 3.21 shows all the scene together.

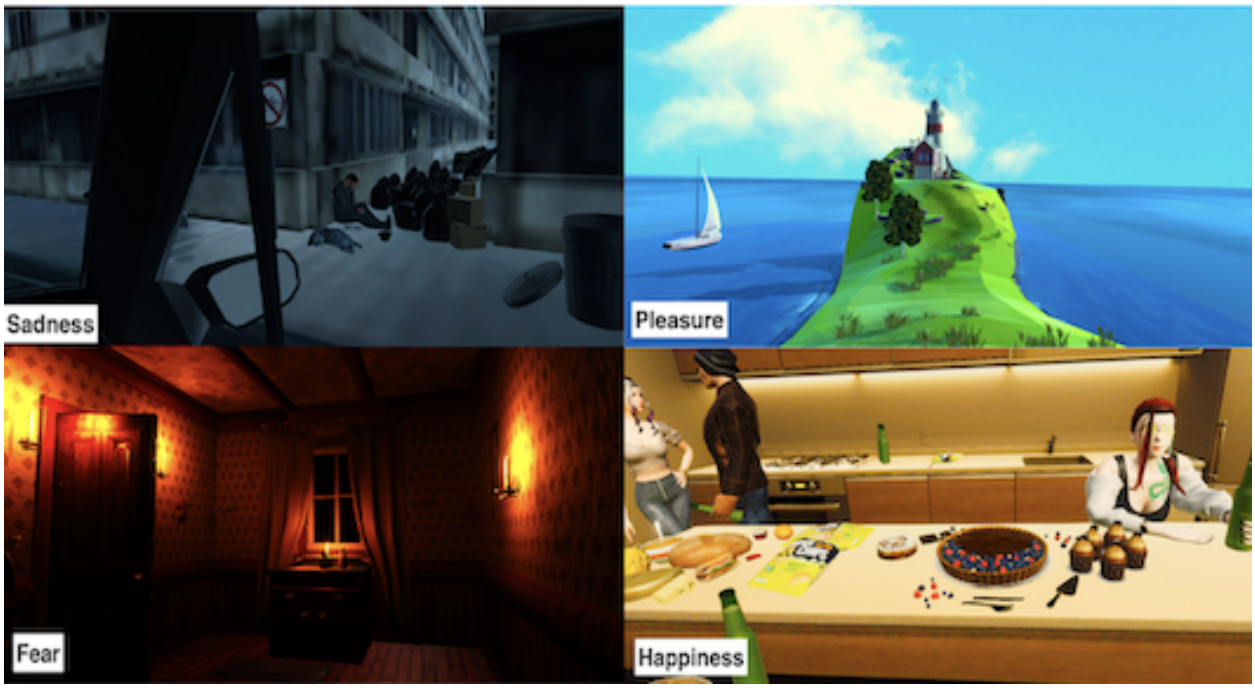

Figure 3.21: One screenshot from each emotional scene.

3.3.3 Statistical analysis and Classification

In order to determine the normality of the data, the Shapiro-Wilk test was applied to each individual feature value in the four emotions. As it was discovered that at least one variable was not normally distributed, a non-parametric and pair-wise Friedman's test was employed. When a comparison was found to be significant, a multcompare comparison was performed to identify which specific phases displayed significant differences. In all cases, the Tukey correction method was implemented as a correction measure.

43 subjects were acquired but only 37 subjects were selected to be part of the dataset for the classification of the four emotions. Several participants were not able to be included in the study as a result of the intensive processing required for their signals, which posed a risk of compromising the validity of the results. The most frequent issues were caused by difficulties encountered during the data acquisition process, such as movement artifacts or network disturbances. Furthermore, some of the subjects had an abnormal number of ectopic beats, which resulted in their exclusion from the study. The primary signals that were analyzed were the ECG and BVP, which resulted in the loss of typical morphology for prolonged periods and the difficulty in identifying specific events. As a rule, when 10% of the signal was found to be corrupted, the subject was discarded from the study.

For the classification purpose, the aim is twofold: not only to attain the highest classification score feasible, but also to adhere to a coherent and adaptable pipeline that can provide an impartial evaluation of the actual efficacy of the employed models and can be reproducible. To accomplish these objectives, the initial phase entailed segregating certain subjects from the primary dataset, to employ them as an unobserved test for the designated models. The distribution is comprised of

80% for training and 20% for testing, with subjects chosen using a randomized sampling method. The dataset's definition permits classification of all 4 emotions or solely valence and arousal by utilizing distinct coding for the targets. As analogous methodologies have been employed in all three scenarios, they will be collectively illustrated. Employing a modular strategy facilitated the parameterization of the process and independent regulation of the data transformations applied. The initial step involved dropping features via a designated threshold for internal correlation, usually around 80%. The feature that possesses the highest correlation with all other features in the dataset is eliminated as a result. The data can undergo standard scaling, and PCA can be executed by specifying the preferred number of output features which manage to explain at least the 80% of the initial variance. Utilizing a particular model as input allows for a grid search to tune the hyper-parameters, ultimately generating the best threshold for optimizing the Receiver operating characteristic curve (ROC) curve. Given balanced data, the evaluation metric employed to assess the models is accuracy. To validate the models, a LOO approach was adopted. The search output provides the chosen hyper-parameters, validation accuracy, and ROC threshold value (if selected for binary arousal and valence classification). Four models have been selected for comparison as they are among the most commonly used models in the literature for emotion recognition: KNN, LDA, LR and SVM. For classification purposes, three distinct feature selection procedures were employed. Sequential Feature Selection (SFS), K best (KB), and SM were utilized in this project. Below is a summary of the different feature selection algorithms and their respective functions:

- SFS: It entails forming a feature subset greedily via the addition (forward selection) or removal (backward selection) of features. At each step, the estimator selects the most suitable feature to add or remove, based on the cross-validation score of an estimator.
- KB: It scores the features by computing the ANOVA F-value for the provided sample, ultimately selecting the most significant features with the highest scores.
- SM: This method is detailed in section 3.1.3.

Figure 3.22 the machine learning pipeline use in this project.

3.4 Signal processing and analysis

In this section, a brief introduction of each signal, including the method of acquisition, will be provided before conducting a comprehensive examination of the physiological signals acquired during the protocols outlined in the preceding sections. Each signal will be thoroughly analyzed with particular emphasis placed on the processing and extraction of the relevant features.

3.4.1 ECG

Signal characterization and acquisition

The ECG is a graphical representation, recorded on the surface of the body, of the electrical activity that occurs in the heart during the contraction and relaxation activity. It is composed of a sequence of waves and complexes. The P wave represents the depolarization of the atria and

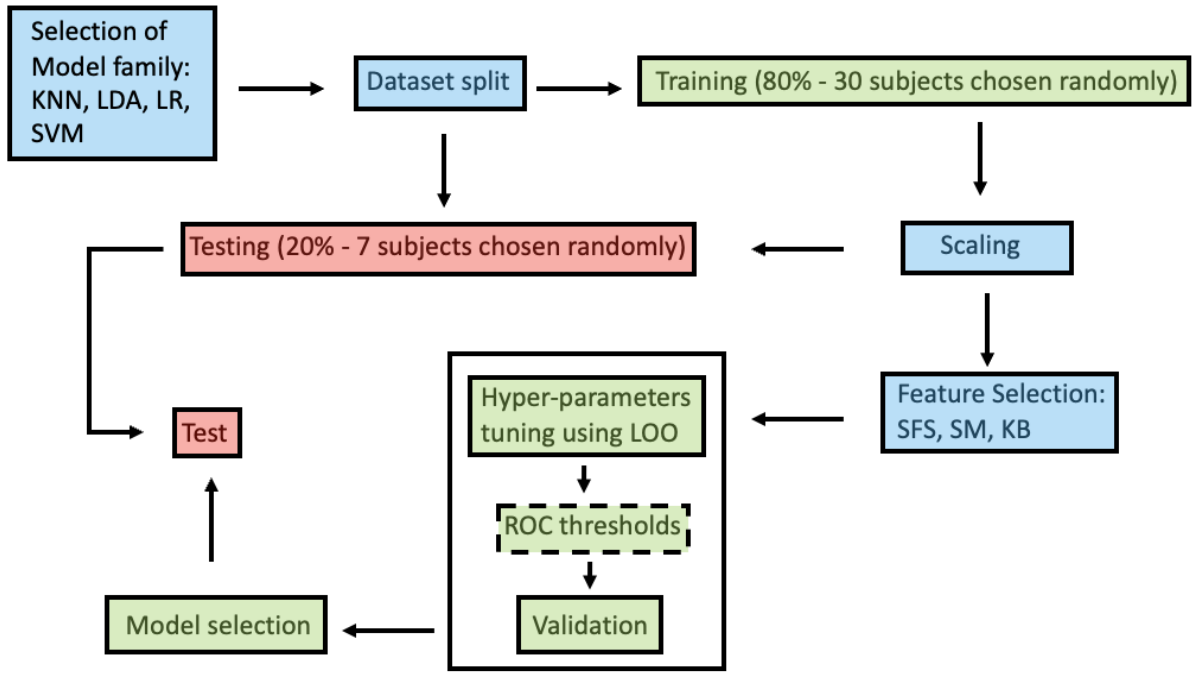

Figure 3.22: Flowchart of the followed machine learning pipeline.

originates in the sinoatrial node. The QRS complex reproduces the depolarization of the ventricles, during this time frame the atrial repolarization is also present but it is masked as it is of much smaller amplitude. Finally, the T wave is present and shows the ventricular repolarization. Figure 3.23 shows the normal waveform for the ECG signal. The signal was acquired using the Procomp Infinity device. The standard ECG electrode placement procedure necessitates the positioning of the negative electrode on the right shoulder, the positive electrode on the lower center or left side of the chest (xiphoid process), and the ground electrode on the left shoulder.

Processing

The initial processing of the ECG signal (fs =2048 Hz) involved application of a zero-phase low-pass Butterworth filter of fourth order, followed by a down-sampling operation at a rate of 250 Hz. The R peaks on the ECG signal were then detected using a Pan-Tompkins algorithm [159]. With the extracted R peak information, a variety of features were computed, including frequency domain features obtained through autoregressive modeling of the RR series using the Yule-Walker method. To ensure optimal performance, the order of the autoregressive model was carefully selected through examination of the residuals, with the goal of achieving white residuals and/or minimizing the Akaike information criterion, within a range of 7 to 15.

Features extraction

Below, all features calculated from the ECG signal will be displayed.

* AVNN: It is obtained as the mean of the HRV time series in the 5-minute scene by 1000 to

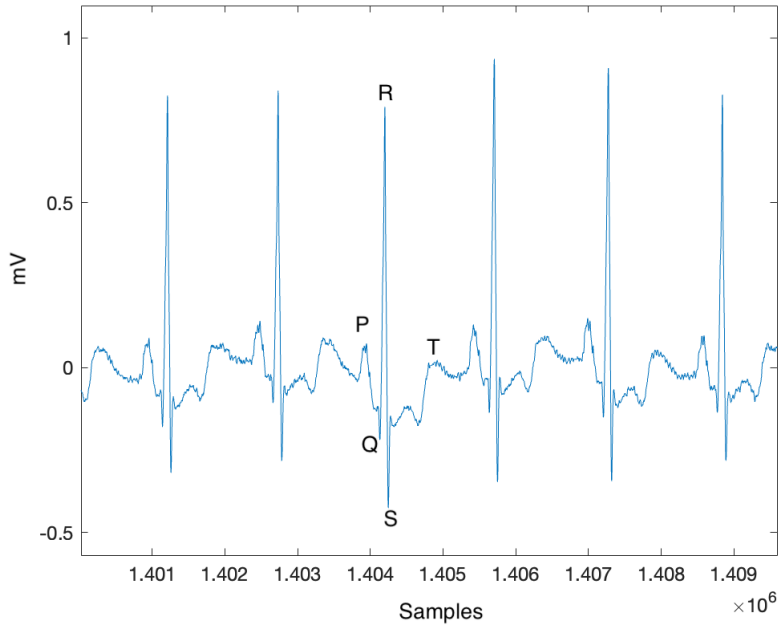

Figure 3.23: The raw ECG signal samples extracted from the Procomp Infinity device with an indication of the main waves.

convert it to milliseconds.

* SDNN: t is obtained as the standard deviation of the HRV time series multiplied by 1000 to convert it to milliseconds.

* RMSSD: is the square root of the mean squared differences between successive RR [ms] intervals.

* LogRMSS: it is the logarithm of the square root of the mean squared differences between successive RR [ms] intervals.

* SD1: the standard deviation related to the points that are perpendicular to the line-of-identity RRn+M = RRn. It describes the HRV short- term variability (computed by PCA).

* SD2: The Standard Deviation which quantifies the dispersion of data points along the identity line and characterizes the long-term variability of the system (computed by PCA).

* SD ratio: the ratio between SD1 and SD2. This feature measures the balance between the HRV long and short-term variability.

* NN20: it is the Number of successive RR interval pairs that differ more than 20 ms.

* pNN20: it is the NN20 divided by the total number of RR intervals [%].

* NN50: it is the Number of successive RR interval pairs that differ more than 50 ms.

* pNN50: it is the NN50 divided by the total number of RR intervals [%].

* RRVLF: It is the integral of the power of the HRV spectrum in the range (0.003 Hz - 0.04 Hz) obtained using an AR model with a variable order obtained maximizing AIC figure of merit.

* RRLF: It is the integral of the power of the HRV spectrum in the range (0.04 Hz - 0.15 Hz) obtained using an AR model with a variable order obtained maximizing AIC figure of merit.

* RRHF: It is the integral of the power of the HRV spectrum in the range (0.15 Hz - 0.4 Hz) obtained using an AR model with a variable order obtained maximizing AIC figure of merit.

* LFtoHF: It is the ratio of LF and HF.

* LFnu: It is the ratio of LF and a normalization term (total power - VLF).

* HFnu: It is the ratio of HF and a normalization term (total power - VLF).

* LyapExp: It is the estimate of the largest Lyapunov exponent of the tacogram.

3.4.2 Univariate Point Process modelling

In the realm of statistics, point processes provide a probabilistic representation of the distribution of points in a specified space [160]. These processes are prevalent in a variety of systems, including the temporal and spatial distribution of neural spikes [161]. This thesis concentrates on modeling heartbeats as a stochastic point process, allowing for the continuous estimation of the average inter-beat-interval and related spectral indices. The log-likelihood function, represented by $L(N_{0:T})$, for a point process that describes a series of events (dN(t)) over a temporal interval T, is mathematically defined as:

$$L(N_{0:T}) = \int_0^T log(\lambda(t|H_t))\, dN(t) - \int_0^T \lambda(t|H_t)\, dt \tag{3.1}$$

The conditional intensity function (CIF), represented as $\lambda(t|H_t)$, represents the predicted frequency of event occurrences and it is linked to the probability of an event happening at a given time, given past events as follows:

$$\lambda(t|H_t) = \frac{f(t|H_t)}{1 - \int_0^t f(u|H_t)\, du} \tag{3.2}$$

Heart cells, as they trigger an action potential, follow a Gaussian random walk with drift [162]. As a result, the time between two consecutive heartbeats follows the Inverse Gaussian (IG) Probability Density Function (PDF) with mathematical expression as (f_{IG}):

$$f_{IG}(x|\mu, k) = \sqrt{\frac{k}{2\pi x^3}} \exp\left(-\frac{k(x-\mu)^2}{2\mu^2 x}\right) \tag{3.3}$$

and it is defined for $x > 0$, $\mu > 0$, and $k > 0$. The aforementioned definition can be applied to the analysis of inter-beat-intervals by considering a set of R-wave events u1, u2, ..., un, ..., uN observed within the temporal interval (0, T], with the waiting time until the next event represented as follow:

$$p_{IG}(t|H_{u_n}, \theta, k) = \sqrt{\frac{k}{2\pi(t-u_n)^3}} \exp\left(-\frac{k}{2}\frac{(t-u_n-\mu_{RR}(t,\theta,H_{u_n}))^2}{(t-u_n)\mu_{RR}(t,\theta,H_{u_n})^2}\right) \tag{3.4}$$

The expected value of the IG can be modeled using an autoregressive (AR) model of order p as follows:

$$\mu_{RR}(t) = \theta_0(t) + \sum_{i=1}^{p} \theta_i(t) RR_{n-i+1} \tag{3.5}$$

RRn is defined as the difference between consecutive inter-beat intervals (u_n-u_{n-1}) to estimate the time-varying spectral representation of the system ((S_{RR})) as $S_{RR}(f,t) = \frac{\sigma^2}{|1-\sum_{i=1}^{p} \theta_i(t) e^{-2\pi j f i \Delta t}|^2}$, enabling real-time ANS activity assessment. [163][164][165] support ANS impact on the heart being evaluated by modeling expected inter-beat-interval ($\mu_{RR}(t)$) through linear relationships and optimizing weights (θ) through maximum likelihood.

Below are the features that were computed through the point process.

* $\mu(t)_{RR}$: Average RR interval of Inverse Gaussian probability distribution.
* $\sigma^2(t)_{RR}$: RR interval variance of Inverse Gaussian probability distribution.
* RRTOT(t): Total power of the continuous RR interval spectrum.
* RRVLF(t): Very low frequency power of the continuous RR interval spectrum.
* RRLF(t): Low frequency power of the continuous RR interval spectrum.
* RRHF(t): How frequency power of the continuous RR interval spectrum.
* RRLFn(t): Normalized low frequency power of the continuous RR interval spectrum.
* RRHFn(t): Normalized high frequency power of the continuous RR interval spectrum.
* RRLFtoRRHF(t): Ratio between RRLF(t) and RRHF(t) .

Figure 3.24 illustrates the real-time monitoring of HRV indices, which have been obtained through the point process modeling of a part of the protocol described in section 3.2 for subject 1.

The primary advantage of this approach is that it enables the calculation and accurate estimation of cardiovascular features and indices related to the ANS over time. This approach benefits from an initial buffer window for training the model, which allows for the calculation of metrics on short time windows, as demonstrated in the experiments conducted in this thesis. Such calculations would be impossible using traditional methods, which are unreliable in accurately estimating features on short time windows.

3.4.3 BVP

Signal characterization and acquisition

The technique for measuring BVP is a non-invasive, optical method employed to identify variations in blood volume within the peripheral circulation. The operation of the sensor system is based on the measurement of light absorption, utilizing light-emitting-diodes emitting at red and infrared wavelengths, to illuminate the skin. Changes in light intensity, corresponding to slight variations in tissue blood perfusion, are detected by a photodiode, resulting in a measurement of blood volume pulse. The signal comprises various reference points, including onset, systole, dicrotic notch, and

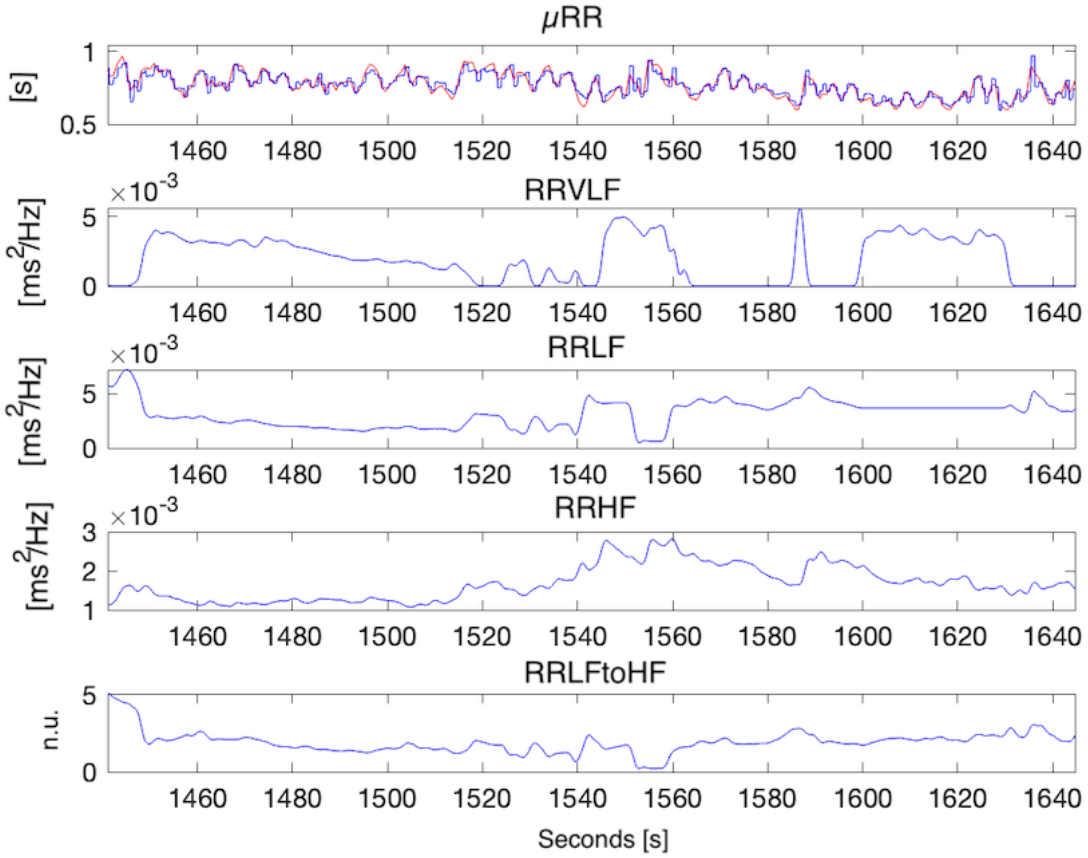

Figure 3.24: Real-time tracking of HRV, utilizing a point process modeling approach. The top panel displays the raw RR series in black and the modeled RR series (μRR) in red. The subsequent panels show the key HRV indices computed.

diastole shown in Figure 3.25. Specifically, the initiation of the BVP signal is referred to as the onset. The systolic phase, which is characterized by an elevation in blood pressure, is a result of the heart contracting and forcibly expelling blood towards the periphery during systole. The amplitude of this phase is influenced by factors such as the cardiac stroke volume and sympathetic activity of the nervous system. A slight increase in blood pressure, known as the dicrotic notch, occurs as a result of the closing of the aortic valve and is observed between the systolic and diastolic phases. The diastolic phase, marked by a decrease in blood pressure, is caused by blood returning to the heart from the periphery during diastole.

The signal was acquired using the Procomp Infinity device. The sensor for measuring BVP is applied to the palmar surface of the fingertip by utilizing an elastic strap.

Processing

In order to analyze the BVP signal (fs = 2048Hz), we utilized a low-pass Butterworth anti-aliasing filter of 4th order and zero-phase with a cut-off frequency of 25 Hz. Subsequently, we downsampled the signal at a rate of 250 Hz. Furthermore, utilizing the R-peaks identified in the ECG signal, we were able to locate and extract the amplitudes of the systolic, diastolic, and

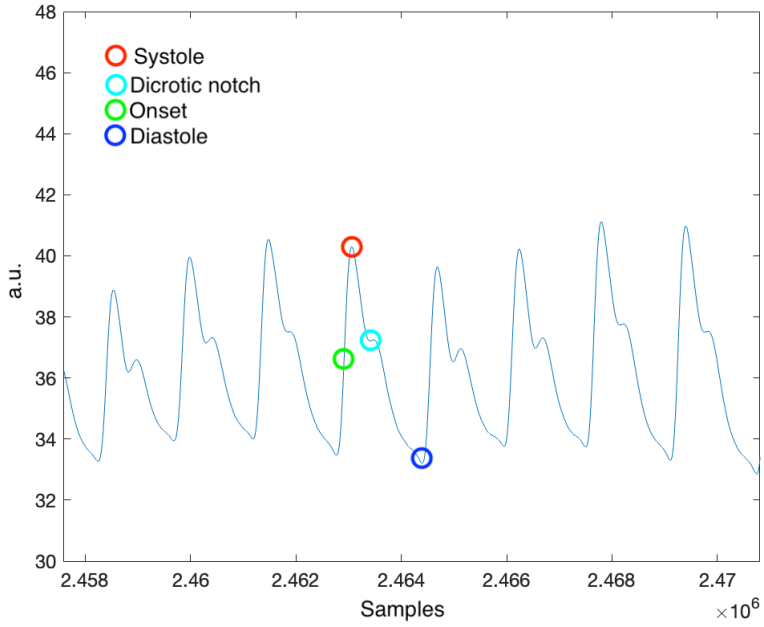

Figure 3.25: Raw BVP signal samples extracted from the Procomp Infinity device with an indication of the fiducial points.

onset fiducial points within the BVP signal. We precisely determined the systolic and diastolic values of the BVP by identifying the maximum and minimum values between consecutive R-peaks, respectively. Furthermore, the onset values were determined by pinpointing the inflection points between each systolic and diastolic location. To guarantee the precision of these determinations, we performed a manual review and rectified any errors by an in-house software whose whose interface is shown in Figure 3.26. The software allows for manual scrolling of the time axis of the synchronized ECG and BVP signals and to examine if there are any annotation errors, and if so, to manually correct them and then save the new annotations. In particular, the interface also allows for the display of the RR series and the pulse arrival time, which is the time elapsed between a peak R in the ECG and the corresponding onset of the pressure wave in the BVP. In this manner, anomalous spikes in the RR series or PAT are often indicative of errors in the signal annotations.

The initial annotations produced by the first processing step were not always adequate for the feature extraction phase, due to the presence of abnormal drifts in the signals. These drifts caused the BVP signal to deviate from its typical morphological form. To resolve this issue, a filter was developed to estimate the missing annotations. The correction procedure is based on the direct correspondence between heartbeats and the BVP waveform. If a segment of the BVP signal was labeled as anomalous through manual inspection, its associated R-peaks were not annotated. The correction algorithm comprised of searching for RR intervals longer than 1.2 seconds in the entire ECG signal. If a gap was found, the R-peak was first identified using the Pan-Tompkins algorithm, and then the timings of the three BVP fiducial markers were estimated by computing the average point-to-point interval of the previous three beats, utilizing the stored index of the beat. The pressure value was estimated by taking into account the previous two beats and the subsequent beat,

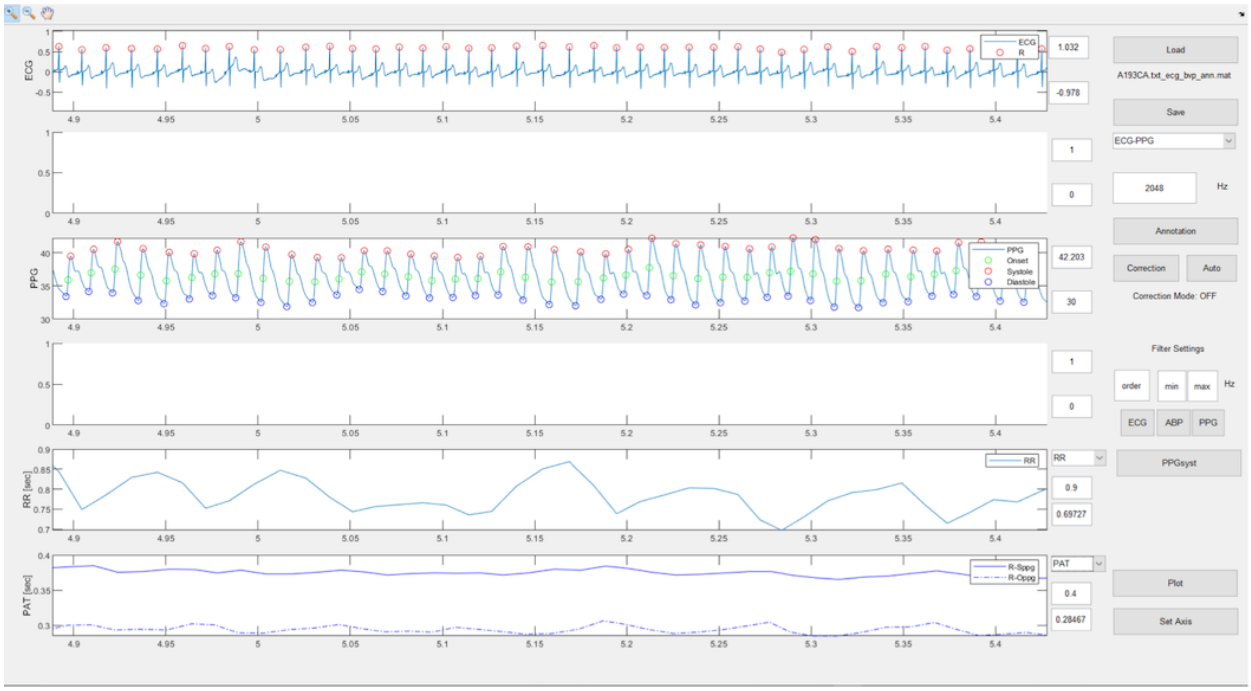

Figure 3.26: The interface used to manually correct the ECG and BVP annotations. At the top, it is possible to see the ECG signal, in the third row the BVP signal, in the fifth the RR series, and in the sixth the PAT calculated using systole in the case of the solid blue line and diastole in the case of the dotted blue line.

weighted by their proximity to the targeted R-peak. Figure 3.27 shows an example of correction for the three occurrence and amplitude of the three fiducial points.

As stated above, the BVP waveform illustrates how the volume of blood in a specific tissue (in our case the finger) fluctuates with each heartbeat. This pulse flow waveform is heavily impacted by the interactions between the ventricles and the blood vessels, much like the forward and backward pressure waves of the pulse [166] [167]. In [168] researchers demonstrated how the changes detected in the BVP amplitude taken at the fingertip are correlated with variations in arterial blood pressure caused by alterations in the vascular tone. In particular, the higher the amplitude of the arterial blood pressure, the lower the amplitude of the BVP waveform. For this reason from now on we will refer to the BVP amplitude as the peripheral blood pressure where low amplitude is mainly related to a symphatetic activation while high amplitude values are linked to symphatetic deactivation.

Features extraction

By analyzing the relationship between the BVP and ECG signals, we extracted two features: the mean amplitude difference between each systolic and its corresponding diastolic value, referred to as the Mean Volume Amplitude Index *(VP)*, and the Mean Pulse Arrival Time *(PAT)*, being the mean temporal difference between each onset value on the BVP signal and its corresponding R-peak on the ECG signal. In order to compute *(PAT)*, we used onsets in relation to diastoles or systoles as these points tend to be more reliable and less susceptible to uncertainty in the most tumultuous segments of the signal.

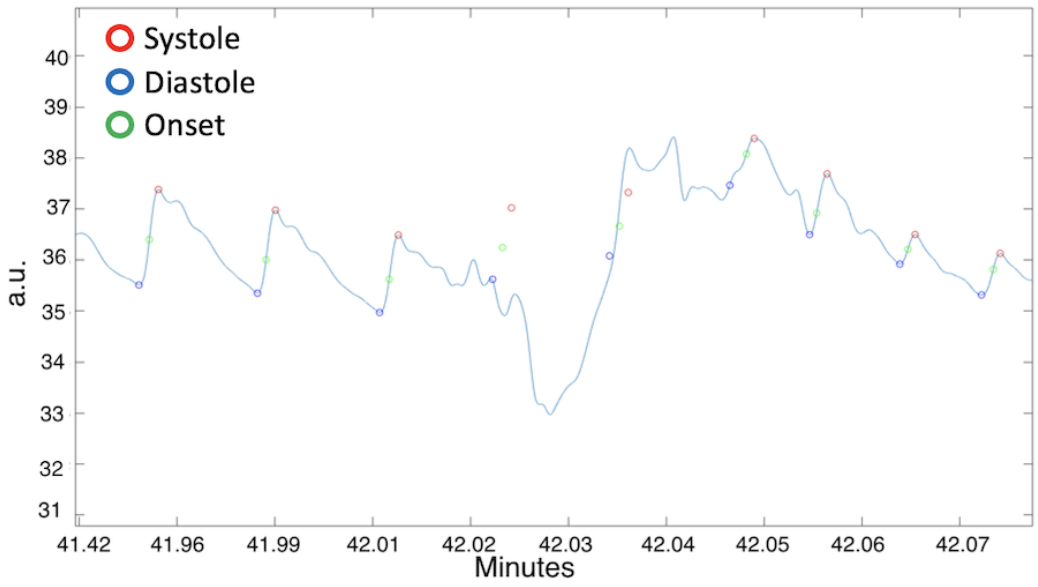

Figure 3.27: The result of the BVP correction algorithm when the signal morphology was contaminated by noise. The three BVP fiducial markers were estimated by computing the average point-to-point interval of the preceding three beats.

3.4.4 GSR

Signal characterization and acquisition

The GSR, also referred to as Electrodermal Activity (EDA), serves as an indicator of the electrical conductivity of the skin through the measurement of skin resistance. The human body's thermoregulation is controlled by the eccrine sweat glands, which are solely innervated by the sympathetic branch of the ANS. These glands act as variable resistors and when sympathetic activity increases, the sweat glands become activated and result in a decrease in skin resistance as sweat secretion increases, leading to an increase in skin conductance.

The GSR encompasses two components: the Tonic Component and the Phasic Component. The Tonic Component, also known as the Skin Conductance Level (SCL), represents the slowly fluctuating baseline level of skin conductance and is influenced by the individual's general level of sympathetic nervous system activation, which can vary in minutes. SCL values tend to be higher during a state of relaxation and lower when the individual is stressed and experiencing increased sweating, resulting in decreased skin resistance and significant differences across individuals. The Phasic Component, referred to as the Skin Conductance Response (SCR), superimposes upon the tonic component and refers to the rapid fluctuations of electrodermal activity. These changes, occurring within seconds, can be elicited by identifiable stimuli within a specified time window (1-3 to 1-5 seconds after stimulus onset) or occur spontaneously.

The signal was acquired using the Procomp Infinity device. A minuscule electrical voltage is applied through two electrodes, typically attached to two fingers of a single hand, in order to establish an electrical circuit in which the individual acts as a variable resistor. The fluctuation of conductance, which is the reciprocal of resistance, is then calculated in real-time.

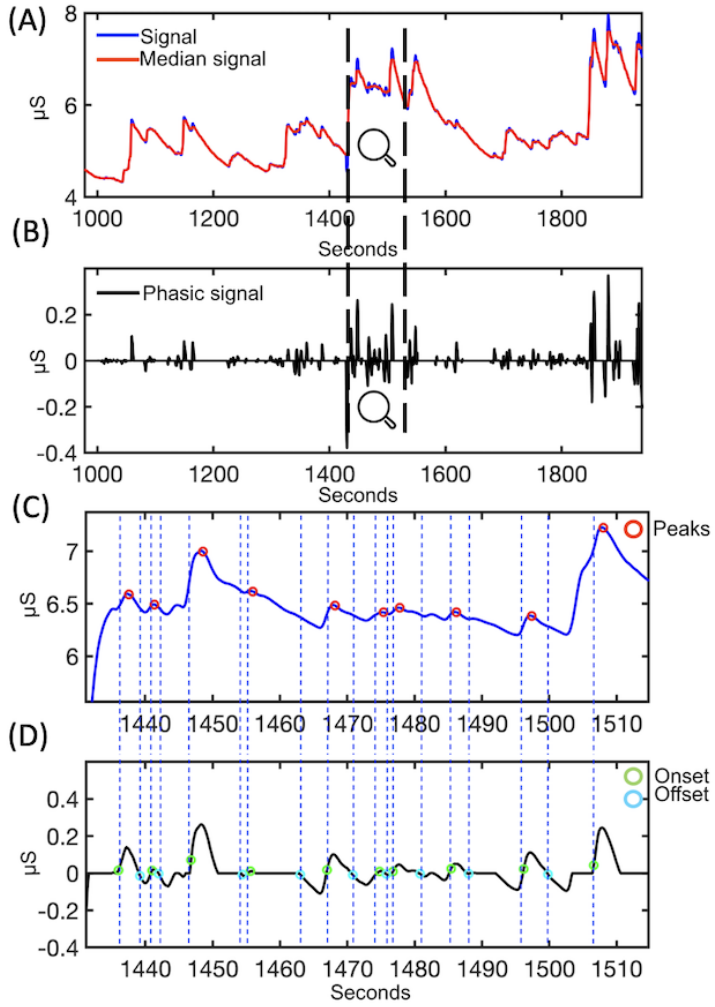

Figure 3.28: In Panel (A), the raw and filtered GSR signals are depicted in blue and red, respectively. In Panel (B), the phasic component of the GSR is presented. The filtered GSR signal with peak amplitudes is shown in red in Panel (C), while Panel (D) displays the phasic component of the GSR with onset and offset marked in green and light blue, respectively. It should be noted that Panels (C) and (D) are focused on the magnified portion of the GSR signals as indicated by the magnifying lens in Panels (A) and (B).

Processing

The Phasic Component of GSR (fs = 256 Hz) signal was extracted by first implementing a low pass Butterworth filter of 4th order with a cutoff frequency of 2 Hz, and then downsampling the signal to a rate of 5 Hz. Subsequently, a median filter was applied to the signal, which replaced each sample with the median value of the neighboring samples within a 4-second window centered around the current sample [169]. The result was a median signal that was subtracted from the filtered signal to derive the Phasic Component. GSR peaks, signifying spikes in the eccrine glands, were identified by detecting local maxima in the filtered signal that occurred between the onset (amplitude greater than 0.01 μS) and offset (amplitude less than 0.0 μS) of the Phasic Component. Figure 3.28 shows all the process to extract GSR peaks on the signal.

Features extraction

Below, all features calculated from the GSR signal will be displayed.

* Avg amplitude peaks: It is computed over the found peaks as the average amplitude in the time window.

* Sd amplitude peaks: It is computed as the standard deviation of peaks' amplitude.

* Avg rise time: It is computed over the found peaks as the average distance in seconds between the peak and the onset.

* Avg recovery time: It is computed over the found peaks as the average distance in seconds between the offset and the peak.

* N peaks: It is computed as the number of peaks identified in the time window.

* Max sign amplitude: it is the maximum signed amplitude between 2 consecutive extremes of bandpass GSR (maximum range with a sign between a consecutive max-min).

* Avg der amplitude: It is the average of the first derivative of Bandpass filtered GSR.

* Sd der amplitude: it is the standard deviation of the first derivative of Bandpass filtered GSR.

* Max der amplitude: It is the maximum of the first derivative of Bandpass filtered GSR.

* Slope GSR: It is the slope of the low pass filtered GSR taking into account both tonic and phasic components.

* Avg GSR: It is the mean of the low pass filtered GSR taking into account both tonic and phasic components.

* Sd GSR: it is the standard deviation of the low pass filtered GSR taking into account both tonic and phasic components.

* Avg abs1: It is the mean of low pass filtered GSR first derivative under absolute value taking into account both tonic and phasic components.

* Avg abs1 norm: it is the mean of low pass filtered GSR first derivative under absolute value and normalized with the standard deviation taking into account both tonic and phasic components.

* Avg abs2: It is the mean of low pass filtered GSR second derivative under absolute value taking into account both tonic and phasic components.

* Avg abs2 norm: it is the mean of low pass filtered GSR second derivative under absolute value and normalized with the standard deviation taking into account both tonic and physic components.

* Env: It is the mean the envelope of the phasic component.

3.4.5 PUPIL

Signal characterization and acquisition

The pupil is a circular opening in the iris, a muscular ring that regulates the amount of light entering the eye by adjusting the size of the pupil. The pupil is located at the center of the iris and allows light to reach the retina. From an optical perspective, the pupil functions as an aperture and the iris as a diaphragm. Variations in pupil diameter are primarily caused by changes in environmental lighting conditions and are due to the activation of the nervous system, which is not directly related to vision control.

To record pupillometric data, the Tobii Pro X2 Compact eye-tracker was used, which has a sampling rate of 60 Hz. The Tobii Pro Lab software was utilized to execute the experimental design, perform a calibration of the eye-tracker using 13 points, and to collect and analyze the data.

Processing

Despite the filtering options available in Tobii Pro Lab software, including a built-in blink removal function, the raw data (fs = 60 Hz), pre-processed with blink compensation, were chosen to be further processed. This decision was made to retain as much valuable information as possible and to apply a custom cleaning technique to the data. First, to analyze the samples, data points with diameters less than 2 mm or greater than 8 mm were designated as "NaN" to indicate they were treated as blinks. This was done as diameters outside the range of 2-8 mm are considered non-physiological. Second, the artefacts due to errors in acquisition were removed: sudden increases or decreases of more than 0.375 mm within a 20 ms interval were judged as artefacts as well [170] [171]. The process of sample replacement and interpolation was designed to address instances of blinks that were detected in the recorded pupillometric data. If a blink was detected in one eye, the sample value for the other eye was substituted. However, if both eyes displayed blinks, a cubic spline interpolation of the signal was performed to estimate the missing data. This method of blink compensation allowed the preservation of as much information as possible from the raw data. After the blink removal process was completed, the samples were filtered using a 4th-order zero-phase low-pass Butterworth anti-aliasing filter, which was designed to effectively eliminate any high-frequency noise that may have been present in the signal. The filter had a cut-off frequency of 5 Hz, which was set to allow for the retention of relevant information in the signal. Finally, the signal was downsampled to a rate of 10 Hz, which was determined to be an appropriate rate for pupillometric data analysis. Figure 3.29 shows the raw and processed signal. The spectral analysis was conducted by computing the Welch's periodogram on the detrended signal using a 1.875-second Hamming window and an overlap of 50%.

Features extraction

With regard to the pupillary signal, upon availability of samples from both eyes, we have taken the average of the diameter values from both eyes to obtain an average diameter. The features were then extracted from the averaged signal of both eyes. The computation of six distinct features was carried out: the Mean Diameter (AVD), the Diameter Standard Deviation (SDD), the Power Spectral Density of the Diameter in the Low Frequency range (0.05-0.15 Hz) denoted as DLF, in the High Frequency range (0.15-0.45 Hz) denoted as DHF, in the Very High Frequency range (0.45-1.5

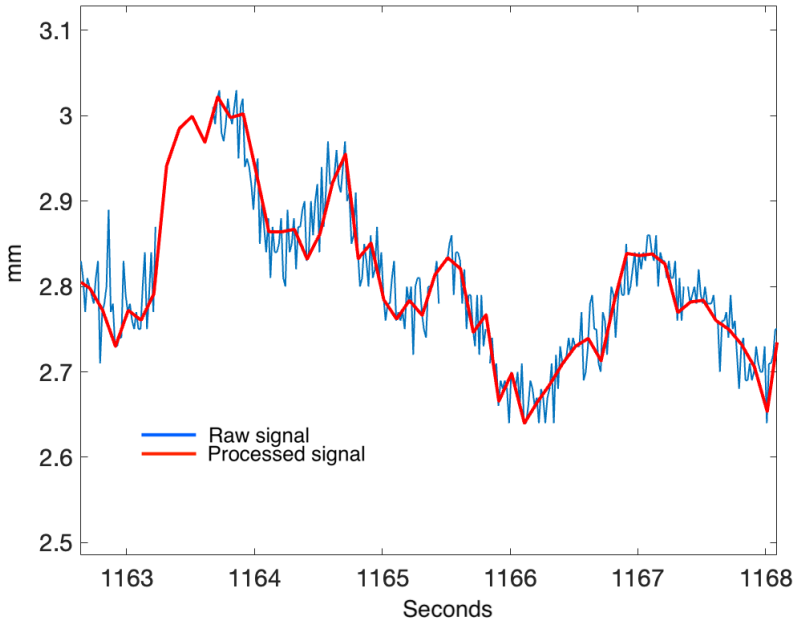

Figure 3.29: The raw pupil signal in blue and the processed signal in red. It can be seen that there are parts of the raw signal that have gaps due to blinks, however, these gaps have been interpolated in the red signal.

Hz) denoted as DVHF, and the ratio of DLFtoDHF.

3.4.6 RESP

Signal characterization and acquisition

The act of breathing, a fundamental physiological function in the human body, refers to the inhalation and exhalation of air into and out of the lungs, which plays a crucial role in facilitating the exchange of gases, primarily oxygen and carbon dioxide. Additionally, this process is vital in maintaining the acid-base balance and other important homeostatic mechanisms within the body, even during stressful situations.

There exist a multitude of methods for obtaining the respiratory signal, some of which are non-invasive and comfortable for the subject. These include the use of chest straps to measure movements, strain sensors, sensorized T-shirts equipped with piezoresistive electrodes, and systems based on measurement of respiratory sounds.

In the present study, the respiratory signal was acquired using using the Procomp Infinity device, specifically a sensor that comprised of a long Velcro strap. This strap was designed to be securely fastened around the chest or abdomen of the participant.

Processing

In our study, the respiration signal was subjected to a zero-phase digital low-pass filtering procedure, utilizing the Parks–McClellan algorithm [172], in order to isolate and capture the desired frequency components. The filtering process involved setting the cutoff frequency to 1 Hz. Once

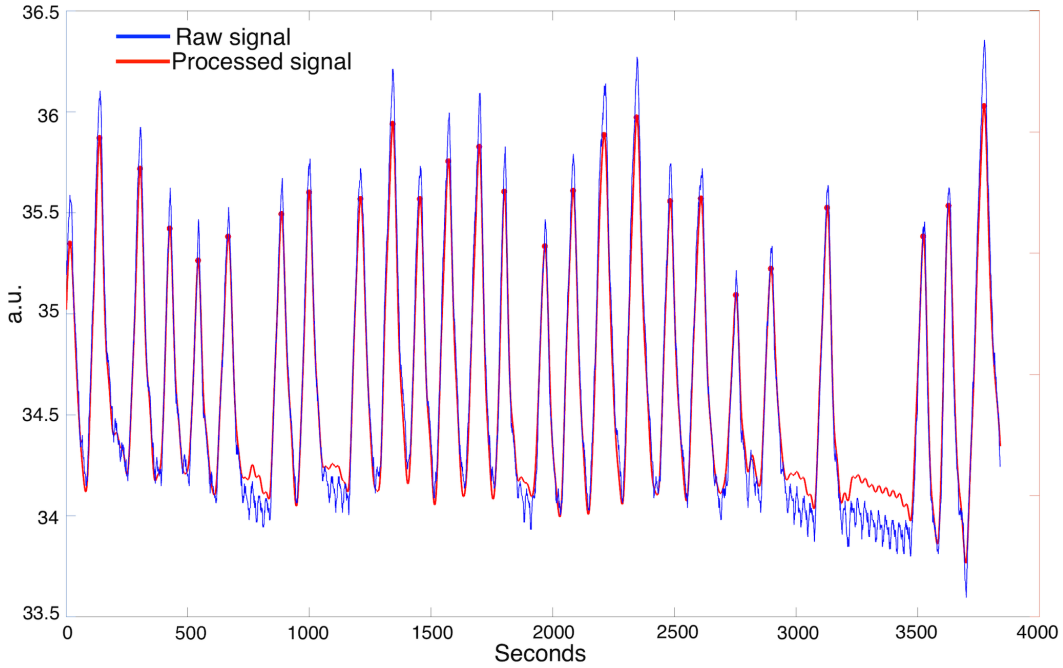

Figure 3.30: The figure displays the raw respiratory signal in blue and the processed signal in red. The red dots highlight the inspiratory peaks identified in the processed signal.

the signal had been filtered, a thresholding technique was applied in order to identify the maximum and minimum values present in the signal. Figure 3.30 shows the raw and processed RESP signal.

Features extraction

For the analysis of the respiration signal, two fundamental features were extracted from the signal: the Respiratory Frequency (fRESP) and the Mean Breath Amplitude (ampRESP). The Mean Breath Amplitude was calculated as the difference between the maximum and minimum values of each breath cycle.

3.4.7 Bivariate Point Process modelling

Similarly to the univariate approach, a bivariate ar model can be used to estimate the autonomic regulation of heartbeat due to changes in the respiration.

$$\mu_{RR}(t) = \theta_{110}(t) + \sum_{i=1}^{p} \theta_{11i}(t) RR_{n-i+1} + \sum_{j=1}^{q} \theta_{12j}(t) RESP_{n-i+1} \tag{3.6}$$

where RR_n=u_n-u_{n-1} and $RESP_n$ represents the breath generated by beat u_{n-1} and corresponding to RR_n.

Equation 3.6 delineates the ans's feedback branch's general functioning, separating the self-regulatory process from the Respiratory sinus arrhythmia's effects. To depict the closed loop system, another equation is required.

$$\mu_{RESP}(t) = \theta_{210}(t) + \sum_{i=1}^{r} \theta_{21i}(t) RR_{n-i+1} + \sum_{j=1}^{s} \theta_{22j}(t) RESP_{n-i+1} \tag{3.7}$$

The parameters θ_2 are estimated using weighted least squares, assuming a Gaussian distribution of the RESP series. Moreover, in this second equation, the variable RR_n is replaced with the estimated expected RR interval at the moment of the RESP event $\mu_{RR}(t = t_{RESP_n})$.

Equation 3.7 outlines the global functionality of the ans's feedforward branch, separating the self-regulation of respiration from the effects induced by changes in heartbeat (FeedForward).

As per [163], closed loop spectral component estimates can be derived from bivariate autoregressive models, with the exception of the zero-order coefficients θ_{110} and θ_{220}:

$$\begin{bmatrix} RR(f) \\ RESP(f) \end{bmatrix} = \begin{bmatrix} H_{11}(f) & H_{12}(f) \\ H_{21}(f) & H_{22}(f) \end{bmatrix} \begin{bmatrix} w_{RR}(f) \\ w_{RESP}(f) \end{bmatrix} = \underline{H}(f) W(f) \tag{3.8}$$

$$\underline{H}(f) = \frac{\begin{bmatrix} 1 - \Theta_{22}(f) & \Theta_{12}(f) \\ \Theta_{21}(f) & 1 - \Theta_{11}(f) \end{bmatrix}}{(1 - \Theta_{11}(f))(1 - \Theta_{22}(f)) - \Theta_{21}(f)\Theta_{12}(f)} \tag{3.9}$$

In this context, $\Theta_{lm}(f)$ denotes the Fourier transform of p autoregressive coefficients ($\Theta_{lm}(f) = \sum_{i=1}^{p} \theta_{lmi} e^{-j2\pi fi}$) from signal l to signal m. Additionally, w_{RR} and w_{RESP} are two independent sources of white noise for the RR and RESP series.

Subsequently, as the proposed framework accounts for the time-varying nature of the system, the estimated $\Theta(t, f)$ will also vary as a function of time. Hence, we can obtain a time-frequency representation of the RR and RESP series ($S_{RR}(t, f)$, $S_{RESP}(t, f)$) and the corresponding cross-spectrum ($S_{CROSS}(t, f)$) using the following expressions.

$$S(t, f) = H^*(t, f) \Sigma(t) H^T(t, f) \tag{3.10}$$

$$S(t, f) = \begin{bmatrix} S_{RR}(t, f) & S_{CROSS}(t, f) \\ S^*_{CROSS}(t, f) & S_{RESP}(t, f) \end{bmatrix} \tag{3.11}$$

$$\Sigma(t) = \begin{bmatrix} \sigma^2_{RR} & 0 \\ 0 & \sigma^2_{RESP} \end{bmatrix} \tag{3.12}$$

$$S(t, f) = \begin{bmatrix} |h_{11}|^2 \sigma^2_{RR} + |h_{12}|^2 \sigma^2_{RESP} & h^*_{11} h_{21} \sigma^2_{RR} + h^*_{12} h_{22} \sigma^2_{RESP} \\ h^*_{21} h_{11} \sigma^2_{RR} + h^*_{22} h_{12} \sigma^2_{RESP} & |h_{21}|^2 \sigma^2_{RR} + |h_{22}|^2 \sigma^2_{RESP} \end{bmatrix} \tag{3.13}$$

Lastly, the directional gains from respiration to heartbeat ($G_{21}(t, f)$) and vice versa ($G_{12}(t, f)$), which are referred to as RSA and Feedforward gains, respectively, can be calculated as:

$$G_{12}(t,f) = G_{RR\to RESP}(t,f) = \sqrt{\frac{S_{RR} - |h_{11}|^2\sigma^2_{RR}}{S_{RESP} - |h_{21}|^2\sigma^2_{RR}}} \tag{3.14}$$

$$G_{21}(t,f) = G_{RESP\to RR}(t,f) = \sqrt{\frac{S_{RESP} - |h_{22}|^2\sigma^2_{RESP}}{S_{RR} - |h_{12}|^2\sigma^2_{RESP}}} \tag{3.15}$$

This modeling approach enables obtaining a high-resolution time-varying estimation of the average RR-interval. Additionally, it provides the spectral representation of both RR and RESP, their cross-spectrum, and the RSA and Feedforward gains.

Moreover the coherence in time and frequency (COH(t,f)) is computed between RR and RESP series as follows:

$$COH(t,f) = |\sqrt{S_{CROSS}(t,f)^2/(S_{RR} * S_{RESP})}| \tag{3.16}$$

Below are the features that were computed through the bivariate point process. For the calculation of features, we have excluded the calculation of feedforward gain ($G_{12}(t,f)$) from RR series to RESP, as opposed to RSA which is a validated index in literature for studying vagal tone. The contribution of RR on respiratory signal does not have a physiological interpretation. The θ_{21} coefficients correct the closed loop estimation mathematically, but there is no corresponding physiological index despite existing correlations between RR series and RESP.

* $\mu(t)_{RESP}$: Average resp series of Gaussian probability distribution.
* $\sigma^2(t)_{RESP}$: resp variance of Gaussian probability distribution.
* $RESP_{HF}(t)$: High frequency power of the continuous resp spectrum.
* $G_{21,HF}(t)$: RSA gain in the high frequency range.
* $COH_{HF}(t)$ The coherence of the two time series computed in the high frequency range.
* $fmax_{HF}(t)$ The RESP frequency computed at the point at maximum coherence in the HF band.

3.4.8 EEG

Signal characterization and acquisition

The EEG technique is widely used for the investigation of brain functionalities. More specifically, EEG is the non-invasive measurement of the brain's electric fields. The measurement of electrical activity generated by the brain is captured by the placement of electrodes on the scalp, which record the voltage potentials that are produced by the flow of current in and around neurons.

The collection of the EEG signal was performed using a DSI 24 headset that was equipped with 19 dry electrodes. These electrodes were positioned at specific locations on the scalp in accordance with the international 10-20 system, which is widely accepted as a standard for EEG electrode

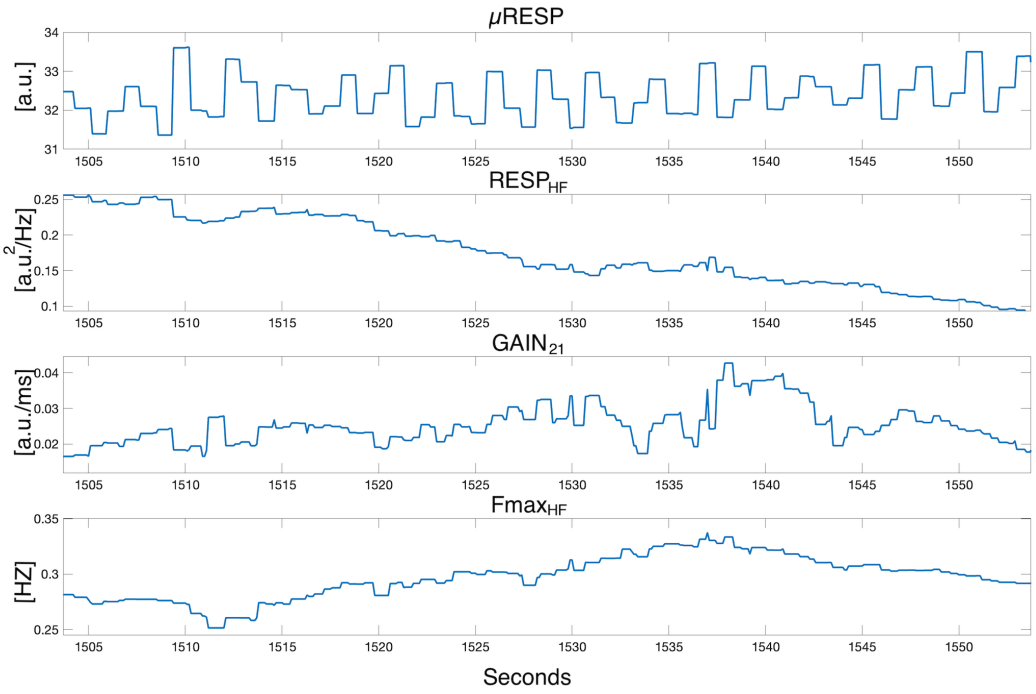

Figure 3.31: Real-time tracking of RESP features, utilizing a bivariate point process modeling approach.

placement. The headset utilized a sampling rate of 300 Hz and was integrated with a 16-bit analog-to-digital (A/D) converter to facilitate the accurate conversion of the analog signals into digital signals for further analysis and processing. Fig 3.32 shows the electrode configuration.

Processing

The EEG signals were imported and preprocessed in a MATLAB environment utilizing the EEGLAB toolbox. As a preliminary step of preprocessing, all EEG data were filtered between 1 Hz and 45 Hz using a finite impulse response, zero-phase filter. The signal recorded by the Pz electrode was found to be faulty, and therefore it was temporarily removed and later interpolated. Then, the Independent Component Analysis (ICA) method with the Extended Infomax algorithm was applied. This method is based on the natural gradient approach as described in [173] where distributions are modeled as sub-Gaussian and super-Gaussian, and separated through the utilization of a simple learning rule, which is grounded in the negentropy as a projection search index [174]. A comprehensive evaluation of the extracted components was carried out through a systematic categorization process into seven distinctive classes, including Brain, Eye, Muscle, Heart, Line Noise, Channel Noise, and Other, utilizing the IClabel plugin [175]. This method aimed to classify the components based on their inherent characteristics and origin. Our study aimed to assess the accuracy of the categorization process by analyzing the probability of the components being classified as relative to brain or other sources. Afterward, the preprocessing phase involved removing components that were deemed as artifacts, determined through the application of default threshold values, resulting in the generation of cleaned datasets that were then subjected to further analysis. The electrode Pz, which had been previously identified as producing a corrupted signal and temporarily removed,

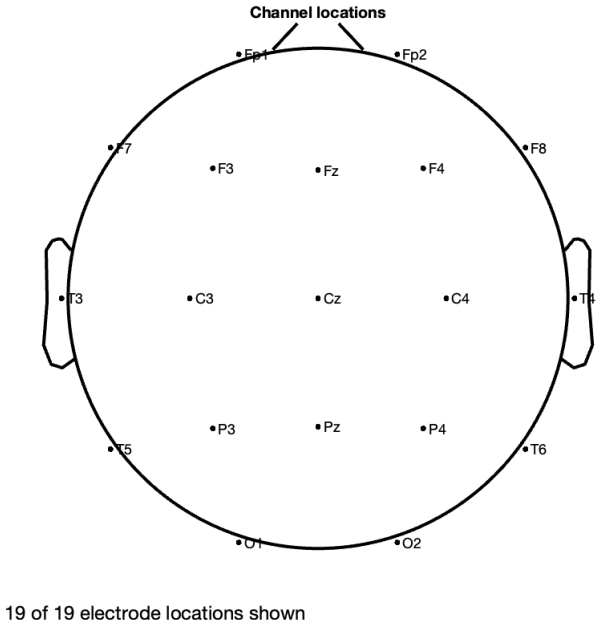

Figure 3.32: Electrode configuration of the DSI 24 headset.

was subsequently subjected to interpolation through the utilization of the spherical interpolation method. Finally, to reduce the common noise present in the recorded signals, the Common-Average Referencing (CAR) method was employed. This approach involves subtracting the average potential of multiple electrodes from each individual electrode, thus reducing the impact of common sources of noise on the recorded signals. The entire pipeline for EEG signal processing was chosen based on comparisons of different pipelines described in [176]. Figure 3.33 displays a screen shot from EEGLAB related to the raw EEG signal in panel (A) and the processed signal in panel (B). As can be observed from the image in panel (A), the Pz electrode was corrupted, displaying a periodic sinusoidal signal that has been interpolated in panel (B).

Features extraction

Regarding the analysis of EEG signals during the different experiments, several features were extracted to gain a deeper understanding of the underlying physiological processes. The Normalized Power Spectral Density with respect to the total power (1-45 Hz) was computed for each of the frontal and parietal regions, in each of the following frequency bands: δ (1-3 Hz), θ (4-7 Hz), α (8-12 Hz), and β (16-38 Hz). This information provides a comprehensive view of the distribution of energy across different frequency ranges, and enables the identification of patterns and trends in the EEG signal that are characteristic of different emotional states.

Furthermore, to assess the level of attention of the subjects during the experiments, the ratio of the Power Spectral Density in the β frequency band over the one in the θ frequency band was also calculated for both the frontal and parietal regions. The β/θ F and β/θ P ratios are widely recognized as indicators of attention [177], and it is believed that the ratio increases during attentive states, providing a more precise measure of the level of attention during each trial. Additionally,

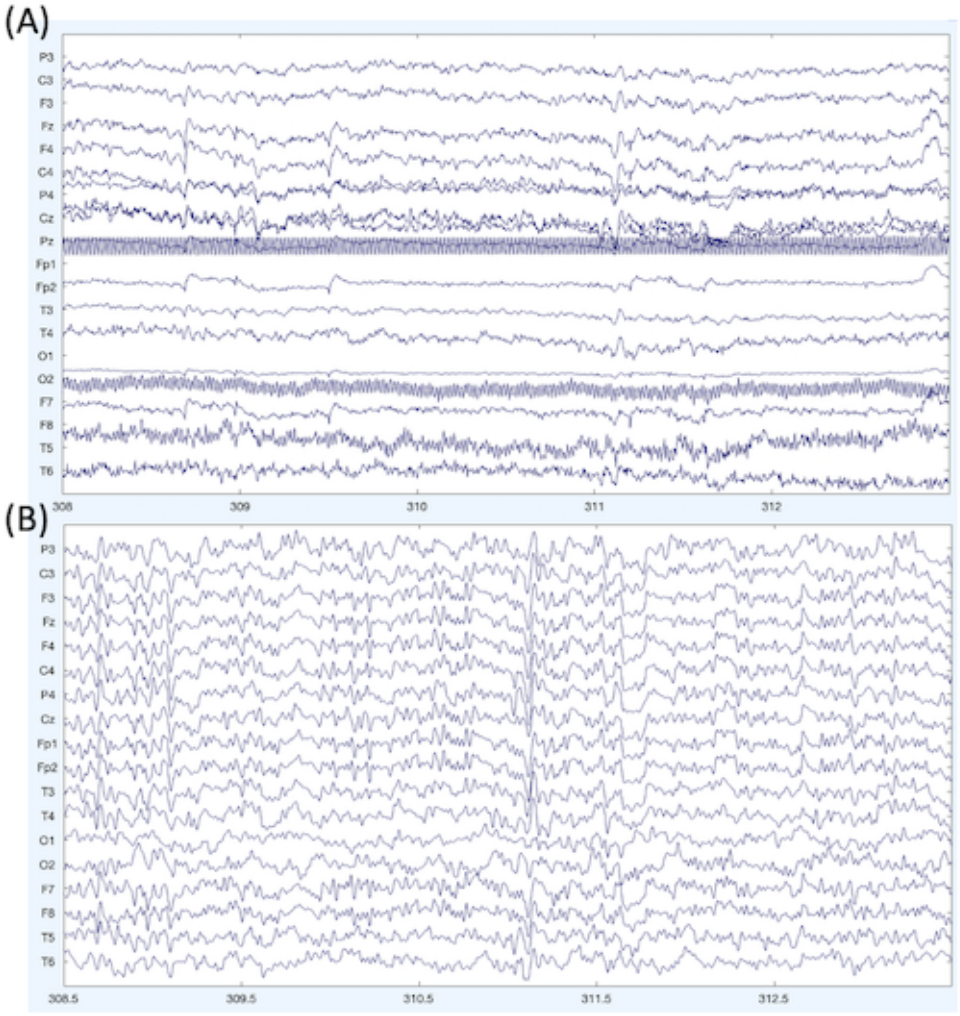

Figure 3.33: Panel (A) shows the raw EEG signal and panel (B) shows the EEG signal after undergoing the entire processing phase.

in the same two regions an index was calculated to represent the level of task engagement being performed. The index β/α is built on the premise that an increase in Beta power is associated with increased brain activity during a mental task, as evidenced by studies linking the Beta frequency band to various cortical contributions, such as activation of the visual system, motion planning activity, and other functions mainly tied to an attentive state of the brain [178]. On the other hand, increases in Alpha activity is related to lower levels of mental vigilance and alertness.

To ensure that the results are robust and accurate, the Power Spectral Densities were calculated using the Welch method, a well-established approach in signal processing that is widely used for the analysis of EEG signals.

Chapter 4

Results

4.1 Listening effort protocol and emotions in music

Listening effort

In the previously discussed section 3.1.2, a thorough examination of the extracted features from ECG and BVP, GSR, RESP, PUPIL and EEG signals was conducted using statistical testing methodologies. The aim of this analysis was to differentiate and compare the features obtained from subjects under two levels of effort (Low=L and High=H) with those recorded during the baseline (B) phase. The results of this analysis were aimed to provide insights into the variations in physiological responses associated with changes in effort levels.

In Table 4.1, we present a comprehensive summary of the median values and the median absolute deviations for all the computed features across all two levels of effort and the baseline. The median absolute deviation, on the other hand, quantifies the variability of the features around the median and provides an indication of the spread of the data.

During this experiment, a one and a half minute buffer was available prior to the actual experiment in order to facilitate the initialization process of both univariate and bivariate point process models. As a result, ECG and RESP point process features were computed with respect to less robust traditional features, using brief time window, where at least 5 minutes are necessary [179].

Starting from the PUPIL-related measures, they do not not seem to significantly distinguish between the two effort phases, although the AVD, in accordance with the literature, shows higher values during high effort compared to low effort. In frequency, it appears that regardless of the frequency ranges, the spectral power is always higher during the baseline compared to the two test phases. In general, regarding the PUPIL signal, the baseline may not be consistent with the two test phases as during the baseline subjects focus on a gray screen while during the two test phases the test screen, including three buttons that change the stimuli shown with each trial, is in front of them.

Regarding the cardiac features calculated with point processes, $\mu(t)_{RR}$ appears relevant as it significantly distinguishes between the two test phases. During high effort, there is indeed an acceleration of the heartbeat, even compared to the initial baseline phase although not significantly. The total spectral power and in the very low frequency range is significantly higher in baseline

compared to the high effort phase. Typically, a decrease in these two features is interpreted in relation to sympathetic activation or parasympathetic deactivation.

Regarding RESP, the breath amplitude is significantly higher during the high-effort phase, linked although not significantly to a higher breathing rate. Therefore, at high effort, subjects tend to take bigger and closer breaths. The other features seem very similar in all three phases, except for GAIN21, which represents an estimate of the vagal tone or RSA, which shows a higher median value during the low-effort phase, as if there was a greater parasympathetic activation in this phase.

The GSR signal appears to be very similar in most features across the three phases, with the exception of the mean value of the signal in the three time windows, which is significantly higher in the low-effort phase compared to the other two phases. This is also evident in the ENV feature, which represents the envelope of the phasic component of the signal, which is distinctly higher in the low-effort phase compared to the high-effort phase, suggesting contrary to the respiratory signal, higher arousal and sweating in the low-effort phase.

As with cardiac features, the BVP signal also seems to show greater sympathetic control, as indicated by significantly lower VA and PAT values in the high-effort phase. VA represents the volume of blood in the periphery, and a lower value indicates an increase in peripheral blood pressure and vasoconstriction. The same is evident in PAT, where during high effort, the time it takes for the pressure wave to travel from the heart to the periphery is reduced, indicating, in the same way as VA, greater sympathetic activation.

As for the EEG signal, the normalized power spectral densities in α, β and θ bands both in parietal and frontal regions exhibit a significant increase during the two effort phases, but only when compared to the baseline. The normalized power spectral density of the more relevant frequency bands (i.e., α, β, θ) are not reported in Table 4.1 and are instead depicted in Figure 4.1. With regards to the computed EEG indexes, both at frontal and parietal levels, the β/θ attention index is significantly higher in the low-effort phase, even if only compared to the baseline. Moreover, the level of engagement (i.e., β/α) is significantly higher in both effort phases when compared to the baseline.

Figure 4.2 shows boxplots related to all the features which showed statistical differences between low and high effort phases.

Figure 4.3 shows the main results obtained in the study. In particular, both 2D and 3D boxplots were created. In particular features from cardiovascular, central and autonomic domains. In order to summarize the most significant results, features related to different systems were selected to create a general physiological picture. Specifically, the VA feature was chosen for the cardiovascular part, the EEG signal attention index for the central part, and Env for the autonomic part.

Table 4.1 The first column displays calculated features, while the second, third, and fourth columns respectively show the median and in parentheses the median absolute deviation of the features in Baseline (B), low effort (L), and high effort (H). The last column shows the pairs for which there is a significant difference.

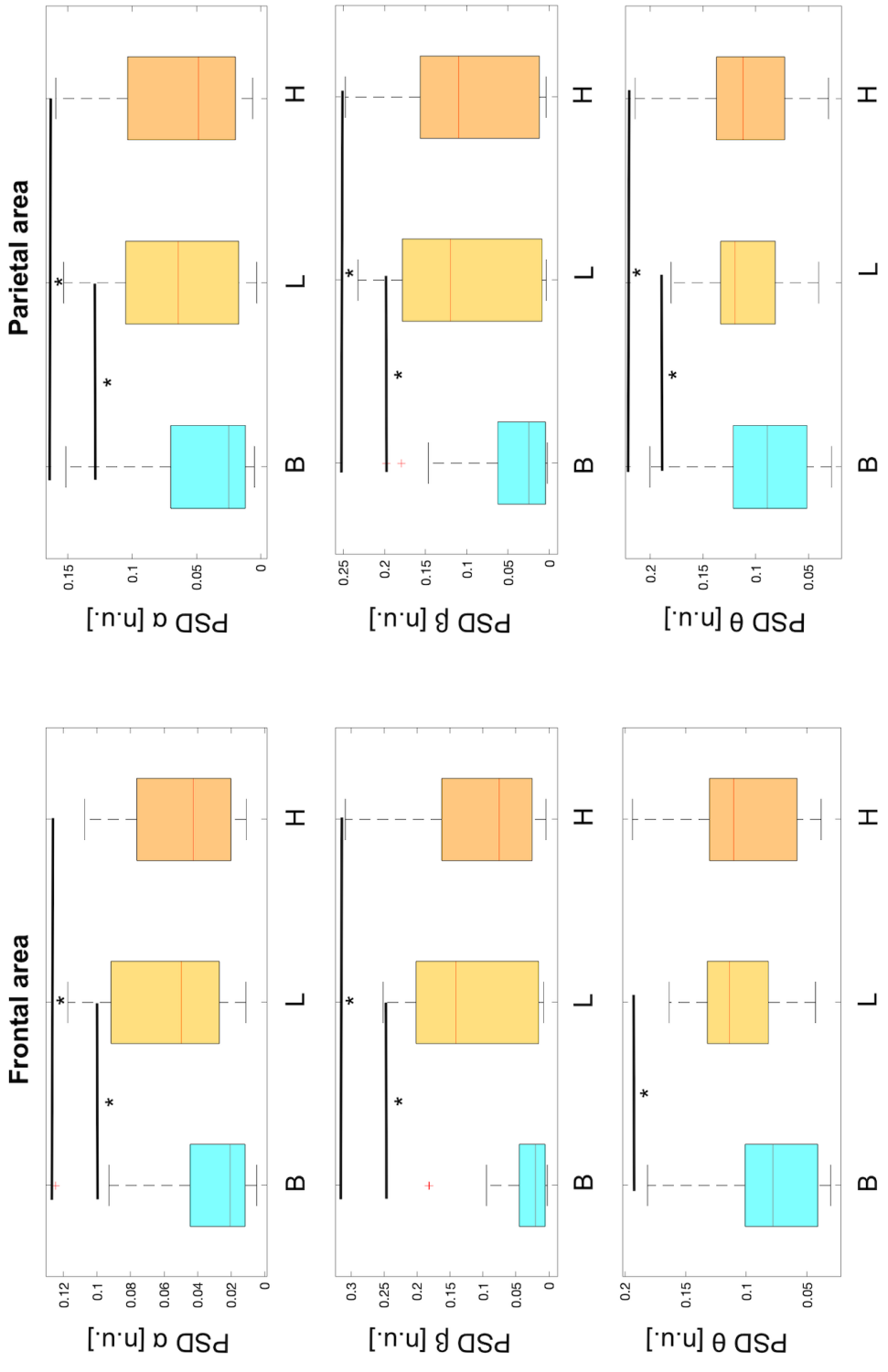

Figure 4.1: The normalized power spectral density of the most significant frequency bands, specifically the α, β, and θ bands, in both the frontal and parietal regions. Statistically significant differences are marked with *.

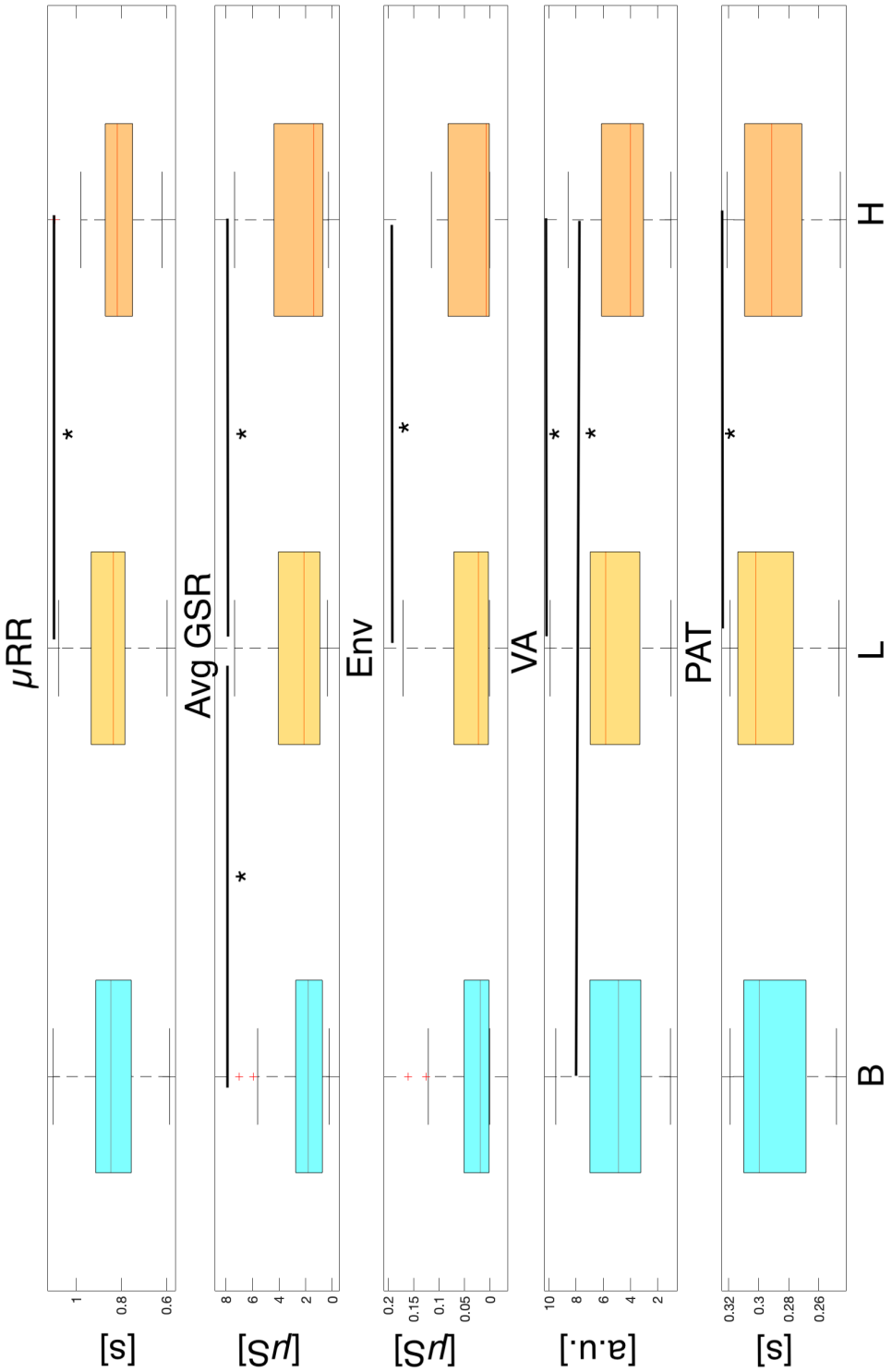

Figure 4.2: The figure reports, all features which show statitical significant differences between low effort (L) and high effort (H). B rapresents the baseline phase. From top to buttom ECG (μ_{RESP}), GSR (AVG GSR and ENV), BVP (VA and PAT). Statistically significant differences are marked with *.

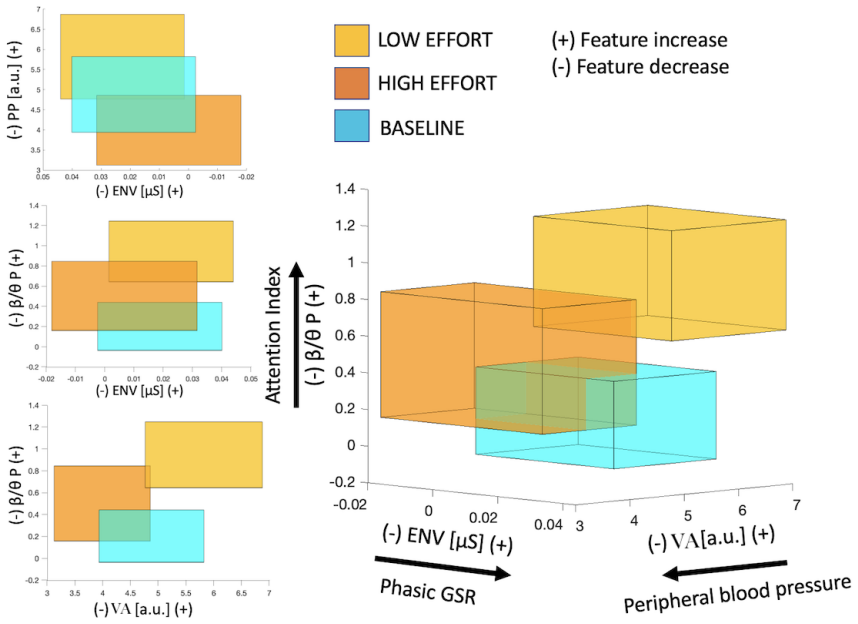

Figure 4.3: 2D and 3D boxplots are built by centering the rectangles around the coordinates of the corresponding median values, where the length of the sides of the rectangles is set equal to median absolute deviation.

Emotions and music (AuBT protocol)

Regarding the AuBT dataset, unfortunately the signals were already cut into two minute segments, approximately the length of songs listened to by the subject. With two minutes per signal, it was impossible to evaluate point process features, as using a 1.5-minute buffer would leave only 30 seconds of stimulation. As for ECG and RESP, traditional features explained in sections 3.4.1 and 3.4.6, respectively, were calculated. Table 4.2 the results of the extracted features are presented and statistically compared for the four emotions. The statistical analysis has shown that the majority of features have proven to be significant in at least three out of the six comparisons made between emotions (joy vs anger, joy vs sadness, joy vs relaxation, anger vs sadness, anger vs pleasure, and sadness vs pleasure). Specifically, AVNN, RRVLF, RESP f, and RESP amp have been found to be the most effective features, showing statistically significant differences in five out of six comparisons. Figure 4.4 displays a graphical representation of the four emotions, including projections and a 3D representation in the plane of the most significant features.

Overall, features associated with the GSR signal can differentiate between the emotional dimension of arousal, distinguishing emotions with high excitement (such as joy and anger) from those with low excitement (such as pleasure and sadness).

At the level of the heart, AVNN exhibits excellent performance and effectively distinguishes between joy and sadness. These two emotions are antithetical to the Russel's circumplex model of affect used to describe emotions in terms of valence and arousal. Joy, characterized by high arousal and positive valence, exhibits the greatest acceleration of heart rate, while sadness, characterized by low arousal and negative valence, exhibits the greatest deceleration.

The respiratory signal appears to be particularly useful for identifying anger, as it exhibits a much higher average rate compared to the other three emotions.

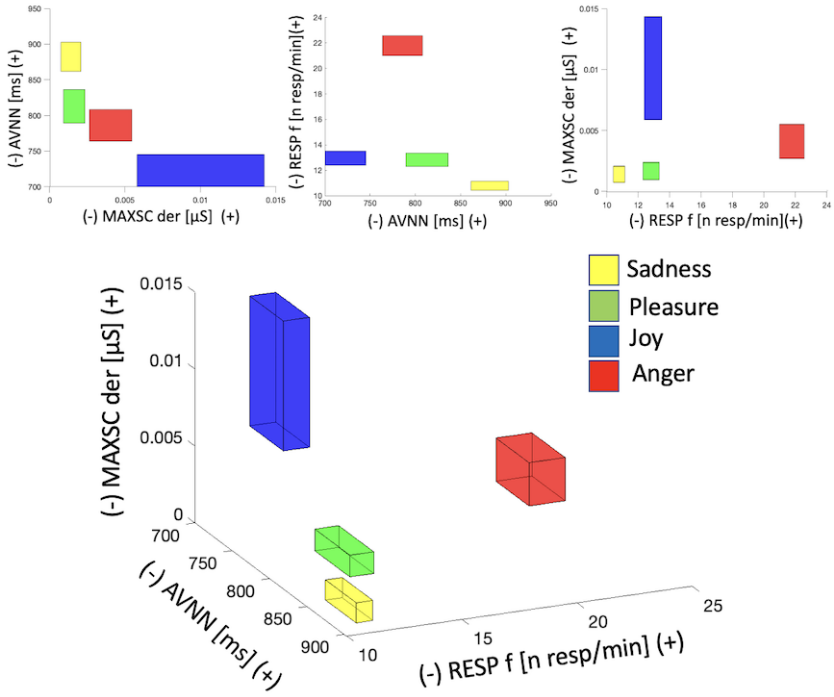

Figure 4.4: The figure displays 3D boxes representing each emotion, with the center of each box representing the mean of three features and the length of the sides representing the 95% confidence limits for the mean estimate. The main figure shows the combination of the three best-performing features in the four emotions' space. Additionally, two-dimensional projections of the 3D boxes are presented.

Classification

As explained in section 3.1.3, a graphical feature selection method was employed to identify the most effective features that could separate the four emotions. The AVNN from ECG, Max der amplitude from GSR, and RESP f from RESP were determined to be the best features, which could separate the four emotions in a 2D plane with minimal correlation. To ensure a more interpretable model, without the application of dimensionality reduction methods, only these features were used to classify the four emotions. As mentioned earlier in section 3.1.3, different models were tested using leave-1-day-out and leave-5-day-out logic, as the dataset consisted of repeated measurements on the same subjects over 21 days. The models were applied both for the 4-class problem, i.e., the four emotions under study, as well as for classifying the two dimensions of the Russell's circumplex model of affect, i.e., high/low arousal and positive/negative valence. The average validation accuracy for classifying the four emotions, as well as arousal and valence, was calculated and reported in Table 4.3.

Table 4.1: The caption of the Table can be found at the end of page 77.

PUPIL	**B**	**LOW**	**HIGH**	*
AVD [mm]	2.9 (0.36)	2.99 (0.299)	3.04 (0.26)	–
SDD [mm]	0.22 (0.10)	0.17 (0.03)	0.17 (0.07)	BL-BH
DLF [mm^2]	0.92 (3.33)	0.41 (0.23)	0.28 (0.1)	BL-BH
DHF [mm^2]	0.78 (1.25)	0.55 (0.37)	0.50 (0.38)	–
DVHF [mm^2]	0.49 (0.50)	0.28 (0.18)	0.29 (0.34)	BL
DLFtoHF	0.98 (1.20)	0.63 (0.48)	0.58 (0.29)	BL-BH
ECG				
$\mu(t)_{RR}$ [s]	0.85 (0.09)	0.84 (0.09)	0.82 (0.09)	L-H
σ^2_{RR} [s^2]*10^-3	0.88 (0.79)	0.80 (0.63)	0.62 (0.61)	–
RRTOT [s^2]*10^-2	46 (63)	36 (19)	21 (17)	BH
RRVLF [s^2]*10^-3	1.3 (5.6)	1.5 (1.4)	0.7(0.9)	BH
RRLF [s^2]*10^-3	0.9 (1.4)	0.6 (1)	0.8 (0.8)	–
RRHF [s^2]*10^-3	0.41 (0.60)	0.42 (0.32)	0.31 (0.37)	–
RRLFn	0.63 (0.15)	0.57 (0.12)	0.54 (0.17)	–
RRHFn	0.37 (0.15)	0.43 (0.12)	0.34 (0.13)	–
RRLFtoHF	2.45 (3.31)	1.71 (1.621)	2.67 (2.06)	–
RSA				
μ_{RESP} [a.u.]	31.07 (3.25)	31.30 (3.39)	31.39 (3.54)	BH-LH
$GAIN21_{HF}$ [a.u./ms]	1.85 (2.60)	4.25 (3.03)	2.78 (2.03)	–
$RESP_{HF}$ [$a.u.^2$]*10^-3	0.4 (1.7)	0.5 (12)	0.5 (0.9)	–
COH_{HF}	0.55 (0.11)	0.60 (0.11)	0.68 (0.10)	–
$fmax_{HF}$ [Hz]	0.31 (0.05)	0.30 (0.04)	0.32 (0.04)	–
GSR				
Avg amplitude peaks [μS]*10^-2	1.73 (4.85)	2.51 (5.41)	0.36 (5.69)	–
Avg rise time [s]	0.82 (0.57)	0.82 (0.44)	0.63 (0.59)	–
Avg recovery time [s]	2.38 (3.13)	3.17 (5.27)	1.95 (2.43)	–
N peaks	2 (2.60)	2 (3.30)	1 (2.77)	–
Max sign amplitude [μS]	0.05 (0.10)	0.01 (0.10)	0.01 (0.12)	–
Avg der amplitude [μS]*10^-6	-0.26 (27.64)	-1.66 (22.98)	-0.64 (21.24)	–
Sd der amplitude [μS]*10^-2	0.14 (0.21)	0.14 (0.23)	0.03 (0.25)	–
Max der amplitude [μS]*10^-2	0.64 (0.76)	(0.56)	0.69 (0.17)	–
Slope GSR [μS/s]*10^-7	2.79 (73.91)	0.60 (54.47)	0.61 (80.84)	–
Avg GSR [μS]	1.81 (2.13)	2.12 (2.22)	1.40 (2.32)	BL-LH
Avg abs1 [μS]*10^-2	0.18 (0.32)	0.26 (0.34)	0.05 0.40	B-L
Avg abs1 norm *10^-2	1.98 (1.65)	1.44 (2.16)	1.77 (1.30)	–
Avg abs2 [μS]*10^-3	0.23 (0.57)	0.22 (0.74)	0.06 (0.69)	–
Avg abs2 norm *10^-2	0.41 (0.53)	0.18 (0.48)	0.47 (0.28)	–
Env [μS]*10^-2	1.88 (3.78)	2.27 (4.25)	0.67 (4.97)	LH
BVP				
VA [a.u.]	4.88 (1.89)	5.82 (2.11)	3.99 (1.73)	BH-LH
PAT [s]*e-2	29.96 (1.93)	30.21 (1.95)	29.13 (1.93)	LH
EEG				
β/θ F	0.11 (0.74)	1.01 (0.81)	0.97 (1.51)	BL
β/θ P	0.20 (0.48)	0.94 (0.60)	0.50 (0.69)	BL-BH
β/α F	0.59 (1.79)	1.75 (1.24)	1.56 (2.70)	BL-BH
β/α P	0.56 (0.88)	1.56 (0.56)	1.41 (0.86)	BL-BH

Table 4.2: The first column displays calculated features, while the second, third, fourth and fifth columns respectively show the median and in parentheses the median absolute deviation of the features in Joy, Anger, Sadness and Pleasure. For the sake of clarity, only features that showed statistically significant differences are shown. The last column shows the pairs in which there is a significant difference.

	Joy	**Anger**	**Sadness**	**Pleasure**	*
			ECG		
AVNN [ms]	713.11 (40.24)	799.78 (41.51)	882.26 (37.88)	828.40 (48.78)	Jvs.all,A-S,S-P
SD1	20.02 (4.32)	22.52 (7.03)	23.78 (7.98)	15.78 (4.85)	A-P,S-P
SD2	65.08 (11.35)	56.09 (12.02)	57.29 (12.45)	45.86 (13.39)	J-A,J-P
SD ratio	0.30 (0.07)	0.42 (0.14)	0.42 (0.07)	0.34 (0.09)	J-A,J-S,A-P
RRVLF $[s^2/Hz]$*10^-5	1.21 (0.20)	1.61 (0.25)	2.19 (0.27)	1.82 (0.31)	Jvs.all,S-P
RRLF $[s^2/Hz]$*10^-4	0.94 (0.11)	1.01 (0.11)	1.33 (0.14)	1.10 (0.12)	J-S,J-P,A-S,S-P
RRHF $[s^2/Hz]$*10^-3	2.54 (0.50)	2.67 (0.39)	3.81 (0.75)	2.90 (0.36)	J-S,A-S,S-P
RRLFtoHF	3.50 (0.40)	3.85 (0.28)	3.49 (0.44)	3.71 (0.25)	A-S
RRLFnu *e-2	23.38 (1.5)	21.98 (1.40)	23.72 (1.33)	23.06 (1.25)	J-A,A-S
RRHFnu *e-2	6.68 (0.81)	5.69 (0.49)	6.77 (0.92)	6.10 (0.41)	J-A,A-S,S-P
			GSR		
Avg amplitude peaks [μS]*10^-2 23.73 (32.08)	8.42 (14.88)	3.01 (5.25)	4.06 (4.32)	J-S,J-P,A-S	
Sd amplitude peaks [μS]*10^-2	28.82 (35.61)	7.09 (17.59)	2.14 (6.74)	3.17 (5.40)	J-S,J-P,A-S
Avg rise time [s]	0.88 (0.18)	0.73 (0.23)	0.61 (0.37)	0.68 (0.26)	J-S
Avg recovery time [s]	1.93 (0.51)	2.50 (1.78)	1.79 (1.02)	2.00 (0.72)	A-S,A-P
N peaks	13.00 (4.82)	6.00 (2.54)	2.00 (2.33)	4.00 (2.36)	Jvs.all,A-S
Max sign amplitude [μS]*10^-2	58.59 (108.07)	22.12 (45.82)	7.63 (17.88)	10.19 (18.97)	Jvs.all,A-S
Sd der amplitude [μS]*10^-2	1.59 (2.15)	0.52 (0.54)	0.22 (0.18)	0.22 (0.25)	Jvs.all,A-S
Max der amplitude [μS]*10^-2	7.12 (8.01)	2.82 (2.73)	0.94 (1.16)	1.08 (1.16)	Jvs.all,A-S
Avg GSR [μS]	2.05 (0.74)	1.86 (0.83)	1.43 (0.43)	1.48 (0.43)	J-S,J-P,A-S
Avg abs1 [μS]*10^-2	1.69 (2.38)	0.86 (0.62)	0.39 (0.24)	0.49 (0.37)	J-S,J-P,A-S
Avg abs1 norm *10^-2	20.90 (8.51)	11.72 (6.84)	7.22 (4.76)	7.27 (5.76)	Jvs.all
Avg abs2 [μS]*10^-3	4.78 (4.75)	2.94 (0.92)	1.71 (0.66)	1.88 (0.80)	Jvs.all
Avg abs2 norm *10^-2	5.33 (2.60)	3.93 (2.97)	2.68 (2.31)	2.77 1.95	J-S,J-P
Env [μS]*10^-2	24.44 (36.59)	8.76 (9.38)	3.44 (2.34)	3.53 (4.39)	J-S,J-P,A-S
			RESP		
fRESP [breaths/min]	12.50 (1.25)	22.00 (1.71)	11.00 (0.69)	13.00 (1.03)	J-A,J-S,A-S,A-P,S-P
ampRESP [a.u.]	0.83 (0.30)	0.50 (0.18)	0.72 (0.23)	0.73 (0.26)	Jvs.all,A-S,A-P

Table 4.3: Accuracy as mean (s.d.) obtained for four emotions, arousal and valence classification using KNN, LDA, SVM and DT.

KNN	**4 emotions**	**Arousal**	**Valence**
leave-5days-out	85 (0.07)	90 (0.03)	92 (0.09)
leave-1day-out	85 (0.17)	92 (0.11)	93 (0.13)
leave-one-out	85 (0.30)	92 (0.27)	93 (0.25)
LDA	**4 emotions**	**Arousal**	**Valence**
leave-5days-out	84 (0.07)	89 (0.04)	77 (0.07)
leave-1day-out	83 (0.20)	91 (0.12)	77 (0.17)
leave-one-out	83 (0.37)	89 (0.31)	75 (0.43)
SVM	**4 emotions**	**Arousal**	**Valence**
leave-5days-out	84 (0.02)	88 (0.06)	73 (0.07)
leave-1day-out	80 (0.19)	90 (0.12)	76 (0.19)
leave-one-out	80 (0.40)	88 (0.32)	75 (0.43)
DT	**4 emotions**	**Arousal**	**Valence**
leave-5days-out	80 (0.07)	87 (0.04)	87 (0.06)
leave-1day-out	78 (0.19)	86 (0.16)	90 (0.16)
leave-one-out	78 (0.41)	88 (0.32)	91 (0.29)

4.2 Emotional protocol with visual, auditory, and combined stimuli

Table 4.4 and 4.5 show a summary of important features that were found when comparing three different phases of stimulation for both arousal and valence levels. The results in the two tables are divided into emotional dimension and signal type.

GSR: Table 4.4 show that the physiological response to different types of stimuli varies. The GSR response is generally higher during the IADS-only phase compared to the IAPS-only phase. High valence stimuli elicit a higher GSR response, with higher GSR Amp peaks and AV env, GSR n peaks, and AVGSR rise time. Conversely, during the image viewing phase, GSR activity is generally lower.

ECG: Table 4.4 shows that μRR is generally higher in the IADS-only phase compared to the other two phases, with higher median values in all arousal sessions. This suggests a slowing down of the heartbeat when subjects listen to sounds alone, indicating lower sympathetic activity. In A4, RR LFn shows significantly lower values in the IADS-only phase compared to the IAPS+IADS phase, while RR HFn shows the opposite trend. In the same arousal session, RR LF/HF is also significantly lower in the IADS-only phase compared to the other two phases (see Figure 4.6). No statistically significant differences were observed in the valence dimension.

BVP: Table 4.4 shows that the computed PAT is significantly lower in the IADS-only phase compared to the other two phases, with a trend of lower VA values in the IADS-only phase. The blood propagation time from the heart to the periphery narrows, and the volume of blood decreases during the beats related to the IADS-only phase. In Table 4.5, significantly lower PAT and VA values are observed in the IADS-only phase during high valence stimuli in all arousal sessions compared to the other two phases. In general, lower values are observed during the IADS-only phase, and this trend seems to be attributed to high valence stimulation.

EEG: During EEG arousal sessions, higher average activation in the δ frequency range was observed in frontal and central regions, with more pronounced activation in the IADS-only phase (see Figure 4.5). Although no significant differences were found in power spectral density, attention levels were consistently higher in the IADS-only phase in the frontal and parietal regions, except for the fourth session (see Figure 4.7). In terms of valence, low and high valence stimulation resulted in significantly higher attention levels in the frontal region during the IADS-only phase, and in the parietal region during the first two arousal sessions. Higher attention levels were also observed in the IADS-only phase compared to the other two phases in both regions during high valence stimulation in A1-A3.

PUPIL: During arousal sessions, the average pupil diameter was significantly lower in the IADS-only phase than in the other two phases, but there were no significant differences in pupil size between the IAPS-only and IAPS+IADS phases. The standard deviation of pupillary diameter was also lower in the IADS-only phase in A3. DHF was significantly higher during the IAPS-only phase in A3, and DLF was higher during the IAPS-only phase in both A3 and A4. However, there were no significant differences in normalized power spectral density or balance index between the three phases in A3. During the IADS-only phase in A3, increases were observed in DLFn and DLF/DHF, while decreases were observed in DHFn. For low valence stimuli, the DLF band was significantly higher during the IAPS-only phase compared to both other phases in A4, and the standard deviation of pupillary diameter was higher in the IAPS-only phase compared to the other two phases in A4. Additionally, the very high frequency band was higher in the IAPS+IADS phase compared to the IAPS-only phase in A4 (see Figure 4.6).

RESP: Regarding the respiratory signal, the power in the HF band is found to be higher in the IAPS-only phase compared to the other two phases in A3 (see Figure 4.6). Although the median is higher in the IAPS-only phase, we can see from Figure 4.6 that there is a decreasing trend from IAPS-only, IADS-only, and IAPS+IADS. It seems, therefore, that the parasympathetic system is more active in the phase of only images. In terms of valence, $GAIN21_{HF}$, which represents an estimate of RSA, is higher in high valence in IAPS+IADS compared to IADS-only, with a trend of lower values in general in the IADS-only phase, further supporting the notion that the parasympathetic system is less active in the phase of only sounds, as RSA is an index of vagal activation. The higher power in the HF frequency band in IAPS, as seen from Table 4.5, is relative to the low valence part. At low valence, the frequency at maximum coherence is significantly lower in IAPS+IADS compared to the other two phases, indicating a slowed respiratory rate in this phase.

Figure 4.8 shows the main results obtained in the study. In particular, both 2D and 3D box-plots were created by joining all feature values of the four arousal sessions of each phase (mean $\pm$ standard error).

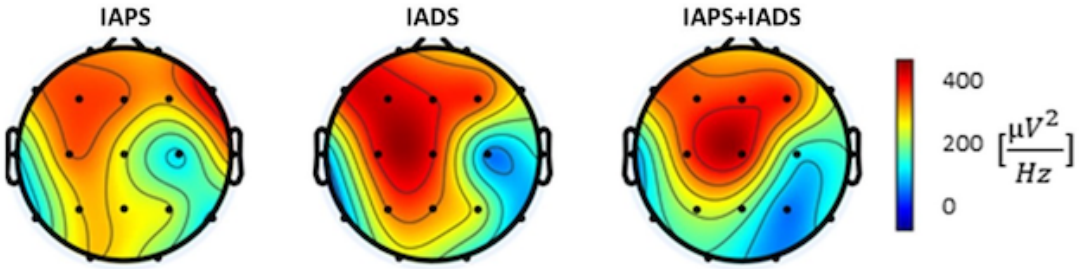

Figure 4.5: Activation on the scalp in terms of Power Spectral Density in δ band for each type of stimulation mode averaged among subjects.

Table 4.4: Median and ranges of all the features computed in low (L) and high (H) valence sessions for the three stimuli. For the sake of clarity, only features that showed statistically significant differences are shown except for μRR. Statistically significant differences are shown in bold and the last column specifies between which phases these differences are observed (1: IAPS-only, 2: IADS-only, 3: IAPS+IADS).

AROUSAL	IAPS-only				IADS-only				IAPS+IADS				*
	A1	A2	A3	A4	A1	A2	A3	A4	A1	A2	A3	A4	
ECG													
μRR [ms]	812 (131)	817 (154)	847 (189)	835 (128)	862 (198)	847(200)	858 (187)	856 (154)	837 (146)	846 (130)	848 (162)	841 (139)	–
RR LFn [ms^2/Hz]	0.78(0.29)	0.69(0.26)	0.74(0.26)	0.71(0.38)	0.64(0.31)	0.74(0.23)	0.73(0.23)	**0.65(0.34)**	0.74(0.20)	0.68(0.20)	0.69(0.24)	**0.71(0.20)**	2vs.3(A4)
RR HFn [ms^2/Hz]	0.22(0.29)	0.31(0.26)	0.26(0.26)	0.28(0.38)	0.36(0.31)	0.26(0.23)	0.27(0.23)	**0.35(0.34)**	026(0.20)	0.32(0.21)	0.31(0.24)	**0.29(0.20)**	2vs.3(A4)
RR LF/HF	3.81(4.49)	2.49(1.99)	2.65(4.17)	**3.22(4.49)**	2.02(4.01)	3(3.1)	2.99(3.993)	**2.03(2.76)**	3.02(3.89)	2.21(1.98)	2.80(3.59)	**2.78(2.93)**	2vs.1,3(A4)
BVP													
VA [a.u.]	6.87(4.35)	5.93(4.87)	6.72(4.97)	5.38(4.44)	5.02(3.96)	4.78(5.43)	4.62(4.52)	5.64(3.99)	6.19(3.66)	5.66(4.39)	5.89(4.27)	5.33(3.48)	–
PAT [ms]	292(43)	292(43)	**299(44)**	300(42)	296(41)	**295(41)**	**296(46)**	**297(46)**	296(46)	**296(44)**	296(45)	**298(45)**	2vs.1(A3)/2vs.3(A2,A4)
GSR													
GSR Amp peaks [nS]	**0(6.9)**	0(12)	**0(24)**	1.4(22)	**20(49)**	5.7(29)	**4.2(44)**	3.2(33)	2.7(35)	0(60)	1.1(10)	2.5(53)	1vs.2(A1,A3)
GSR n peaks	**0(1.37)**	0(2)	0.5(3.12)	1(2.5)	**0.5(3.25)**	1(3)	0.5(3.12)	0.5(2.62)	**0.5(3)**	0(3)	0.5(2.75)	0.5(3.25)	2vs.1,3(A1)
AV env [nS]	**3.4(10)**	**2.4(16)**	5(23)	4.4(30)	**3.8(43)**	**5.6(32)**	3.3(37)	5.3(27)	4.4(22)	3.1(47)	7.2(20)	17(63)	1vs.2(A1,A2)
PUPIL													
Std diameter [mm]	0.22 (0.11)	0.24 (0.10)	**0.26 (0.12)**	0.28 (0.10)	0.22 (0.13)	0.24 (0.11)	**0.23 (0.10)**	0.26 (0.13)	0.23 (0.11)	0.26 (0.10)	**0.25 (0.11)**	0.26 (0.14)	2vs.1,3(A3)
DVHF [mm^2/Hz]	0.26 (0.27)	0.31 (0.31)	0.31 (0.35)	**0.36 (0.37)**	0.26 (0.49)	0.38 (0.35)	0.41 (0.37)	0.31 (0.34)	0.32 (0.39)	0.28 (0.37)	0.40 (0.46)	**0.41 (0.42)**	1vs.3(A4)
EEG													
β/θ frontal	**0.18 (0.10)**	**0.18 (0.15)**	**0.18 (0.20)**	0.24 (0.11)	**0.42 (0.17)**	**0.38 (0.21)**	**0.47 (0.16)**	0.26 (0.08)	**0.16 (0.11)**	**0.18 (0.13)**	**0.17 (0.13)**	0.19 (0.10)	2vs.1,3(A1,A2,A3)
β/θ parietal	**0.30 (0.20)**	**0.33 (0.23)**	**0.29 (0.20)**	0.35 (0.04)	**0.52 (0.20)**	**0.47 (0.18)**	**0.46 (0.22)**	0.33 (0.13)	**0.30 (0.11)**	**0.32 (0.08)**	**0.31 (0.10)**	0.35 (0.14)	2vs.1,3(A1,A2,A3)
RESP													
$RESP_{HF}$ [$a.u.^2$]	0.20 (0.85)	0.26 (0.59)	**0.23 (0.93)**	0.28 (0.98)	0.16 (0.58)	0.17 (0.33)	**0.29 (0.84)**	0.18 (0.94)	0.15 (0.41)	0.25 (0.43)	**0.11 (0.40)**	0.15 (0.57)	1vs.2,3(A3)

Table 4.5: Median and ranges of all the features computed in low (L) and high (H) valence sessions for the three stimuli. For the sake of clarity, only features that showed statistically significant differences are shown. Statistically significant differences are shown in bold and the last column specifies between which phases these differences are observed (1: IAPS-only, 2: IADS-only, 3: IAPS+IADS).

VALENCE	IAPS-only				IADS-only				IAPS+IADS				*
	A1	A2	A3	A4	A1	A2	A3	A4	A1	A2	A3	A4	
BVP													
L: VA [a.u.]	6.92 (4.76)	7.11 (4.50)	6.79 (5.20)	5.31 (3.95)	**4.89 (4.02)**	4.38 (5.77)	4.61 (4.80)	5.54 (4.12)	**6.60 (3.76)**	6.45 (4.42)	6.58 (4.81)	6.11 (3.89)	2vs.3(A1)
H: PAT [ms]	293 (40)	293 (39)	**300 (45)**	299 (42)	296 (41)	**298 (40)**	**297 (47)**	**296 (45)**	297 (43)	**295 (46)**	296 (43)	**298 (47)**	1vs2(A3)/2vs.3(A2,A4)
GSR													
H: GSR Amp peaks [nS]	**0 (1)**	**0 (8)**	0 (20)	0 (22)	**27 (54)**	**4 (34)**	0 (45)	0 (36)	0 (14)	0 (52)	0 (14)	5 (61)	1vs.2(A1,A2)
H: GSR n peaks	**0 (1.25)**	0 (1.25)	0 (2.25)	1 (2)	**1 (2.50)**	1 (3)	**0 (3.25)**	0 (3)	0 (2.5)	0 (3)	**0 (2)**	1 (3.25)	1vs.2(A1),2vs3(A3)
H: AVenv [nS]	**3 (9)**	2 (13)	3 (22)	6 (19)	**6 (36)**	4 (37)	3 (44)	4 (24)	5 (24)	3 (46)	7 (26)	22 (68)	1vs.2(A1)
H: GSR rise Time [ms]	**0 (271)**	0 (644)	0 (944)	375 (881)	**850 (1125)**	875 (1311)	0 (1039)	0 (1063)	0 (750)	0 (1025)	0 (960)	750 (961)	1vs.2(A1)
PUPIL													
L: DLF [mm^2/Hz]	0.28 (0.23)	0.33 (0.51)	0.38 (0.63)	**0.45 (0.57)**	0.32 (0.42)	0.42 (0.43)	0.54 (0.54)	**0.27 (0.73)**	0.26 (0.67)	0.28 (0.48)	0.30 (0.62)	**0.30 (0.50)**	2vs1,3(A4)
L: SDD [mm]	0.22 (0.11)	0.24 (0.10)	0.26 (0.12)	**0.28 (0.10)**	0.23 (0.13)	0.24 (0.10)	0.23 (0.10)	**0.26 (0.13)**	0.23 (0.11)	0.26 (0.10)	0.25 (0.11)	**0.26 (0.14)**	2vs1,3(A4)
H: DLF [mm^2/Hz]	0.28 (0.23)	0.33 (0.51)	**0.38 (0.63)**	0.45 (0.57)	0.32 (0.42)	0.42 (0.43)	0.54 (0.54)	0.27 (0.73)	0.26 (0.67)	0.28 (0.48)	**0.30 (0.62)**	0.30 (0.50)	1vs.3(A3)
H: DHF [mm^2/Hz]	0.34 (0.22)	0.33 (0.32)	**0.41 (0.56)**	0.44 (0.44)	0.35 (0.49)	0.44 (0.35)	**0.34 (0.42)**	0.26 (0.43)	0.32 (0.44)	0.39 (0.36)	0.40 (0.25)	0.47 (0.42)	1vs.2(A3)
EEG													
H:β/θ frontal	**0.20 (0.14)**	**0.21 (0.14)**	**0.16 (0.19)**	0.21 (0.11)	**0.35 (0.17)**	**0.36 (0.23)**	**0.34 (0.21)**	0.19 (0.15)	**0.16 (0.11)**	**0.19 (0.10)**	**0.17 (0.17)**	0.18 (0.05)	2vs.1,3(A1,A2,A3)
H:β/θ parietal	0.30 (0.24)	**0.33 (0.13)**	**0.29 (0.37)**	0.32 (0.10)	**0.48 (0.16)**	**0.43 (0.20)**	**0.50 (0.21)**	0.32 (0.13)	**0.27 (0.10)**	**0.32 (0.09)**	**0.33 (0.09)**	0.32 (0.08)	2vs.3(A1,A2)/2vs.1,3(A3)
L:β/θ F	**0.19 (0.23)**	**0.17 (0.17)**	0.17 (0.23)	0.26 (0.14)	**0.41 (0.22)**	**0.35 (0.20)**	0.42 (0.35)	0.27 (0.24)	**0.20 (0.12)**	**0.16 (0.14)**	0.15 (0.06)	0.18 (0.11)	2vs.1,3(A1,A2)
L:β/θ P	**0.32 (0.16)**	0.33 (0.18)	0.31 (0.21)	0.34 (0.05)	**0.53 (0.21)**	**0.42 (0.25)**	**0.40 (0.33)**	0.36 (0.16)	**0.33 (0.16)**	**0.31 (0.08)**	**0.28 (0.07)**	0.36 (0.08)	2vs.1,3 (A1)/2vs.3(A2,A3)
RESP													
H: $GAIN21_{HF}$ [a.u./ms] *10^{-2}	3.59 (4.39)	4.25 (3.71)	3.68 (2.57)	4.41 (3.84)	**2.76 (1.19)**	3.78 (4.98)	2.51 (3.62)	3.60 (7.23)	**3.86 (3.98)**	5.65 (5.68)	3.75 (3.18)	3.77 (2.40)	2vs.3 (A1)
L: $RESP_{HF}$ [$a.u.^2$]	0.24 (0.69)	0.26 (0.83)	**0.24 (1.05)**	0.27 (0.50)	0.18 (0.53)	0.11 (0.36)	**0.24 (0.58)**	0.11 (0.43)	0.12 (0.34)	0.21 (0.41)	**0.10 (0.48)**	0.18 (0.40)	1vs.2,3(A3)
L: $fmax_{HF}$ [Hz]	0.32 (0.10)	0.35 (0.08)	**0.34 (0.09)**	0.31 (0.11)	0.30 (0.08)	0.31 (0.11)	**0.35 (0.09)**	0.29 (0.08)	0.31 (0.10)	0.36 (0.15)	**0.32 (0.06)**	0.32 (0.10)	1,2vs.3(A3)

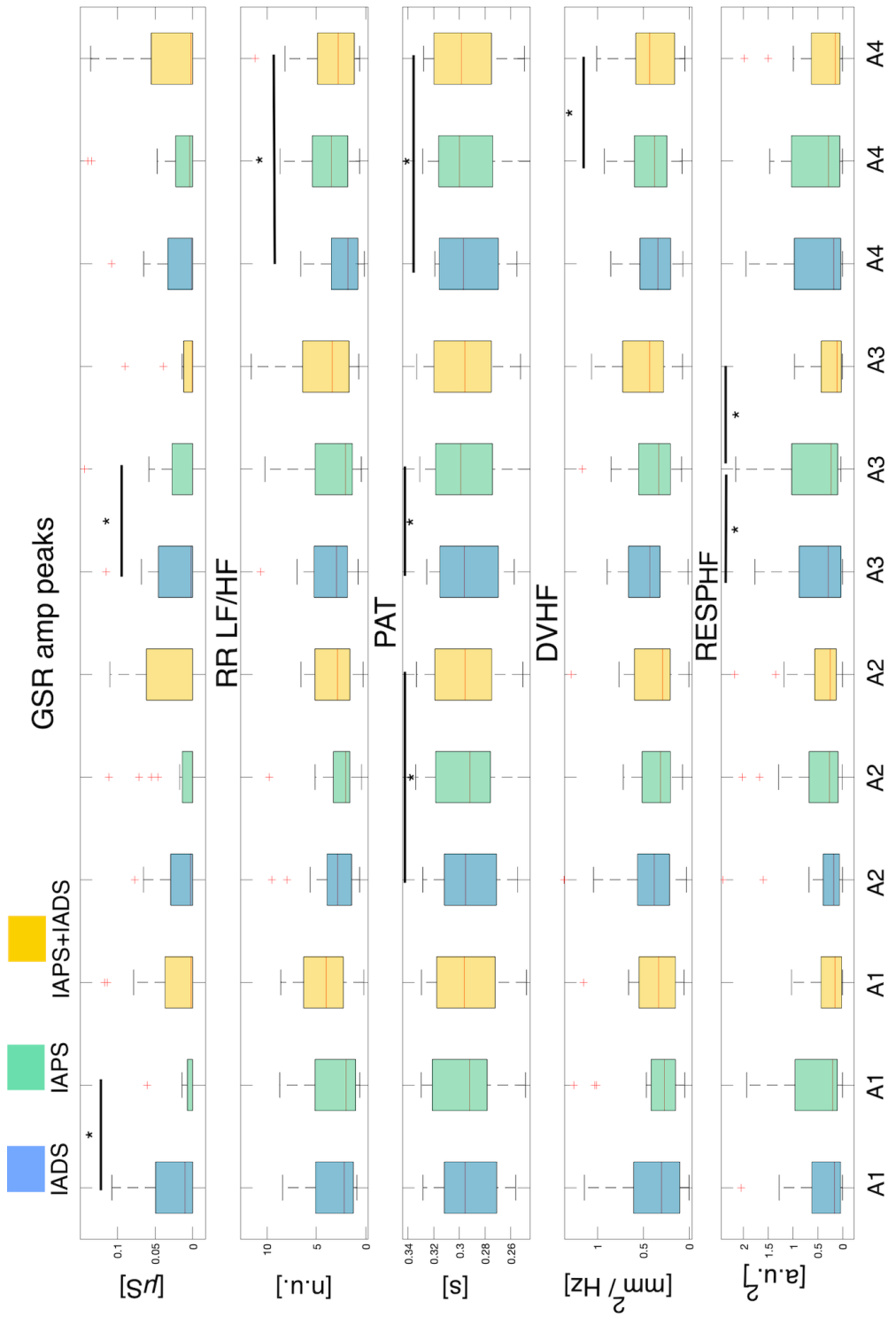

Figure 4.6: The figure reports, for each signal linked to autonomic activity, one of the statistically significant feature found in the arousal comparison among the three phases. From top to buttom GSR (GSR Amp peaks), ECG (RR LF/HF), BVP (PAT), PUPIL (DVHF) and RESP ($RESP_{HF}$). Statistically significant differences are marked with *

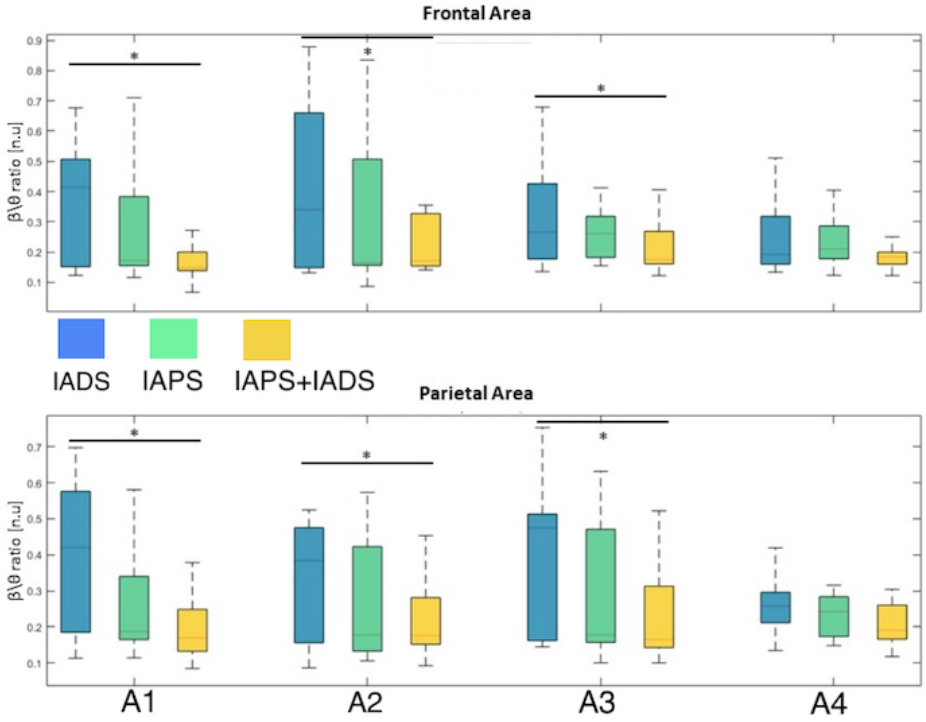

Figure 4.7: Attention index β/θ found significant in both parietal and frontal areas. Statistically significant differences are marked with *

Classification

Table 4.6 displays the best results obtained for the four-class classification, referring to the four quadrants of Russel's circumplex model, considering only the first and last arousal sessions for all stimuli together, as well as for the three individual phases IAPS-only, IADS-only, and IAPS+IADS.

The top-performing model on the entire dataset, considering all three methods of stimulation and in terms of accuracy, was the random forest, with a test accuracy of 52% (twice the performance of a random choice classification method) and an average validation accuracy of 73% for the four-class problem. For the four-class problems applied to each emotion stimulation phase, the k nearest neighbor model achieved the best performance for visual stimulation, with a training accuracy of 39% and an average validation accuracy of 35%. On the other hand, for auditory stimulation, the best model was AdaBoost, with a training accuracy of 55% and an average validation accuracy of 43%. Finally, for the combined (visual+auditory) stimulation, the top-performing model was once again AdaBoost, with a training accuracy of 56% and an average validation accuracy of 44%. The most accurate results obtained by combining the three different stimulation methods can likely be attributed to a physiological agreement among the same emotions elicited by different stimuli, but primarily due to the larger amount of data available for training the machine learning models.

4.3 Virtual reality elicitation protocol

The results of the PAM model are presented in Figure 4.9, based on the surveys given to the participants following the auditory and visual stimuli for emotion labeling. It is apparent that the scenes

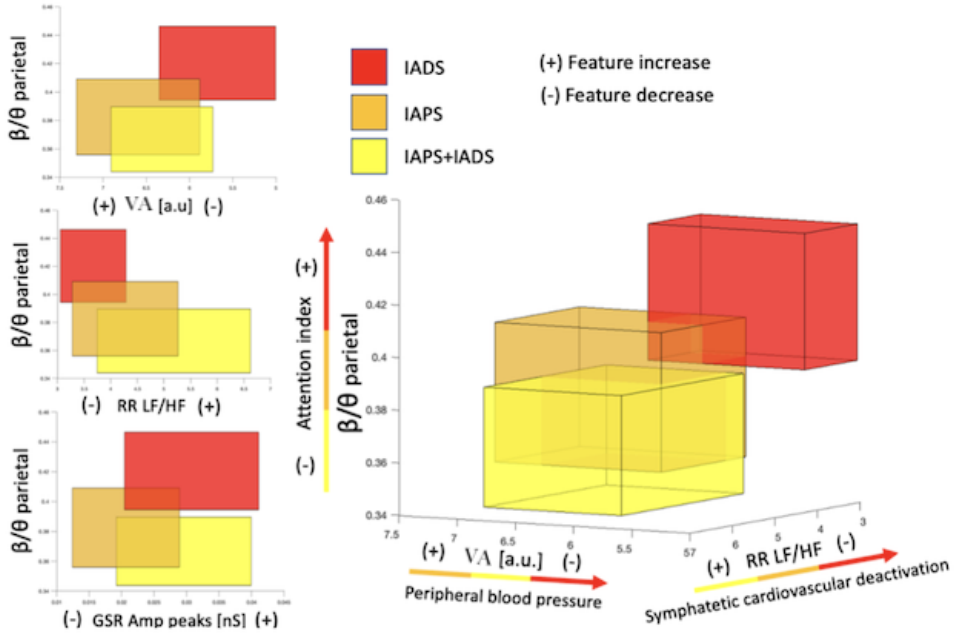

Figure 4.8: 2D and 3D boxplots are built by centering the rectangles around the coordinates of the corresponding average values, where the length of the sides of the rectangles is set equal to the standard error of the average estimations.

Table 4.6: Best performing models and the relative train accuracies, average validation accuracies and test accuracies are reported.

Stimulation	Model	Train accuracy	Validation accuracy	Test accuracy
All stimuli	RF	0.95	0.73	0.52
IAPS-only	KNN	0.39	0.35	–
IADS-only	ADB	0.55	0.43	–
IAPS+IADS	ADB	0.56	0.44	–

were comprehended as intended during the protocol design phase, as the majority of participants accurately labeled the auditory and visual stimuli with the associated emotions. Notably, the images in all scenes except for the happy one displayed a greater consistency in emotion labeling compared to the sounds which obtained slightly lower percentages.

In order to compare all features with a common starting point, an analysis was conducted on the four different baselines. Since the baselines did not seem to exhibit a behavioral trend with features related to sympathetic activation, a signal variance analysis was performed to determine the most suitable baseline for detrending all features. Baseline 2 (before the relaxing scene) was selected as it had fewer variable signals and lower sympathetic indices, indicating a calmer virtual room for subjects. To avoid bias in comparing virtual scenes, an initial calibration phase will be necessary to detrend calculated features in a real application.

Table 4.7 shows the features that have been found to be significantly different between at least two emotions.

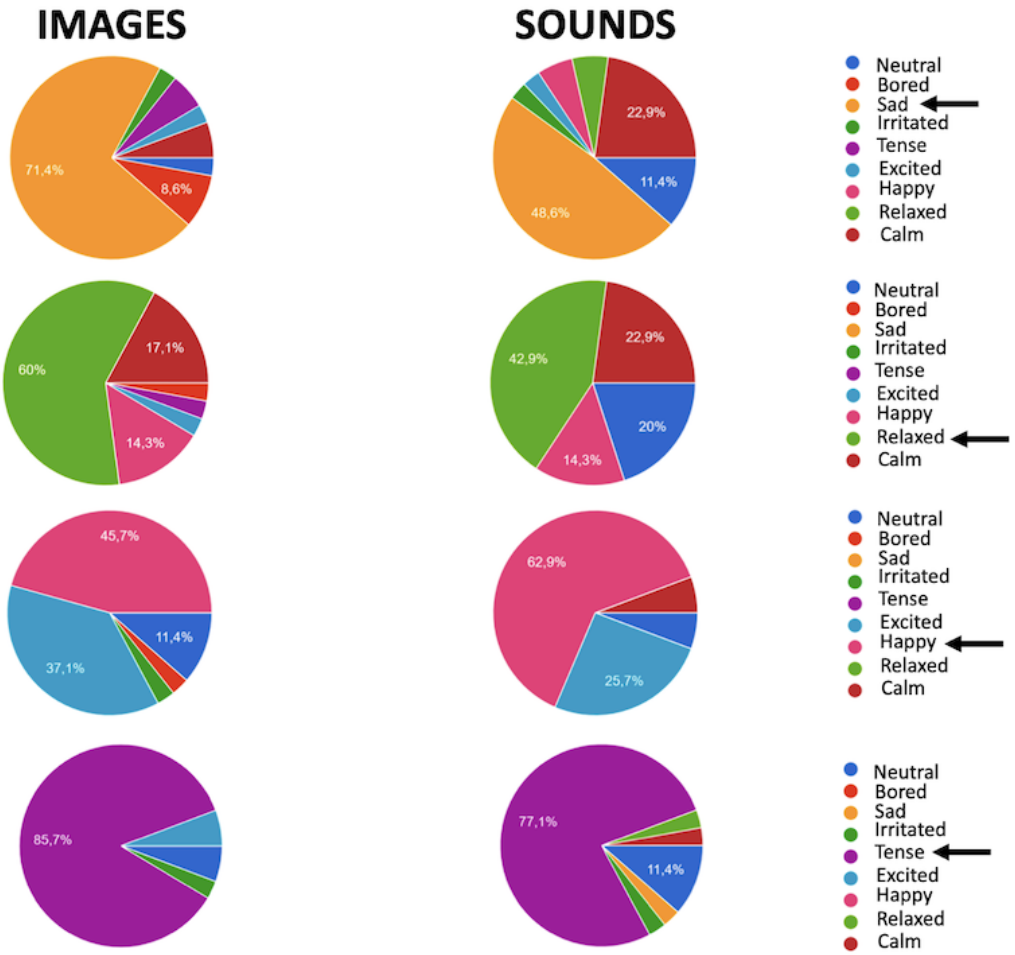

Figure 4.9: Pie charts relating to the post-protocol survey administered to the subjects. On the left and right, pie charts related to the visual and auditory stimuli of the protocol are respectively shown. The arrow next to each legend indicates the evoked emotion.

Due to the duration of each emotional scene being 5 minutes, it was feasible to compute both traditional and point process framework features. Specifically, a minimum signal length of 5 minutes is imperative to obtain reliable traditional measurements, especially in regards to the ECG signal [179].

The following results are presented by signal.

ECG: Concerning the ECG signal, no features showed significant differences between any pairs of emotions, although trends are identifiable. In Table 8, even though it did not display statistically significant differences, the RRLFtoHF feature, which represents a feature closely linked to the sympatho-vagal balance of our organism, was reported. From RRLFtoHF, it is possible to observe a growing trend passing from low-arousal emotions such as sadness and relaxation, to higher arousal emotions (i.e., happiness and fear), with the highest median value in fear (see Figure 4.10).

BVP: The BVP signal turned out to be one of the most important in separating the four emotions. From Table 4.7, it can be observed that both happiness and fear were significantly different from the VA feature in all comparisons with the other emotions. The only non-significant comparison

was between relaxation and sadness. Therefore, it is evident that the VA feature can separate not only the arousal dimension but also the valence dimension if it is at high arousal. Generally, there is a decreasing trend from sadness to fear in the amplitude of this feature, which is always negative, indicating that a smaller blood volume is present compared to the baseline (see Figure 4.10). Since blood volume is closely related to the inverse of peripheral blood pressure, it means that higher arousal emotions generate greater sympathetic activation, increasing the peripheral blood pressure measured.

GSR: The GSR signal produced very interesting results. It is clear how closely related the GSR signal is to the excitement associated with emotions. Indeed, most features show a decidedly increasing trend passing from the first two low-arousal emotions to the last two in increasing arousal order. The fear and relaxation emotions are the only ones that can be significantly separated from all the other emotions, as can be seen from the N peaks and Avg GSR features (see Figure 4.10). In particular, N peaks can significantly separate all emotions except sadness and happiness, which, however, are significantly different in many of the other features.

RESP: At the respiratory level, the cardio-respiratory coupling evaluated through RSA does not seem to discriminate between the different emotions. However, the amplitude of breaths (μ_{RESP}) is significantly different between sadness and fear, thus managing to separate two emotions at the extremes of the circumplex model of Russell. For completeness, Table 4.10 also reports the frequency at maximum coherence calculated in the HF band, which generally shows an increasing trend with arousal, demonstrating that as arousal increases, breath amplitude decreases, but respiratory rate increases (see Figure 4.10). Breaths are therefore less extensive but more frequent.

Figure 4.11, 4.12 and 4.13 show 3D boxplots built using the method described in section 3.1.3, which represent how the three most effective features can separate the four emotions, the arousal dimension, and the valence dimension, respectively.

Table 4.7: Median and ranges of all the features computed in the 4 emotions. For the sake of clarity, only features that showed statistically significant differences are shown except for RRLFtoHF and $fmax_{HF}$. Statistically significant differences are shown in bold and the last column specifies between which emotions these differences are observed (S: Sadness, R: Relaxation, H: Happiness and F: Fear).

	Sadness	**Relaxation**	**Happiness**	**Fear**	*
			ECG		
RRLFtoHF	-0.29 (2.57)	0.06 (2.20)	0.50 (3.10)	0.91 (3.69)	–
			BVP		
VA [a.u.]	-0.33 (1.28)	-0.40 (1.61)	-1.64 (2.07)	-2.22 (3.96)	Svs.H,F-Rvs.H,F-Hvs.all-Fvs.all
			GSR		
Avg amplitude peaks [μS]*10^{-2}	0.19 (0.79)	0.11 (1.44)	0.94 (2.48)	3.10 (4.45)	Svs.H,F-Rvs.H,F-Hvs.S,R-Fvs.S,R
Sd amplitude peaks [μS]*10^{-2}	0 (0.78)	0 (1.67)	0.99 (2.58)	1.96 (3.51)	Svs.H,F-Rvs.H,F-Hvs.S,R-Fvs.S,R
Avg rise time [s]]*10^{-2}	7.64 (19.02)	0 (10.28)	5.21 (23.33)	13.74 (43.23)	Svs.F-Rvs.F-Hvs.none-Fvs.S,R
N peaks	2 (6)	0 (6.25)	4 (6)	9 (15.75)	Svs.R,F-Rvs.all-Hvs.R,F-Fvs.all
Max sign amplitude [μS]*10^{-2}	0.03 (6.07)	0.46 (7.43)	2.24 (9.14)	7.12 (20.40)	Svs.H,F-Rvs.F-Hvs.S-Fvs.S,R
Sd der amplitude [μS]*10^{-5}	13.71 (32.94)	1.50 (59.69)	62.68 (188.70)	214.11 (355.11)	Svs.H,F-Rvs.H,F-Hvs.S,R-Fvs.S,R
Max der amplitude [μS]*10^{-5}	0.01 (0.39)	0.014 (0.44)	0.33 (0.75)	0.87 (1.11)	Svs.H,F-Rvs.H,F-Hvs.S,R-Fvs.S,R
Avg GSR [μS] *10^{-2}	1.15 (26.17)	2.98 (31.09)	27.27 (42.93)	68.97 (95.37)	Svs.F-Rvs.H,F-Hvs.R,F-Fvs.all
Sd GSR [μS] *10^{-2}	1.70 (10.24)	0.78 (5.90)	6.06 (8.39)	11.34 (19.16)	Svs.F-Rvs.F-Hvs.F-F.vs.all
Avg abs1 [μS]*10^{-4}	1.43 (6.51)	0.19 (6.13)	9.46 (19.90)	32.46 (41.78)	Svs.H,F-Rvs.H,F-Hvs.S,R-Fvs.S,R
Avg abs1 norm *10^{-2}	0.13 (1.07)	-0.06 (0.83)	0.27 (0.66)	0.54 (1.10)	Svs.F-Rvs.F-Hvsnone-Fvs.S,R
Avg abs2 [μS]*10^{-5}	1.08 (14.71)	-0.38 (9.73)	9.60 (47.47)	45.12 (70.97)	Svs.H,F-Rvs.H,F-Hvs.S,R-Fvs.S,R
Env [μS]*10^{-2}	0.14 (1.07)	0.03 (0.73)	1.12 (2.97)	3.89 (5.03)	Svs.H,F-Rvs.H,F-Hvs.S,R-Fvs.S,R
			RESP		
μ_{RESP} [a.u.]	0.08 (9.31)	0.79 (9.86)	-1.17 (16.48)	-0.43 (10.86)	Svs.F-Rvsnone-Hvsnone-Fvs.S
$fmax_{HF}$ [Hz]*10^{-2}	-0.64 (10.43)	1.30 (9.01)	1.63 (7.77)	2.15 (10.01)	–

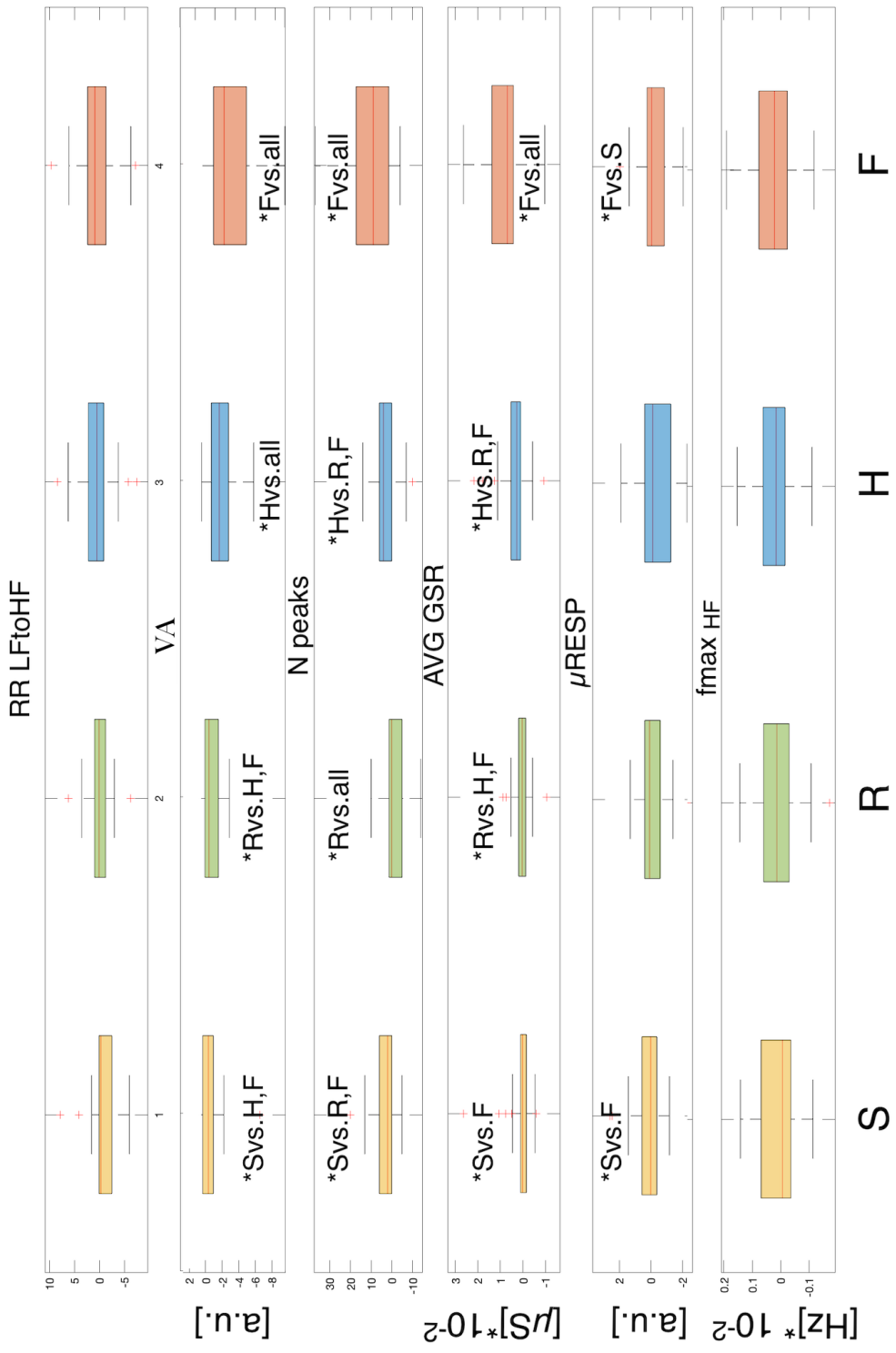

Figure 4.10: The figure reports, for each signal linked to autonomic activity, one or two of among the best representative features found in the comparison among the four emotions (S: Sadness, R: Relaxation, H: Happiness, F: Fear). From top to buttom ECG (RR LF/HF), BVP (VA), GSR (N peaks and AVG GSR) and RESP (μ_{RESP} and $fmax_{HF}$). Statistically significant differences are marked with *.

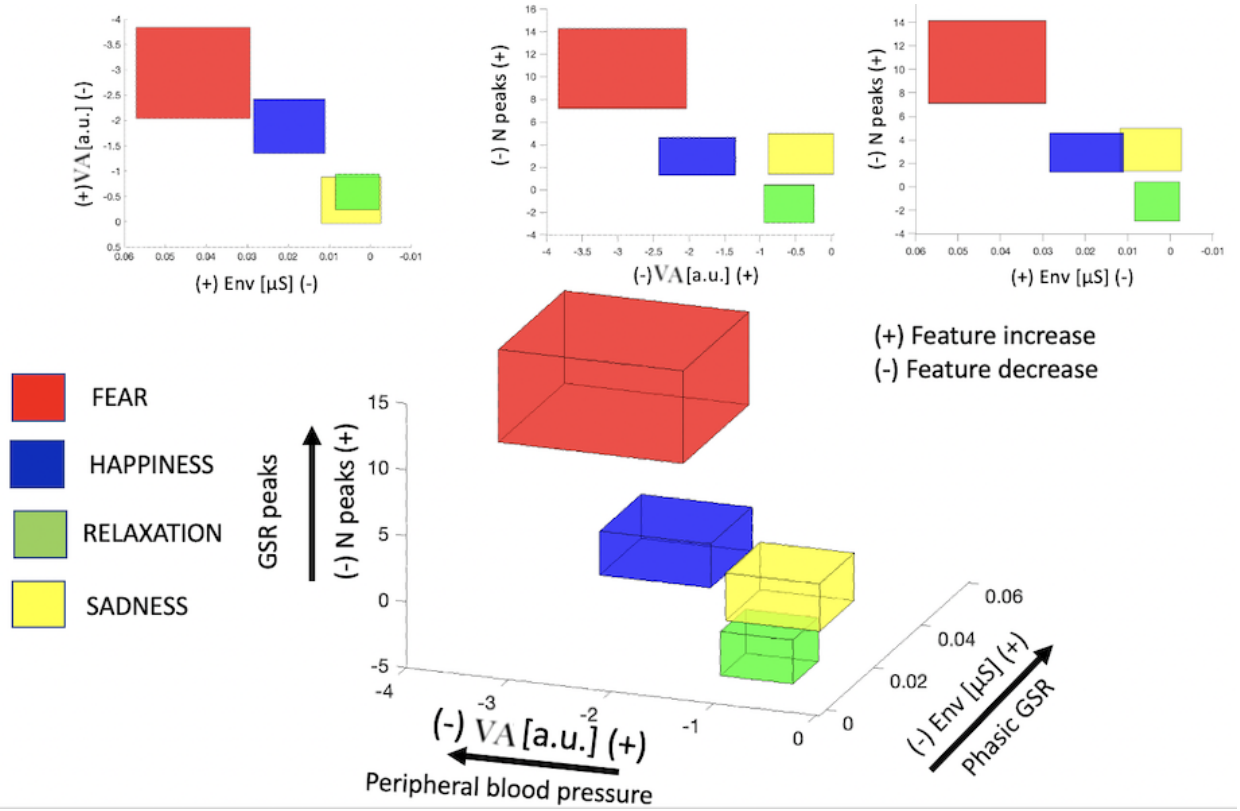

Figure 4.11: 3D boxes are displayed, one for each emotion, where the center of each box (i.e., each emotion) is given by the mean of the three features and the length of the sides is given by the 95% confidence limits for the mean estimate. Specifically, the main figure shows the combination of the three most performing features in the space of the four emotions. Above, the projections of the 3D boxes in two dimensions are presented.

Regarding the ECG signal, having access to a baseline prior to the protocol that allowed the use of the point process as a method for estimating HRV, it was decided to use this method over traditional ones in order to obtain a precise estimation of HRV features, even with future applications for real-time monitoring in mind. However, since this protocol is the only one that provides stimulations (albeit non-stationary) lasting 5 minutes, traditional methods could still be used for HRV estimation. Despite the fact that the point process model has been extensively validated in literature, a comparison was made between one of the more easily calculated traditional indices, i.e. AVNN, and the corresponding index mu_{RR} calculated by the point process, for completeness. For each scene related to the four emotions, a Bland-Altman plot was created to assess the agreement between the two different methods. From Figure 4.14, it is possible to observe that there is agreement between the two measurements. Additionally, a Wilcoxon signed rank test was performed for each emotion between the two variables, which did not show significance for any pair of comparisons ($\text{p}_{sadness} = 0.73$, $\text{p}_{relaxation} = 0.56$, $\text{p}_{happiness} = 0.54$, $\text{p}_{fear} = 0.13$).

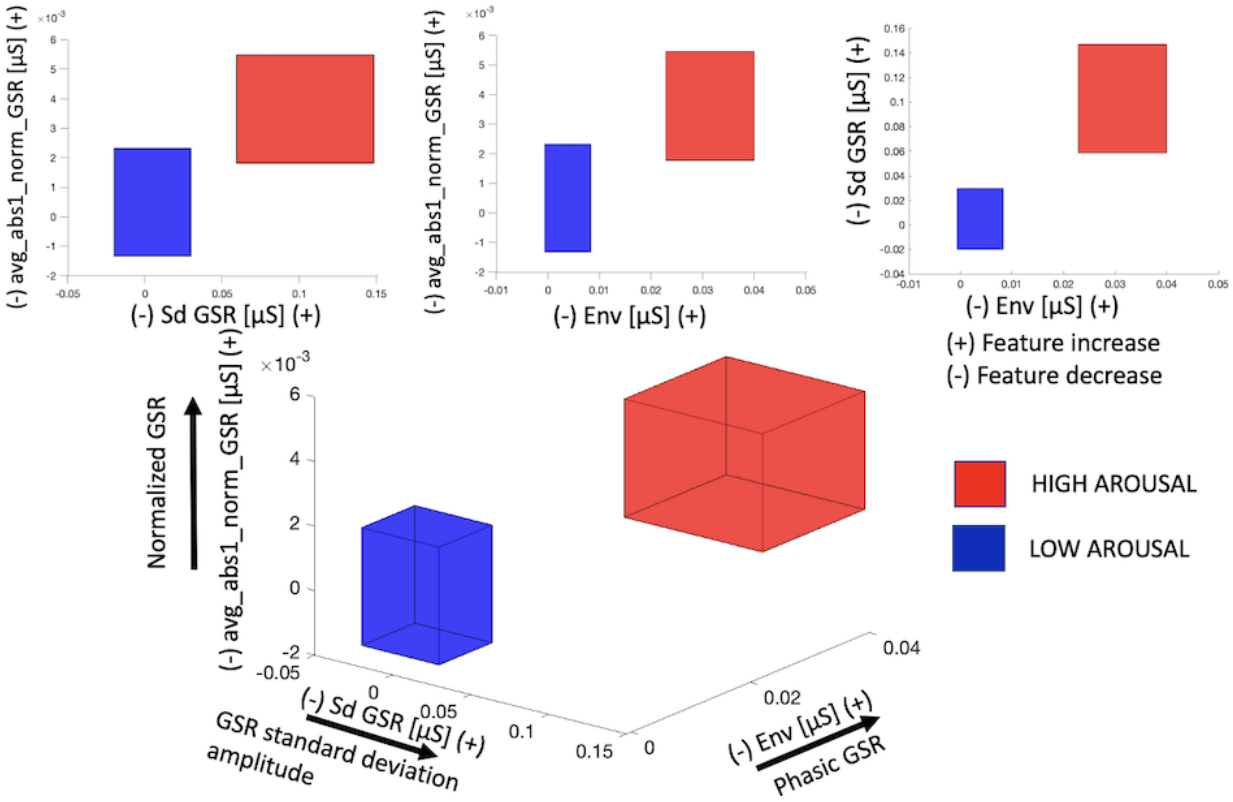

Figure 4.12: 3D boxes are displayed, one for each emotion, where the center of each box (i.e., each emotion) is given by the mean of the three features and the length of the sides is given by the 95% confidence limits for the mean estimate. Specifically, the main figure shows the combination of the three most performing features able to separate the four emotions along the arousal dimension. Above, the projections of the 3D boxes in two dimensions are presented.

Classification

The results obtained through the ML approach need to be sorted based on the selected features and models. Then, a comparison is made between the accuracies of all the models. Since many tests were conducted, the following Tables 4.8, 4.9 and 4.10 only show the most relevant outcomes divided into the three tasks (arousal, valence and 4 emotions). Each table will present the model used, the feature selection method applied, the signals used to select the features, the number of selected features, the hyperparameter values (Hyp), as well as the validation and test accuracies. The Tables contain two additional columns for the AUC in both training and testing. Specifically, in the case of the 4 emotions (sadness, relaxation, happiness, and fear), the values were obtained using the One versus Rest (OVR) criterion for both training and testing.

To simplify the tabular results and better understand the hyperparameters selected for tuning, it is important to explain which parameters were chosen for each model. For more detailed information, please refer to the scikit-learn website. In the tables, the parameters are listed in the same order as the following list.

- KNN: The tuning process involved selecting parameters for the number of neighbors and observation weights. As a general rule, the number of neighbors is often set to the square root of the total number of observations, which was the maximum value we used. In terms

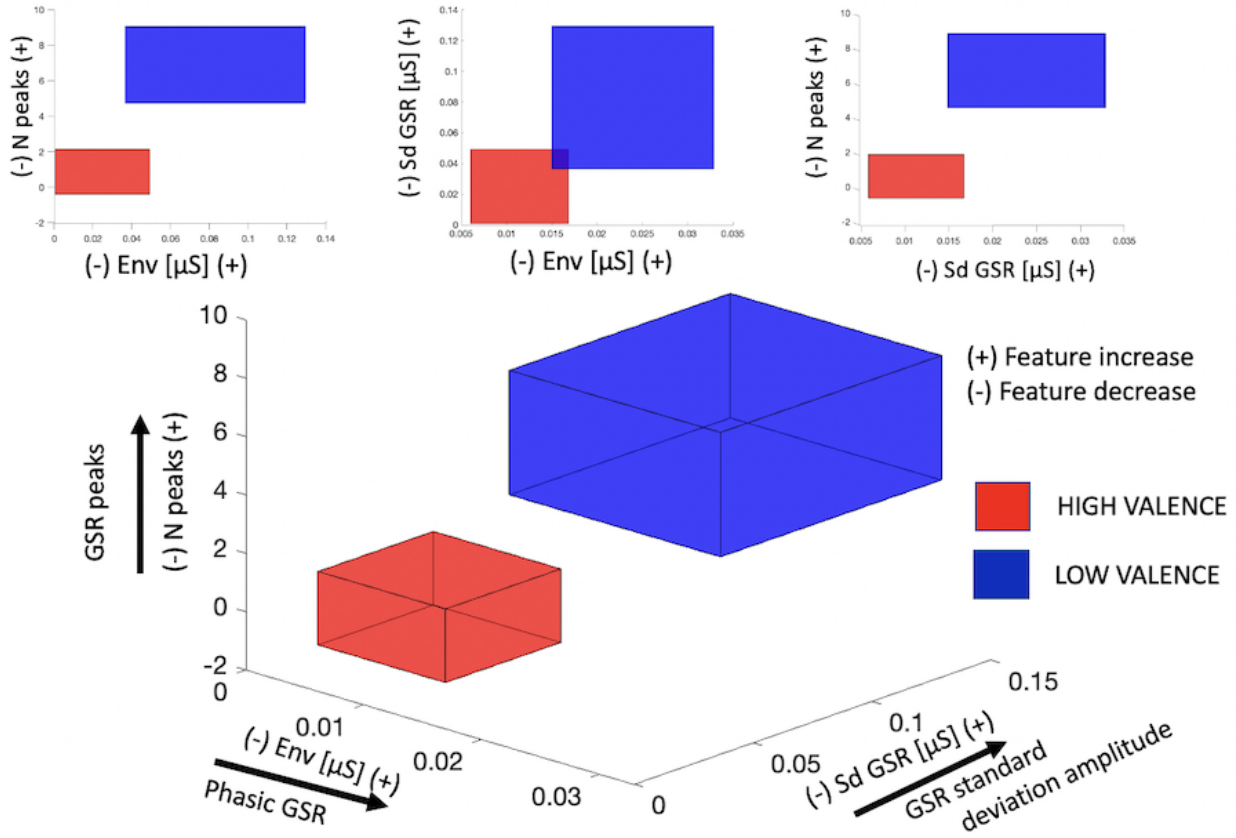

Figure 4.13: 3D boxes are displayed, one for each emotion, where the center of each box (i.e., each emotion) is given by the mean of the three features and the length of the sides is given by the 95% confidence limits for the mean estimate. Specifically, the main figure shows the combination of the three most performing features able to separate the four emotions along the valence dimension. Above, the projections of the 3D boxes in two dimensions are presented.

of observation weights, all tests produced uniform results, which is the default setting, so this parameter was not included in the tables.

- LDA: LDA tuning was performed on three different parameters, namely solver, shrinkage, and store covariance. Solver determines the algorithm to be used, and the options available include Singular Value Decomposition (SVD), Least Squares Solution (LSQR), and Eigenvalue Decomposition (Eigen). Shrinkage can either be set to none or automatic. Finally, the store covariance parameter explicitly calculates the weighted within-class covariance matrix when the solver is set to 'svd'.

- LR: In logistic regression, the focus was on the following parameters: penalty, solver, class weight, multi-class, and max iter. Penalty has several options, including l1, l2, elasticnet, or none. We tested different solvers, such as newton-cg, lbfgs, liblinear, sag, and saga. Class weight can be balanced, none, or a dictionary. Multi-class can be set to auto, ovr, or multinomial. The default value for max iter was 400.

- SVM: For SVM, the focus was on three parameters: kernel, gamma coefficient, and decision function shape. The available kernel options were linear, poly, rbf, and sigmoid. The gamma

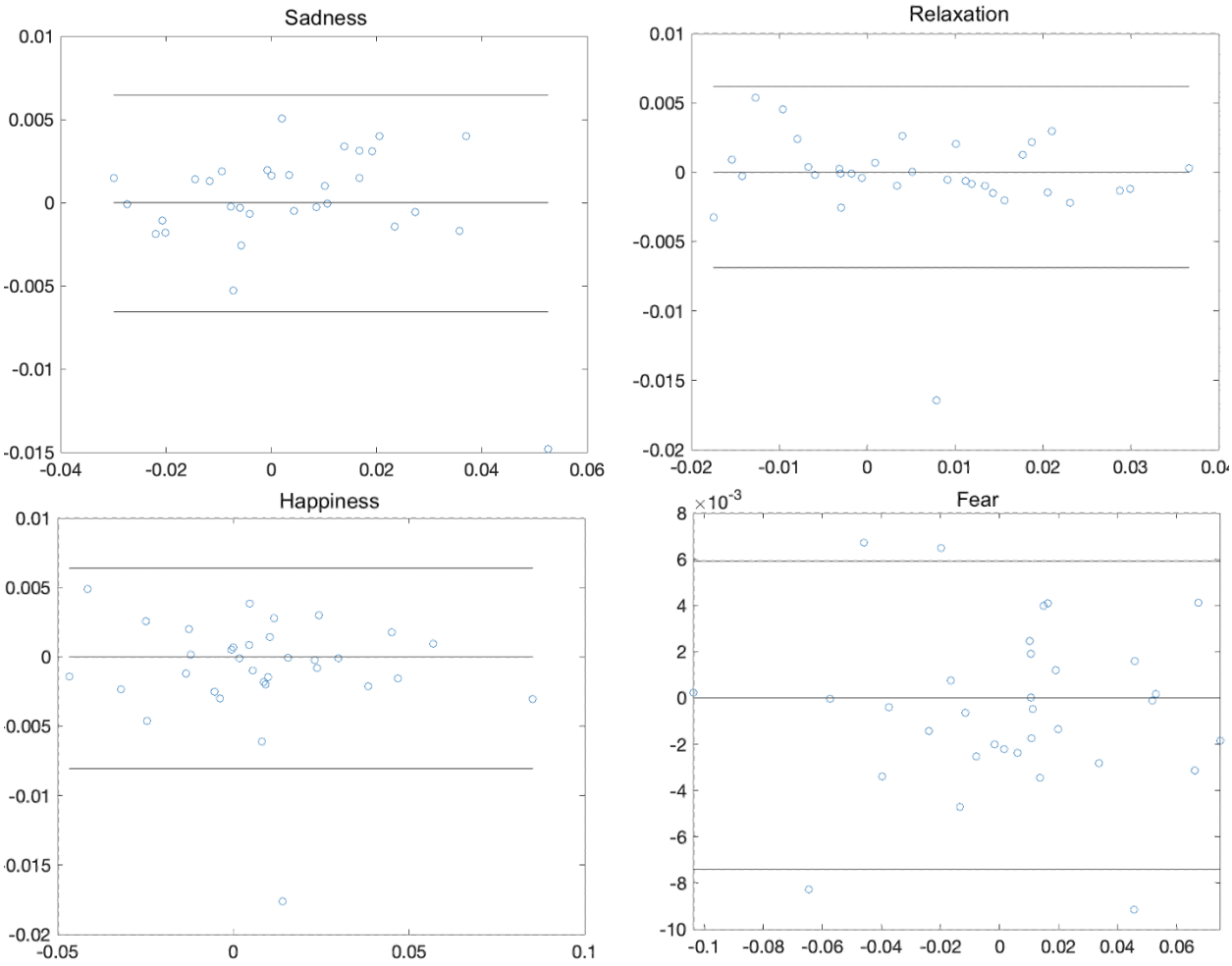

Figure 4.14: The Bland Altman plots illustrate the relationship between AVNN and muRR for each emotion, portraying the differences in seconds for each subject on the y axis and the average of these differences on the x axis. The central line represents the mean difference between AVNN and muRR, while the top and bottom lines display the 95% confidential intervals for the average estimation.

coefficient can take on the values of scale or auto. The decision function shape can be either one-versus-rest or one-versus-one.

In addition, ROC curves for the best performing models in classifying the four emotions, arousal, and valence are presented in Figures 4.15, 4.16 and 4.17 for both training and testing data.

Table 4.8: Machine learning results for classifying the arousal dimension (i.e., low and high). The validation accuracy is average. The AUCs are reported in protocol order for the emotions (i.e., sadness, relaxation, happiness, and fear).

Model	KNN	LDA	lR	SVM
Feature Selection Method	SFS	SM	SFS	SM
Selected signals	GSR,ECG,BVP	GSR,ECG,BVP,RESP	GSR,ECG,RESP	GSR,ECG,BVP,RESP
#features	14	19	12	19
Hyp	8	lsqr, auto, True	12,newton-cg, balanced, auto	rbf, scale, ovr
Validation acc.	0.83	0.78	0.81	0.80
Test acc.	0.89	0.89	0.86	0.86
AUC train	0.91	0.89	0.88	0.95
AUC test	0.94	0.91	0.93	0.92

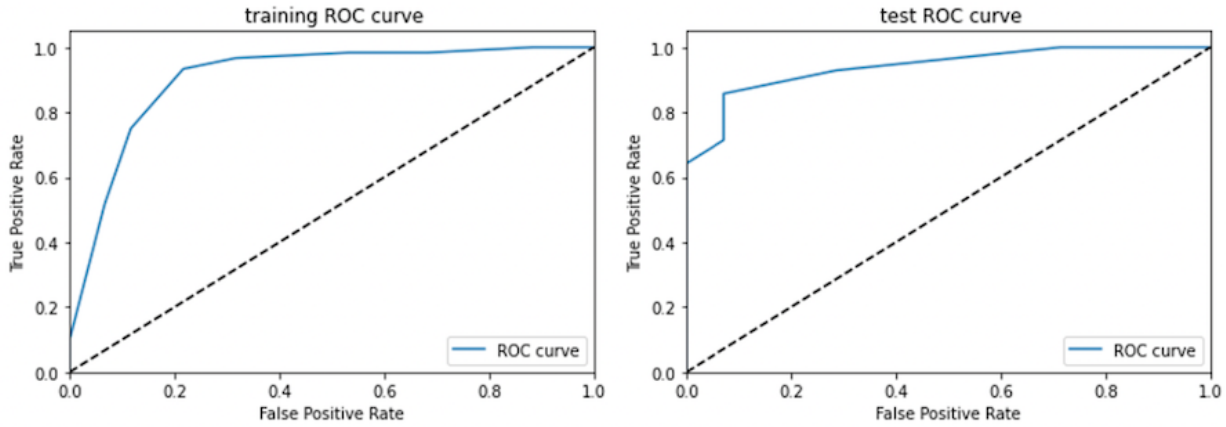

Figure 4.15: ROC curves of the best arousal model (KNN).

Table 4.9: Machine learning results for classifying the valence dimension (i.e., low and high). The validation accuracy is average. The AUCs are reported in protocol order for the emotions (i.e., sadness, relaxation, happiness, and fear).

Model	KNN	LDA	lR	SVM
Feature Selection Method	KB	KB	SFS	SFS
Selected signals	GSR,ECG,RESP	GSR,ECG,RESP	GSR,ECG,BVP	GSR,ECG,BVP
#features	6	18	13	3
Hyp	8	lsqr, auto, True	12,newton-cg, balanced, auto	linear, scale, ovr
Validation acc.	0.71	0.63	0.67	0.70
Test acc.	0.57	0.59	0.64	0.72
AUC train	0.85	0.76	0.78	0.83
AUC test	0.65	0.57	0.66	0.71

Table 4.10: Machine learning results for classifying 4 emotions. The validation accuracy is average. The AUCs are reported in protocol order for the emotions (i.e., sadness, relaxation, happiness, and fear).

Model	KNN	LDA	lR	SVM
Feature Selection Method	SFS	SM	KB	KB
Selected signals	GSR,ECG	GSR,BVP	GSR,ECG,BVP,RESP	GSR,ECG,BVP,RESP
#features	4	3	13	14
Hyp	7svd,None,True	one,newton-cg,balanced,ovr	rbf,scale,ovr	
Validation acc.	0.56	0.53	0.53	0.53
Test acc.	0.50	0.57	0.61	0.54
AUC train	[0.91, 0.83, 0.89, 0.92]	[0.79, 0.79, 0.61, 0.83]	[0.81, 0.84, 0.69, 0.90]	[0.81, 0.76, 0.40, 0.88]
AUC test	[0.74, 0.67, 0.76, 0.7]	[0.76, 0.83, 0.66, 0.93]	[0.79, 0.75, 0.65, 0.96]	[0.78, 0.72, 0.61, 0.86]

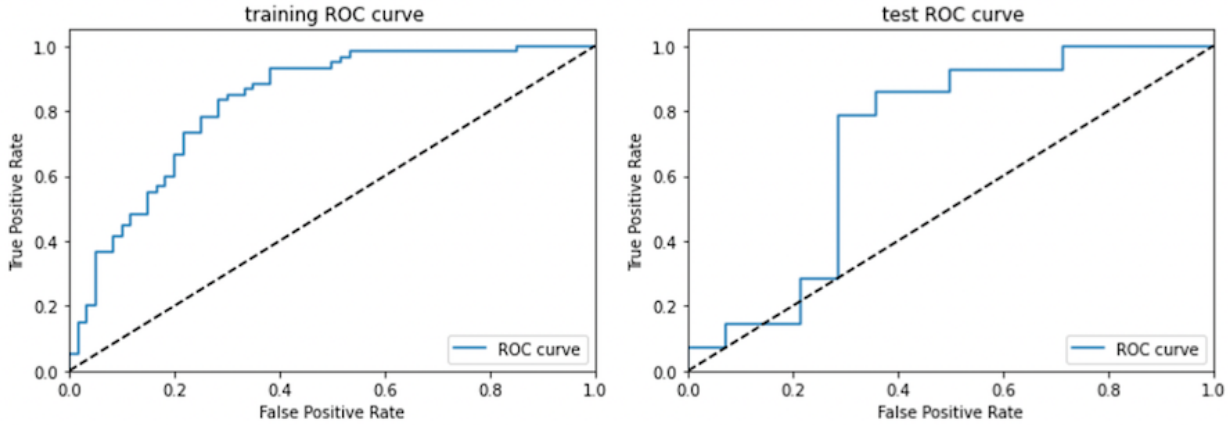

Figure 4.16: ROC curves of the best valence model (SVM).

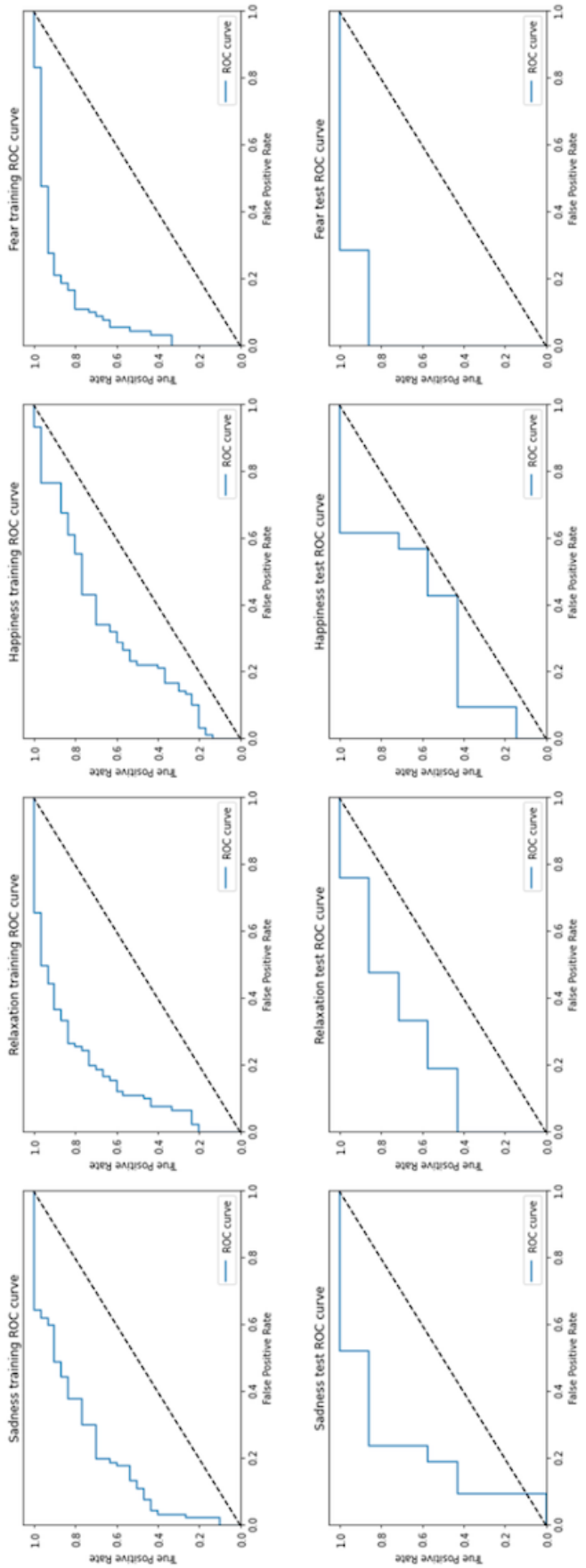

Figure 4.17: ROC curves of the best 4-emotions model (RL).

4.4 Summary of results

Listening effort

The study analyzed the effects of two different levels of listening effort on physiological responses using various physiological signals by using a novel speech-in-noise test. To the best of our knowledge, this is the only study that investigates the issue of listening effort using such a large number of physiological signals.

- PUPIL features did not show significant differences between effort phases, but the average pupil diameter was higher during high effort.
- Cardiac features and spectral power exhibited significant differences between baseline and high-effort phase, with a heartbeat acceleration during high effort.
- RESP features indicated increased breath amplitude during high effort.
- GSR features suggested higher arousal and sweating in the low-effort phase.
- BVP features showed greater sympathetic control during high effort.
- EEG features showed higher attention index and engagement in the low-effort phase.

Results of the listening effort protocol and the creation and validation of the speech-in-noise test used for this purpose can be found in the following studies:

1. **E. M. Polo**, M. Mollura, A. Paglialonga, & R. Barbieri. (2021, September). Listening Effort: Cardiovascular Investigation Through the Point Process. In 2022 Computing in Cardiology (CinC). IN PRESS. DOI: 10.22489/CinC.2022.211
2. **E. M. Polo**, M. Lenatti, M. Zanet, R. Barbieri, A. Paglialonga. 'Preliminary evaluation of the Speech Reception Threshold measured using a new language-independent screening test as a predictor of hearing loss'. Abstract presented at 1st Virtual Conference on Computational Audiology (VCCA). (June 19 2020).
3. **E. M. Polo**, M. Zanet, M. Lenatti, T. van Waterschoot, R.Barbieri, A.Paglialonga, "Development and Evaluation of a Novel Method for Adult Hearing Screening: Towards a Dedicated Smartphone App", Proceedings of the 7th EAI International Conference on IoT Technologies for HealthCare, 2020 [1]. DOI: 10.1007/978-3-030-69963-5_1
4. **E. M. Polo**, M. Zanet, A.Paglialonga, R.Barbieri. "Preliminary Evaluation of a Novel Language Independent Speech-in-Noise Test for Adult Hearing Screening." European Medical and Biological Engineering Conference. Springer, Cham, 2020 [2]. DOI: 10.1007/978-3-030-64610-3_109
5. A. Paglialonga, **E. M. Polo**, M. Zanet, G. Rocco, T. van Waterschoot, R. Barbieri, "An Automated Speech-in-Noise Test for Remote Testing: Development and Preliminary Evaluation", American Journal of Audiology, vol. 29, no. 3S, pp. 564-576, 2020 [3]. DOI: 10.1044/2020_AJA-19-00071

6. M. Zanet*, **E. M. Polo***, M. Lenatti, T. van Waterschoot, M. Mongelli, R. Barbieri, A. Paglialonga. "Evaluation of a Novel Speech-in-Noise Test for Hearing Screening: Classification Performance and Transducers Characteristics." IEEE Journal of Biomedical and Health Informatics (2021) [4]. DOI: 10.1109/JBHI.2021.3100368

7. M. Lenatti, **E. M. Polo**, M. Paolini, M. Mollura, M. Zanet, R. Barbieri, A. Paglialonga. (2021) 'Evaluation of multivariate classification algorithms for hearing loss detection through a speech-in-noise test'. Abstract presented at 2st Virtual Conference on Computational Audiology (VCCA). (June 25 2021).

8. **E. M. Polo**, M. Mollura, R. Barbieri, & A. Paglialonga. (2023, March). Multivariate Classification of Mild and Moderate Hearing Loss Using a Speech-in-Noise Test for Hearing Screening at a Distance. In IoT Technologies for HealthCare: 9th EAI International Conference, HealthyIoT 2022, Braga, Portugal, November 16-18, 2022, Proceedings (pp. 81-92). Cham: Springer Nature Switzerland [5].

9. M. Lenatti, P. A. Moreno-Sánchez, **E. M. Polo**, M. Mollura, R. Barbieri, & A. Paglialonga (2022). Evaluation of machine learning algorithms and explainability techniques to detect hearing loss from a speech-in-noise screening test. American Journal of Audiology, 31(3S), 961-979 [6]. DOI: 10.1044/2022_AJA-21-00194

   * co-first authors

Emotion and music (AuBT protocol)

The study investigated the physiological responses to auditory stimulation with music and yielded noteworthy results through the use of machine learning models on an online dataset that has not been extensively studied before, particularly in terms of the connection between emotions and physiology..

- GSR features distinguished high and low excitement emotions.
- AVNN from the ECG signal characterized joy and sadness in the most different way.
- RESP was important for recognizing anger, with a higher average rate than other emotions.
- At the classification level, the study achieved excellent performance in correctly classifying the four emotions with an average validation accuracy of 85% for the four-class classification and more than 90% for the binary classification of the two dimensions of valence and arousal.

Results of this study were accepted to the Eighth national congress of bioengineering: **E. M. Polo**, M. Mollura, A. Paglialonga & R. Barbieri. (2022) Decoding Emotions through Music: A Physiological Analysis of Emotion Recognition. Proceedings of the VIII Congress of the National Association of Bioengineering(GNB 2023), Jun 18-20 2023, Padova, Italy.

Emotional protocol with visual, auditory, and combined stimuli

The study investigated physiological responses to various types of stimuli and yielded interesting findings. To our knowledge, this is one of the few studies that combines two widely-used databases in the literature, namely IAPS and IADS, separately and jointly to investigate different physiological patterns in response to different types of stimuli.

- GSR amplitude-related features were higher during the phase with only sounds and high valence stimuli elicited a higher GSR response.
- Average modelled R-R interval was higher in the phase with only sounds, indicating a slowing down of the heartbeat compensated by a blood pressure increase.
- Power in the high-frequency band of the respiratory signal was higher in the phase with only images, indicating higher parasympathetic nervous system activity.
- EEG arousal sessions showed higher activation in the delta frequency range in the IADS-only phase and higher attention levels in the frontal and parietal regions in the IADS-only phase.
- The KNN model achieved a test accuracy of 52% in correctly categorizing the four quadrants of Russel's circumplex model using all three methods of stimulation.
- Sound phases were more effective in creating a clear physiological distinction between the four quadrants compared to visual stimuli alone, where the machine learning models achieved lower performance.

Results of this study can be found in the following publications:

1. **E. M. Polo**, Farabbi, A., M. Mollura, R. Barbieri, A. Paglialonga & L. Mainardi. (2022, July). Analysis of the skin conductance and pupil signals for evaluation of emotional elicitation by images and sounds. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 1968-1971) [7].
2. **E. M. Polo**, M. Mollura, A. Paglialonga & R. Barbieri. (2022) 'Quantitative assessment of the influence of sound in affective audio-visual elicitations'. Abstract presented at HEAL 2022 - HEaring Across the Lifespan (16-18 June 2022).
3. **E. M. Polo**, A. Farabbi, M. Mollura, A. Paglialonga, L. Mainardi, R. Barbieri. (2022). Comparative assessment of physiological responses to emotional elicitation by auditory and visual stimuli. IEEE Journal of Translational Engineering in Health and Medicine. SUBMITTED.
4. **E. M. Polo**, A. Farabbi, L. Mainardi, R.Barbieri. Unlocking the Power of Emotion in Marketing: Using Machine Learning to Analyze Neurophysiological Responses to Visual, Auditory, and Combined Stimulation. Abstract accepted for further publication on Frontiers in Human Neuroscience.

5. A. Farabbi, **E. M. Polo**, R. Barbieri, & L. Mainardi. (2022, July). Comparison of different emotion stimulation modalities: an EEG signal analysis. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 3710-3713). IEEE [8].

Emotions and virtual reality

The study investigated how participants' physiological responses varied in reaction to different emotional stimuli, presented in a virtual reality setting. These stimuli included emotions such as sadness, relaxation, happiness, and fear. This study is unique in its approach, as it designed specific scenes in virtual reality to elicit and study emotions while collecting many physiological signals. As far as we know, this is the first study to utilize such a comprehensive approach to understanding the relationship between emotions and physiological signals in a virtual reality environment.

- The ECG signal did not show significant differences between emotions, but trends were observable. RRLFtoHF feature showed a growing trend from low to high arousal emotions, with fear having the highest median value.
- The BVP signal was important in separating emotions, with happiness and fear significantly different from other emotions. BVP amplitude decreased from sadness to fear, indicating greater sympathetic activation in higher arousal emotions.
- The GSR signal showed a close relationship with excitement, with fear and relaxation significantly separated from other emotions in average amplitude and number of peaks.
- Respiratory activity showed that breath amplitude was significantly different between sadness and fear, separating these two emotions at opposite ends of the circumplex model of Russell.
- At the classification level, excellent performance was achieved in correctly classifying the 4 emotions and the two dimensions of valence and arousal, with best models achieving test accuracies of 61%, 89% and 71% for the 4-class, arousal and valence classification respectively.

These results have just been achieved in the last year and they are in the process of being submitted in the following formats:

1. **E. M. Polo**, M. Mollura, A. Paglialonga, & R. Barbieri. Exploring Emotional Responses in Virtual Reality Through Skin Conductance Signal. Proceedings of the 2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE 2023), Milan XXX 2023. Accepted abstract.
2. **E. M. Polo**, M. Mollura, A. Paglialonga, & R. Barbieri. Exploring Emotions in Virtual Reality: Enhancing Recognition through Physiological Signals Acquisition. IEEE Transactions on Affective Computing. To be sumbitted soon.

Below are the methodological studies that served for all the studies:

1. **E. M. Polo**, M. Mollura, Lenatti, M. Lenatti, A. Paglialonga, & R. Barbieri. (2021, November). Emotion recognition from multimodal physiological measurements based on an interpretable feature selection method. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 989-992). IEEE [9]. DOI: 10.1109/EMBC46164.2021.9631019

2. **E. M. Polo**, M. Mollura, M. Zanet, M. Lenatti, A. Paglialonga, & R. Barbieri. (2021, September). Analysis of the Effect of Emotion Elicitation on the Cardiovascular System. In 2021 Computing in Cardiology (CinC) (Vol. 48, pp. 1-4). IEEE [10].
DOI: 10.23919/CinC53138.2021.9662859

In summary, these studies examined the relationship between physiological responses and various stimuli, including speech-in-noise test, music, flat screen emotion elicitation, and virtual reality environments. Using peripheral and central physiological signals and machine learning models, the studies accurately classified emotions and dimensions of valence and arousal. The results demonstrated that listening effort was associated with changes in pupil diameter, cardiac acceleration, and breath amplitude, while higher attention and engagement were linked to lower effort. GSR-related measures distinguished high and low excitement emotions, and BVP-related measures were important in stratifying emotions in valence and arousal dimensions. These studies highlight the importance of using multiple signals and emphasize the innovative and effective use of virtual reality environments to elicit emotions in a more realistic setting.

Chapter 5

Discussion

5.1 Overview of findings and their implications

As expounded in the thesis aims, the central objective of this thesis is to scrutinize emotional and stress responses via the analysis of both peripheral and central physiological signals. To this end, four protocols were used in order to physiologically investigate listening effort (section 3.1), emotions elicited by music (section 3.1), emotions elicited by three distinct types of sensory stimulation, namely auditory, visual, and a combination of both (section 3.2), and emotions elicited by more complex stimuli such as virtual reality (section 3.3). In the subsequent discussion, the outcomes obtained for each implemented protocol will be expounded upon.

Listening effort

As introduced in section 2.2.7, there has been much discussion and debate within the field regarding the most appropriate and effective method to measure listening effort. Despite the fact that pupillary dilation, which increases with increased auditory effort, is the most accredited physiological measure [73] [180], it has not always been found to be a useful marker for this problem, since it is often not correlated to other objective and subjective measures [181] [182]. Specifically, in [182], it is suggested that the issue of measuring listening effort is multidimensional, and that various measures of listening effort using both qualitative and objective indices (such as physiological signals) rarely exhibit consistent correlations and may even yield contradictory results. Therefore, the study proposes that there is no universally applicable "gold standard" measure of listening effort, and emphasizes that different measures of listening effort should not be employed interchangeably.

For this reason, a speech-in-noise test was created to be simple and short, so as limit the occurrence of fatigue. Given the complexity of the problem, we aimed to measure various physiological signals, both peripheral and central, to create a clearer physiological picture of this problem. From the results shown in section 4.1, it is generally evident that sympathetic activation is significantly greater during the high-effort phase. Most of the signals, in fact, agree that, on average, subjects are more physiologically excited during the high-effort phase. Pupillary diameter, indeed, even though not significantly, appears to be larger during the high-effort phase.

In terms of cardiovascular function, time-domain features such as μRR, VA, and PAT were

found to be highly effective in differentiating high-effort from low-effort levels. Specifically, μRR, which represents the modelled RR series, exhibited an average decrease during the high-effort level, indicating an acceleration in heartbeat during the most difficult phase of the test. This acceleration was also reflected in PAT, where there was an acceleration of the pressure wave from the heart to the periphery, which reflects the high-effort sympathetic activation visible also in VA. This activation was significantly different from the high-effort phase, even during the baseline. VA which is the amplitude modulation of the BVP signal represents the volume of blood on the periphery, and a lower BVP amplitude value is linked to a greater peripheral blood pressure, which is associated with vasoconstriction. Our findings suggest that this index is a reliable marker for significantly differentiating the high-effort phase from the other two conditions. In terms of frequency-domain features, significantly lower values were observed in RR VLF and RR TOT during the high-effort phase. Total potency is now regarded as an indication of greater tachogram variability, which reflects a prevalent parasympathetic activation, consistent with the findings in the time domain. Although RR VLF is still a matter of debate and requires further clarification, several studies have associated RR VLF power with parasympathetic activity [183], in agreement with the expected response to the elicited effort levels. RR TOT and RR VLF were significantly different between the high-effort and baseline conditions, but not between the two effort phases of the test. It is worth noting that although the other frequency domain features were not significant, the trends were still consistent. From Table 4.1, we can see that RR LFn and RR LFtoHF, both proportional to the sympathovagal balance, were higher during the higher noise-induced degradation. RR HFn, on the other hand, showed lower values in the high effort phase, where the degradation in speech stimuli increased.

Regarding respiratory and cardiorespiratory coupling, the amplitude of the breaths identified by μRESP was significantly higher during high effort than low effort, with a trend of increasing respiratory frequency from low to high effort, although not significantly. The estimation of RSA indicated by GAIN12 was higher during low effort, which reinforces the idea of greater sympathetic activation during high effort, given that an increase in this index is predominantly associated with increased vagal activation [184].

Interestingly, the GSR signal does not seem to agree with the other signals in defining greater sympathetic activation during high effort. Avg GSR and AV Env which respectively represent the average amplitude of the GSR signal and the envelope of the GSR phasic component, are indeed significantly greater at low effort. However, this can be explained by the EEG derived measures, where the attention index β/θ is lower during the high-effort phase. Moreover, the β/α index representing the task engagement shows a decreasing trend from low to high effort. This physiological response pattern could potentially be explained by attention theory, as evidenced by existing research. Lacey's theory of attention [183] suggests that when an individual is engaged in a task requiring focused attention, the body tends to exhibit a decrease in heart rate, often accompanied by an increase in GSR and this behavior is clearly visible in Figure 4.3. This heart rate deceleration occurs as the body is responding to environmental stimuli that demand attention, such as the perception of a visual or auditory stimulus. The subjects are therefore more attentive and also more engaged during the low-effort phase, indicating that they are better able to understand

the stimuli they are hearing compared to the high-effort phase, in which they are likely to have greater difficulty understanding the stimuli immersed in noise.

As this is a complex problem, there may not be agreement in the literature regarding physiological measures for this reason. An increase in GSR amplitude-related features is commonly associated with greater sympathetic activation and we would indeed expect an increase in the related features during high effort. To our knowledge, this is the only study on listening effort that monitors such a large number of physiological signals. Although high effort seems to be associated with greater sympathetic activation, it may also result in decreased attention and engagement from subjects, who may enter a trade-off phase as explained by motivation theory [61]. Subjects analyze their resources to determine whether to increase effort or give up because the task is too difficult, leading to less attention paid when they do not understand all the stimuli correctly.

This analysis reinforces the notion that the issue of studying listening effort is multidimensional, and perhaps it would be useful to approach the problem also from the perspective of attention and motivation other than simple effort.

Emotion and music (AuBT protocol)

As described in Section 2.2.8, the main aim of this investigation is to explore the correlation between musical stimuli that evoke emotions and physiological responses, using the online AuBT dataset [90]. Through data analysis, our goal is to enhance our comprehension of the capacity of music to elicit emotions and the precise mechanisms that underlie these responses.

Table 4.2 illustrates that the statistical analysis of the GSR, ECG, and RESP signals demonstrates their ability to differentiate between the four emotions under study. The study found that the AVNN feature, which relates to the average of normal heartbeats, along with the RESP f and RESP amp features, were the most effective at distinguishing between emotions. These features were able to yield statistically significant differences in all comparisons between emotions, with the exception of one. The findings presented in Table 4.2 and Figure 4.4 demonstrate that the features associated with the GSR have the ability to distinguish the emotional dimension of arousal, differentiating emotions with the highest excitement (joy and anger) from those with the lowest (pleasure and sadness). This is in line with previous research, which has shown that higher values are indicative of higher arousal [185] [186]. Of particular interest is the observation that joy, a positive-valence emotion, appears to elicit higher levels of excitement than anger. Indeed, the majority of GSR-related features exhibit substantially higher values for joy compared to the other emotions, as clearly illustrated in Figure 4.4, where the GSR derivative stands out in terms of height on the z-axis compared to the other three emotions.

Examining the data at the cardiac level, we can observe that the AVNN feature is highly effective in distinguishing joy and sadness in opposite ways. These two emotions are situated at opposite ends of the Russel's circumplex model of affect that is often used to describe emotions in terms of their valence and arousal dimensions. Specifically, joy, which is characterized by high arousal and positive valence, elicits the highest acceleration of heart rate, while sadness, with its low arousal and negative valence, is associated with the highest deceleration.

In terms of the respiratory signal, it appears to be particularly relevant in identifying anger,

which exhibits a much higher average rate than the other three emotions. Physiologically speaking, this study suggests the following results:

- Joy is distinguished by a greater sympathetic activation, resulting in much higher GSR values than the other emotions, as well as a higher heart acceleration.
- Anger, on the other hand, is well characterized by the frequency and amplitude of breath, which are much higher and much narrower, respectively, than the other emotions.
- Sadness is the emotion associated with the lowest responses in terms of respiratory frequency, heart acceleration, and sweating.
- Pleasure, which is very similar to sadness on a physiological level, is characterized by feature values that fall in the middle range when compared to the two high arousal emotions and sadness.

In terms of classification, the results have shown high performance due to both the ability of music to elicit emotions and the design of the protocol. In fact, a mean validation accuracy of 85% was achieved in the 4-class problem and over 90% in the binary problems related to the arousal and valence dimensions. Of particular interest is the valence dimension, which is usually more difficult to discriminate than arousal, but in this case, even higher performance was achieved for valence. Therefore, music seems to be an excellent stimulus for eliciting emotions. It is true that the study was conducted on a single subject, and achieving high classification performance with repeated measures from the same subject is much simpler, but the performance is still very high compared to the literature [187] [188].

Emotional protocol with visual, auditory, and combined stimul

An examination of physiological signals, including GSR, ECG, BVP, PUPIL, and EEG, indicates that sounds generally elicit stronger emotional responses than visual stimuli alone or in conjunction with sounds.

The analysis of GSR, a measure associated with sympathetic activation, indicated that listening to sounds caused higher sympathetic activity than watching images. This higher excitation was linked more to stimuli with positive emotional content.

Counterintuitively, the ECG-derived measures showed a heart rate deceleration during the listening to sounds, reflected in lower normalized power spectral density in the low frequency range and higher in the high frequency range. The BVP-derived measures instead showed increase values (i.e., a decrease in blood pressure), especially at high valence in IADS-only. The heart rate deceleration during the sound phase, as stated also for the listening effort experiment, support Lacey's theory of attention, which suggests that heart rate deceleration is accompanied by increased GSR in response to attention-requiring stimuli. The findings suggest that auditory stimuli are more effective in eliciting emotional responses compared to visual stimuli. Moreover, based on the feedback from the study participants, it seems that sounds alone are perceived as more unpredictable stimuli and capture greater attention from the subjects. This could be due to the lack of a visual counterpart

associated with the sound, which forces the subjects to rely more on their memory and cognitive processing to make sense of the stimulus, potentially resulting in stronger emotional responses. This hypothesis is also supported by the brain activity recorded by the EEG during the experiment. Although there were no significant differences in brain activity in the emotion-processing areas and frequency bands typically associated with emotions, the EEG-derived measures showed a more localized and stronger activation along the scalp during tests where only auditory stimuli (IADS-only) were presented. These findings suggest that auditory stimuli may elicit a more focused neural response in the brain compared to visual stimuli. Moreover, the subjective reports from the subjects indicated a higher level of attention during IADS-only tests, supporting the idea that auditory stimuli can capture and hold more attention than visual stimuli.

The pupillary signal results were challenging to interpret due to the lack of visual stimuli presented to the subjects during IADS-only. Nonetheless, the data revealed that the pupillary diameter was consistently lower, and the standard deviation of the diameter was lower as well, during IADS-only in comparison to the other phases. The difference in pupillary amplitude between IAPS-only and IAPS+IADS was not found to be significant, suggesting that sounds do not have a significant effect on pupil diameter. However, in some arousal sessions, the power spectral density of the pupillary signal was observed to be higher during IAPS-only in both low and high frequency ranges compared to the other phases. This suggests that visual stimuli alone may elicit a greater overall activation of the pupil compared to when combined with auditory stimuli. A detailed examination of the normalized power spectral densities revealed that the IADS-only condition exhibited higher values of DLFn and lower values of DHFn in the most arousal-inducing sessions (e.g. A3 and A4). This finding suggests that there may be a stronger sympathetic influence on the pupil when listening to sounds without any accompanying visual stimuli. Moreover, our analysis revealed a statistically significant disparity in the very high frequency range during IAPS+IADS in contrast to IAPS-only. However, it remains uncertain if this frequency range is associated with the autonomic nervous system, and it necessitates further inquiry.

From a respiratory perspective, a greater vagal activation appears to be present during the viewing of sole images. In the IAPS-only phase, indeed, the spectral power in the HF range is significantly higher compared to the other two phases at low valence. Additionally, the estimation of RSA at high valence shows a general lower trend during the IADS-only which is significant compared to the IAPS+IADS.

Overall, as shown in Figure 4.8, our findings demonstrate that listening to sounds alone (IADS-only) leads to sympathetic cardiovascular deactivation, accompanied by a rise in peripheral blood pressure, and heightened GSR peak amplitudes, particularly when compared to viewing images alone (IAPS-only).

These findings are in line with Lacey's attention theory and are also reflected in the higher attention indices recorded in the EEG during sound exposure. In summary, our results indicate that auditory stimuli have a greater physiological impact compared to visual stimuli or a combination of both.

Moreover, this study conducted a classification analysis to examine the generalizability of machine learning models in distinguishing different types of stimulation based on physiological signals.

The results showed that the models were able to achieve good classification performance, suggesting that machine learning can be a useful tool in analyzing physiological responses to various stimuli.

In support of the statistical analysis favoring auditory stimuli, our findings suggest that the presence of sound in the stimulation leads to higher classification performance compared to the absence of sound. Specifically, during the two test phases that included auditory stimuli, we observed higher classification performance. These results suggest that sound has a more significant impact on eliciting specific emotions and accurately distinguishing their effects on human physiology, compared to visual stimuli.

Emotions and virtual reality

The latest project pertains to the construction of emotional scenes in virtual reality. The aim was to generate a novel dataset for an emotion recognition experiment that utilizes physiological signals, which would be conducted in a virtual reality setting. The application of virtual reality to affective computing is still in its nascent stage. Few studies have concentrated on integrating it with rigorous quantitative methodologies, such as machine learning and advanced signal processing, to extract high-quality features. The implementation of a more immersive environment, in contrast to utilizing passive stimuli, could potentially enhance the assessment of the correlation between physiological signals and emotions. Additionally, it could enable a better comprehension of whether this type of environment is more efficacious in eliciting emotional responses in individuals.

The outcomes derived from the post-experience survey are in line with the anticipated scores in the circumplex plane for all emotions, with the exception of Sadness, which exhibits a minor inclination towards heightened arousal. On average, the perceived duration of the protocol is 8 minutes shorter than the factual duration, implying that virtual reality could potentially be utilized for lengthier protocols in contrast to solely-image datasets, such as IAPS.

From a physiological standpoint, the ECG-derived measures appear to be of limited relevance in discriminating the four emotions, as no statistically significant differences were found in the ECG features. Nevertheless, it is possible to observe a trend of growth in the sympathovagal balance index RR LFtoHF, which shows an increasing trend from the first two emotions in the protocol order with low arousal (i.e., sadness and relaxation) to those with high arousal (i.e., happiness and fear).

However, one of the most relevant features for emotion separation is VA, specifically the amplitude of the BVP signal. This feature was found to be significantly different in almost all comparisons except for the two low arousal emotions. In particular, fear is characterized by an increase in VA that completely sets it apart from all other emotions, indicative of sympathetic activation resulting in vasoconstriction.

The GSR signal, together with the BVP signal, is also very interesting, especially regarding the dimension of arousal but not limited to it. All GSR features show a clear separation concerning the dimension of arousal, being much higher and also significantly in the two high arousal emotions compared to the two low arousal emotions. The most interesting feature is the number of peaks, which is also capable of discriminating emotions in the valence dimension. Indeed, it can separate fear from happiness and relaxation from sadness uniquely. It is the only feature that can separate

the two low arousal emotions concerning the valence dimension.

At the respiratory level, sadness and fear are significantly separated by the amplitude of breaths, which is higher in sadness compared to fear. This is compensated by the respiratory rate, which instead shows an opposite trend.

From a physiological perspective, we can summarize the findings as follows:

- Fear is associated with higher sympathetic activation according to all signals, as evidenced by increased peripheral blood pressure identified by VA, higher GSR features, and increased respiratory rate.
- Happiness is in line with fear but is consistently one step below in terms of features' magnitude. Therefore, with respect to both BVP and GSR, it differentiates from low arousal emotions but not to the same extent as fear.
- Relaxation is characterized by lower feature amplitudes both in BVP and GSR, except for the average amplitude of the GSR signal which is higher in relaxation compared to sadness. Moreover, respiratory features are higher in relaxation compared to sadness, specifically respiratory rate and breath amplitude.
- Sadness falls in between the two high arousal emotions and relaxation. It is well characterized by the respiratory signal, particularly in terms of respiratory rate, which is the lowest among the four emotions and even has a negative median, indicating a decrease from the baseline.

From a classification standpoint, the obtained performances are very good both in the 4-class problem and when considering the individual dimensions of valence and arousal. This is particularly noteworthy since it concerns a user-independent problem without data leakage, given that in the test set entire subjects were kept out instead of random observations. One of the most notable findings pertains to the dimension of arousal. It is worth emphasizing that a test accuracy of 89% is remarkably high, especially in the context of user-independent emotion recognition studies involving multiple subjects. It appears, therefore, that the virtual reality environment holds promise for emotional recognition studies.

Considering both statistical analysis and machine learning, the results have shown that GSR and BVP signals are the most relevant in this specific context, as illustrated in Figure 4.11, 4.12 and 4.13 where the highest-performing features for distinguishing emotions are associated with these two signals. These findings can be linked to research on the physiology of emotions outlined in [33]. Both the BVP and GSR measurements were found to possess a high level of intrasubject variability associated with different emotional states.

5.2 Innovations

This thesis presents several key innovations. With regard to the listening effort protocol, it is one of the few studies to investigate the problem through the acquisition of six peripheral and central physiological signals. Additionally, a custom speech-in-noise test was developed specifically

to study this problem and reduce the occurrence of fatigue, which is often confused with listening effort. This test is much shorter than traditional speech-in-noise tests, taking only half the time to obtain the SRT, an index of auditory quality that could be used in clinical settings as an alternative to longer tests. In fact, in section 3.1.1, a comparison was made between the new algorithm and the traditional approach, using the same VCV stimuli, and the new approach proved to be more repeatable in terms of SRT calculated in two separate sessions on the same subjects. Furthermore, studies not reported in this thesis, but conducted outside the topic, have demonstrated the relevance of this new approach in predicting the pure-tone audiometry outcome, a clinical measure considered the gold standard in audiology for assessing auditory quality [2] [4] [6].

With the AuBT protocol, the main innovation lies in the fact that it has not been extensively explored in literature from the physiological point of view. Among the studies that have actually used it, different analysis has been conducted in terms of machine learning models, which still achieved comparable accuracy to that of this study, without delving into the actual relationship between emotions and physiological patterns [189] [190].

In relation to the IAPS and IADS protocol, this study stands out as one of the very few that uses both images and sounds from these widely validated datasets in conjunction [191]. Furthermore, few studies analyze the different physiological activation that occurs in response to various stimuli. The main innovation of this study lies in its use of validated emotional stimuli from the literature to compare different types of stimulation through an extensive range of both peripheral and central physiological signals. By doing so, this study is able to provide a comprehensive analysis of emotional responses.

The virtual reality protocol, is the first one, to our knowledge, that has a specially designed setup to elicit four specific emotions that span across all quadrants of Russell's circumplex model. In literature, there are primarily studies related to video games and virtual reality. One of the few study which examined emotions, physiological signals and classification [119], even if it was focused on stimuli related to four architectural environments, yielding lower recognition performance than ours - 75% in arousal and 71% in valence, compared to our 89% in arousal and consistently 71% in valence. The main innovation of our application lies in the creation of the dataset itself which is one of the first to join emotion recognition, VR and the acquisition of various physiological signals. Moreover, the innovative idea behind this study was to create real-life situations that can better elicit emotions by creating realistic scenarios. Overall, this study offers a unique and highly informative exploration of emotional responses within a virtual reality context, with a focus on creating an experience that is as close to real life as possible.

5.3 Limitations and future directions

Starting with the listening effort protocol, it should be noted that while the physiological measures were effective in discriminating between the two levels of effort, the analysis was successful for stimuli involving words immersed in Gaussian noise. It would be worthwhile to explore the application of the attention theory to different levels of difficulty, not just low and high effort, by testing with various types of noise and incorporating subjective measures alongside physiological markers. Additionally,

it would be interesting to include motivation, such as a reward system, in the protocol to consider this dimension, which has been shown in the literature to be crucial for studying this issue and validating physiological measures.

Regarding the AuBT protocol, the analysis of physiological signals proved to be crucial in distinguishing between emotions. However, it is important to consider that the same subject was used throughout the study. To validate music as an effective stimulus for eliciting various emotions, a similar protocol with multiple subjects should be developed.

For the IAPS and IADS protocol, the main limitation was that data acquisition was conducted in a laboratory environment, which may not induce natural emotions in the subjects. To address this issue, a second protocol was developed using virtual reality to immerse the subject without seeing the laboratory environment. Moreover, a noteworthy point is that the comparison between images and sounds was always made using matched images with relevant sounds, but not with videos, which may have been more arousing as they are more complex stimuli.

As for the VR protocol, the primary limitations are associated with the time-consuming nature of data acquisition, resulting in a restricted dataset. A larger subject pool would be useful to achieve better results. It would also be valuable to adopt more powerful machine learning approaches and to further investigate signals in the frequency domain, especially by extracting more features. Furthermore, since this preliminary study has yielded promising results, it would be helpful to standardize the stimuli with a greater number of subjective ratings and, above all, a psychological evaluation to strengthen the validity of the emotional scenes based on the protocol design. In addition, as numerous research studies have proven that the feeling of presence tends to grow with the accuracy of a replication or imitation of the physical world, therefore it could be compelling to gauge the sense of presence for each emotional scenario in the protocol. This could be accomplished by utilizing tools such as the ITC-Sense of Presence Inventory [192], which assesses various factors including sense of space, engagement, attentiveness, distractions, control and manipulation, authenticity, naturalness, and perception of time. Such an evaluation could be beneficial in comprehending how each scene is perceived and in improving it, as well as identifying the significant factors that contribute to the sense of presence. It would be also fascinating to divide the signals into sub-regions to investigate the specific effects of mapped events and potentially apply the algorithms in an online setting. In conclusion, due to the defined dataset and access to supplemental information, such as the precise timeline of events within the protocol, a diverse range of analyses remain possible.

Chapter 6

Conclusion

The results obtained from the different protocols in this study highlight the importance of analyzing physiological signals in investigating listening effort and emotional recognition. As it pertains to listening effort, it is clear that this is a complex problem that requires monitoring more than one signal to be explained from a physiological point of view. While many studies have attempted to identify a single physiological signal that can explain listening effort, the multidimensional nature of the problem is clearly demonstrated in this study, which uses various peripheral and central physiological signals to examine the issue.

The results of this study show that effort measures alone, such as increased sympathetic activation, are not always indicative of greater effort. This is evident in the behavior of the GSR-related measures, which usually serves as an indicator of sympathetic activation, but behaves in the opposite way in our case when it increases. This finding can be explained by the interpretation made in this thesis regarding the theory of attention. By acquiring various signals, we were able to understand that despite greater sympathetic activation under high effort, some signals may not agree with others. This is because the attention and engagement dimensions in the task cannot be underestimated.

What makes this study innovative is the use of multiple physiological signals to explore the complex nature of the problem. By employing innovative methodologies to acquire diverse signals, such as EEG and sympathetic response measures, this study provides a more comprehensive understanding of listening effort.

The study of emotions is undoubtedly a challenging task, especially when attempting to establish rules to explain the behavior of individuals with varying responses to different stimuli. The difficulty in standardizing the study of emotions is evident from the comparison of the AuBT and VR protocols. Although the elicited emotions are similar, except for Anger in AuBT and Fear in VR, both with high arousal and low valence, the physiological patterns are not completely different, but not the same either. While Sadness and relaxation are more easily distinguishable from high arousal emotions in terms of sympathetic response, the amplitude of features linked to GSR and BVP signals differ between the two protocols. In general, it is essential to consider that the observations may vary from protocol to protocol as the stimuli used are different.

Despite the variability in physiological responses observed across the different protocols and

stimuli, the study on the AuBT protocol, which elicited four separate emotions using musical stimuli, demonstrated how music can be an effective stimulus for this purpose. The high performance of the machine learning models employed using a limited number of relevant features supports this claim. The study also employed a modified cross-validation technique, which used leave-one-out logic to test the influence of different days on classification performance. This technique demonstrated the robustness of the analysis even over time, as it did not significantly impact the classification performance.

An ad hoc protocol was constructed using IAPS and IADS, which are rarely used in conjunction in the literature to study the different physiological activation patterns depending on the type of stimulus (visual, auditory, or combined). Thanks to the innovative design of the protocol, which combines different types of stimulation and the acquisition of many peripheral and central signals, it was possible to investigate diverse physiological activation patterns. The results showed that the auditory stimuli were the most effective in eliciting emotional responses, both sympathetic and attentive, and Lacey's theory of attention proved useful for comparing the different types of stimulation. The protocol's novelty lies in the integration of various types of stimuli and the acquisition of multiple signals, which allowed for a more comprehensive investigation of physiological activation patterns. Furthermore, this study contributes to the limited literature on the combined use of IAPS and IADS, highlighting their potential for future research.

In the realm of VR, the protocol created in our study represents a significant innovation. To our knowledge, it is the first protocol to create four different immersive environments to elicit four distinct emotions while simultaneously acquiring four peripheral physiological signals. The design of our study aimed to depart from traditional laboratory measurements, where subjects may be less likely to experience emotions due to the unrealistic environment. Instead, we created immersive environments that mimicked real-life situations. By acquiring physiological signals, we were able to investigate different physiological activation patterns depending on emotional stimuli, and our classification models achieved high accuracies, particularly in the arousal dimension when compared with the literature.

The success we achieved in characterizing different emotional scenarios suggests that the design choices made during the VR protocol development phase were valid and effective. In particular, delivering stimuli in a life-like scenario and combining visual and audio layers to enhance immersion were the most relevant and original ideas. Compared to traditional flat-screen methods, the VR stimuli were more effective in eliciting clearer emotions, as well as at the classification level, where higher accuracies were obtained in the VR protocol.

Our study also highlights the potential of non invasive physiological signals, such as GSR and BVP, to discriminate between stress and emotions effectively. In particular, we found that the BVP amplitude, which is not commonly calculated, is an excellent feature to separate high and low effort, high from low arousal emotions, and valence when arousal is high.

Ultimately, acquiring many physiological signals both peripheral and central has proved to be of great help in creating a more defined physiological picture in terms of both listening effort and emotion recognition together with the understanding physiological activation patterns according to different types of stimulation.

In conclusion, this study emphasizes the importance of using multiple physiological signals to investigate complex phenomena such as listening effort and emotional recognition. The multidimensional nature of the problems is highlighted by the different patterns of physiological activation observed across various protocols and stimuli. Innovative methodologies, including the use of VR and diverse stimuli, demonstrate the potential for more effective emotion elicitation and classification. Additionally, non-invasive physiological signals such as GSR and BVP prove to be valuable tools for discriminating between emotions and stress levels. Overall, this study provides a more comprehensive understanding of physiological responses to complex stimuli and lays the foundation for future research in this field.

Bibliography

[1] Edoardo Maria Polo, Marco Zanet, Marta Lenatti, Toon van Waterschoot, Riccardo Barbieri, and Alessia Paglialonga. Development and evaluation of a novel method for adult hearing screening: Towards a dedicated smartphone app. In *IoT Technologies for HealthCare: 7th EAI International Conference, HealthyIoT 2020, Viana do Castelo, Portugal, December 3, 2020, Proceedings 7*, pages 3–19. Springer, 2021.

[2] Edoardo Maria Polo, Marco Zanet, Alessia Paglialonga, and Riccardo Barbieri. Preliminary evaluation of a novel language independent speech-in-noise test for adult hearing screening. In *8th European Medical and Biological Engineering Conference: Proceedings of the EMBEC 2020, November 29–December 3, 2020 Portorož, Slovenia*, pages 976–983. Springer, 2021.

[3] Alessia Paglialonga, Edoardo Maria Polo, Marco Zanet, Giulia Rocco, Toon van Waterschoot, and Riccardo Barbieri. An automated speech-in-noise test for remote testing: Development and preliminary evaluation. *American Journal of Audiology*, 29(3S):564–576, 2020.

[4] Marco Zanet, Edoardo M Polo, Marta Lenatti, Toon van Waterschoot, Maurizio Mongelli, Riccardo Barbieri, and Alessia Paglialonga. Evaluation of a novel speech-in-noise test for hearing screening: Classification performance and transducers' characteristics. *IEEE Journal of Biomedical and Health Informatics*, 25(12):4300–4307, 2021.

[5] Edoardo Maria Polo, Maximiliano Mollura, Riccardo Barbieri, and Alessia Paglialonga. Multivariate classification of mild and moderate hearing loss using a speech-in-noise test for hearing screening at a distance. In *IoT Technologies for HealthCare: 9th EAI International Conference, HealthyIoT 2022, Braga, Portugal, November 16-18, 2022, Proceedings*, pages 81–92. Springer, 2023.

[6] Marta Lenatti, Pedro A Moreno-Sánchez, Edoardo M Polo, Maximiliano Mollura, Riccardo Barbieri, and Alessia Paglialonga. Evaluation of machine learning algorithms and explainability techniques to detect hearing loss from a speech-in-noise screening test. *American Journal of Audiology*, 31(3S):961–979, 2022.

[7] Edoardo Maria Polo, Andrea Farabbi, Maximiliano Mollura, Riccardo Barbieri, Alessia Paglialonga, and Luca Mainardi. Analysis of the skin conductance and pupil signals for evaluation of emotional elicitation by images and sounds. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1968–1971. IEEE, 2022.

[8] Andrea Farabbi, Edoardo M Polo, Riccardo Barbieri, and Luca Mainardi. Comparison of different emotion stimulation modalities: an eeg signal analysis. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3710–3713. IEEE, 2022.

[9] Edoardo Maria Polo, Maximiliano Mollura, Marta Lenatti, Marco Zanet, Alessia Paglialonga, and Riccardo Barbieri. Emotion recognition from multimodal physiological measurements based on an interpretable feature selection method. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 989–992. IEEE, 2021.

[10] Edoardo Maria Polo, Maximiliano Mollura, Marco Zanet, Marta Lenatti, Alessia Paglialonga, and Riccardo Barbieri. Analysis of the effect of emotion elicitation on the cardiovascular system. In *2021 Computing in Cardiology (CinC)*, volume 48, pages 1–4. IEEE, 2021.

[11] Jennifer Sorinas, Jose Manuel Ferrández, and Eduardo Fernandez. Brain and body emotional responses: Multimodal approximation for valence classification. *Sensors*, 20(1):313, 2020.

[12] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80, 2001.

[13] Margaret M Bradley and Peter J Lang. Emotion and motivation. 2007.

[14] Karin Roelofs. Freeze for action: neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1718):20160206, 2017.

[15] Patricia L Lockwood, Mathilde Hamonet, Samuel H Zhang, Anya Ratnavel, Florentine U Salmony, Masud Husain, and Matthew AJ Apps. Prosocial apathy for helping others when effort is required. *Nature human behaviour*, 1(7):1–10, 2017.

[16] Klaus R Scherer. What are emotions? and how can they be measured? *Social science information*, 44(4):695–729, 2005.

[17] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[18] Paul Ekman and Wallace V Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.

[19] Jonathan Posner, James A Russell, and Bradley S Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3):715–734, 2005.

[20] Albert Mehrabian. Framework for a comprehensive description and measurement of emotional states. *Genetic, social, and general psychology monographs*, 1995.

[21] John T Cacioppo, David J Klein, Gary G Berntson, and Elaine Hatfield. The psychophysiology of emotion. *New York: Guilford*, 1993.

[22] Robert Plutchik. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pages 3–33. Elsevier, 1980.

[23] Albert Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292, 1996.

[24] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1): 49–59, 1994.

[25] Marie-Christine Clugnet and Joseph E LeDoux. Synaptic plasticity in fear conditioning circuits: induction of ltp in the lateral nucleus of the amygdala by stimulation of the medial geniculate body. *Journal of Neuroscience*, 10(8):2818–2824, 1990.

[26] Bruce S Kapp, William F Supple Jr, and Paul J Whalen. Effects of electrical stimulation of the amygdaloid central nucleus on neocortical arousal in the rabbit. *Behavioral neuroscience*, 108(1):81, 1994.

[27] Peter J Lang and Margaret M Bradley. Emotion and the motivational brain. *Biological psychology*, 84(3):437–450, 2010.

[28] Yong-Sook Park, Francesco Sammartino, Nicole A Young, John Corrigan, Vibhor Krishna, and Ali R Rezai. Anatomic review of the ventral capsule/ventral striatum and the nucleus accumbens to guide target selection for deep brain stimulation for obsessive-compulsive disorder. *World neurosurgery*, 126:1–10, 2019.

[29] Wolfram Schultz. Neuronal reward and decision signals: from theories to data. *Physiological reviews*, 95(3):853–951, 2015.

[30] Amit Etkin, Tobias Egner, and Raffael Kalisch. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in cognitive sciences*, 15(2):85–93, 2011.

[31] Joshua A Waxenbaum, Vamsi Reddy, and Matthew Varacallo. Anatomy, autonomic nervous system. 2019.

[32] Phillip Low. Overview of the autonomic nervous system - neurologic disorders - merck manuals professional edition, 2022. URL `https://www.merckmanuals.com/professional/neurologic-disorders/autonomic-nervous-system/overview-of-the-autonomic-nervous-system`.

[33] Sylvia D Kreibig. Autonomic nervous system activity in emotion: A review. *Biological psychology*, 84(3):394–421, 2010.

[34] Maria Egger, Matthias Ley, and Sten Hanke. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, 343:35–55, 2019.

[35] Xin Hu, Jianwen Yu, Mengdi Song, Chun Yu, Fei Wang, Pei Sun, Daifa Wang, and Dan Zhang. Eeg correlates of ten positive emotions. *Frontiers in human neuroscience*, 11:26, 2017.

[36] Christos D Katsis, Nikolaos S Katertsidis, and Dimitrios I Fotiadis. An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders. *Biomedical Signal Processing and Control*, 6(3):261–268, 2011.

[37] Ali Azarbarzin, Michele Ostrowski, Patrick Hanly, and Magdy Younes. Relationship between arousal intensity and heart rate response to arousal. *Sleep*, 37(4):645–653, 2014.

[38] Margaret M Bradley and Peter J Lang. Affective reactions to acoustic stimuli. *Psychophysiology*, 37(2):204–215, 2000.

[39] Mimma Nardelli, Gaetano Valenza, Alberto Greco, Antonio Lanata, and Enzo Pasquale Scilingo. Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing*, 6(4):385–394, 2015.

[40] John T Lanzetta and Scott P Orr. Excitatory strength of expressive faces: Effects of happy and fear expressions and context on the extinction of a conditioned fear response. *Journal of Personality and Social Psychology*, 50(1):190, 1986.

[41] Robert W Levenson, Paul Ekman, Karl Heider, and Wallace V Friesen. Emotion and autonomic nervous system activity in the minangkabau of west sumatra. *Journal of personality and social psychology*, 62(6):972, 1992.

[42] Scott R Vrana. The psychophysiology of disgust: Differentiating negative emotional contexts with facial emg. *Psychophysiology*, 30(3):279–286, 1993.

[43] Paul Ekman, Robert W Levenson, and Wallace V Friesen. Autonomic nervous system activity distinguishes among emotions. *science*, 221(4616):1208–1210, 1983.

[44] Mimoun Ben Henia Wiem and Zied Lachiri. Emotion assessing using valence-arousal evaluation based on peripheral physiological signals and support vector machine. In *2016 4th International Conference on Control Engineering & Information Technology (CEIT)*, pages 1–5. IEEE, 2016.

[45] Qiang Zhang, Xianxiang Chen, Qingyuan Zhan, Ting Yang, and Shanhong Xia. Respiration-based emotion recognition with deep learning. *Computers in Industry*, 92:84–90, 2017.

[46] Jun-Wen Tan, Steffen Walter, Andreas Scheck, David Hrabal, Holger Hoffmann, Henrik Kessler, and Harald C Traue. Repeatability of facial electromyography (emg) activity over corrugator supercilii and zygomaticus major on differentiating various emotions. *Journal of Ambient Intelligence and Humanized Computing*, 3(1):3–10, 2012.

[47] Jeff T Larsen, Catherine J Norris, and John T Cacioppo. Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40(5):776–785, 2003.

[48] Khadidja Gouizi, Choubeila Maaoui, and Fethi Bereksi Reguig. Negative emotion detection using emg signal. In *2014 International Conference on Control, Decision and Information Technologies (CoDIT)*, pages 690–695. IEEE, 2014.

[49] Johannes Wagner, Jonghwa Kim, and Elisabeth André. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *2005 IEEE international conference on multimedia and expo*, pages 940–943. IEEE, 2005.

[50] Craig A Smith. Dimensions of appraisal and physiological response in emotion. *Journal of personality and social psychology*, 56(3):339, 1989.

[51] Jonghwa Kim and Elisabeth André. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12): 2067–2083, 2008.

[52] David John Baker. Review of karen burland & stephanie pitts (editors), coughing and clapping: Investigating the audience experience. oxford: Routledge, 2016. *Empirical Musicology Review*, 13(3-4):164–165, 2018.

[53] William B Davis and Michael H Thaut. The influence of preferred relaxing music on measures of state anxiety, relaxation, and physiological responses. *Journal of music therapy*, 26(4):168–187, 1989.

[54] Carol L Krumhansl. An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 51 (4):336, 1997.

[55] Chantal Martin-Soelch, Markus Stöcklin, Gerhard Dammann, Klaus Opwis, and Erich Seifritz. Anxiety trait modulates psychophysiological reactions, but not habituation processes related to affective auditory stimuli. *International journal of psychophysiology*, 61(2):87–97, 2006.

[56] Alexander L Francis and Jordan Love. Listening effort: Are we measuring cognition or affect, or both? *Wiley Interdisciplinary Reviews: Cognitive Science*, 11(1):e1514, 2020.

[57] Ronan McGarrigle, Kevin J Munro, Piers Dawes, Andrew J Stewart, David R Moore, Johanna G Barry, and Sygal Amitay. Listening effort and fatigue: What exactly are we measuring? a british society of audiology cognition in hearing special interest group 'white paper'. *International journal of audiology*, 53(7):433–445, 2014.

[58] Barbara Ohlenforst, Adriana A Zekveld, Thomas Lunner, Dorothea Wendt, Graham Naylor, Yang Wang, Niek J Versfeld, and Sophia E Kramer. Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, 351: 68–79, 2017.

[59] Jonathan E Peelle. Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and hearing*, 39(2):204, 2018.

[60] Benjamin WY Hornsby. The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and hearing*, 34(5):523–534, 2013.

[61] M Kathleen Pichora-Fuller, Sophia E Kramer, Mark A Eckert, Brent Edwards, Benjamin WY Hornsby, Larry E Humes, Ulrike Lemke, Thomas Lunner, Mohan Matthen, Carol L Mackersie, et al. Hearing impairment and cognitive energy: The framework for understanding effortful listening (fuel). *Ear and hearing*, 37:5S–27S, 2016.

[62] Emery Schubert. Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives. *Frontiers in psychology*, 4:837, 2013.

[63] Deborah Tannen. The pragmatics of cross-cultural communication. *Applied linguistics*, 5(3): 189–195, 1984.

[64] Margaret A Koeritzer, Chad S Rogers, Kristin J Van Engen, and Jonathan E Peelle. The impact of age, background noise, semantic ambiguity, and hearing loss on recognition memory for spoken sentences. *Journal of Speech, Language, and Hearing Research*, 61(3):740–751, 2018.

[65] Matthew B Winn, Dorothea Wendt, Thomas Koelewijn, and Stefanie E Kuchinsky. Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in hearing*, 22:2331216518800869, 2018.

[66] Jack W Brehm and Elizabeth A Self. The intensity of motivation. *Annual review of psychology*, 40(1):109–131, 1989.

[67] Benjamin WY Hornsby, Graham Naylor, and Fred H Bess. A taxonomy of fatigue concepts and their relation to hearing loss. *Ear and hearing*, 37(Suppl 1):136S, 2016.

[68] Erin M Picou, Todd A Ricketts, and Benjamin WY Hornsby. How hearing aids, background noise, and visual cues influence objective listening effort. *Ear and Hearing*, 34(5):e52–e64, 2013.

[69] Anthony Hogan, Kate O'Loughlin, Peta Miller, and Hal Kendig. The health impact of a hearing disability on older people in australia. *Journal of Aging and Health*, 21(8):1098–1111, 2009.

[70] Mary Rudner, Thomas Lunner, Thomas Behrens, Elisabet Sundewall Thorén, and Jerker Rönnberg. Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology*, 23(08):577–589, 2012.

[71] Alexander L Francis, Megan K MacPherson, Bharath Chandrasekaran, and Ann M Alvar. Autonomic nervous system responses during perception of masked speech may reflect constructs other than subjective listening effort. *Frontiers in psychology*, 7:263, 2016.

[72] Thomas Lunner, Mary Rudner, Tove Rosenbom, Jessica Ågren, and Elaine Hoi Ning Ng. Using speech recall in hearing aid fitting and outcome evaluation under ecological test conditions. *Ear and hearing*, 37:145S–154S, 2016.

[73] Adriana A Zekveld, Sophia E Kramer, and Joost M Festen. Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and hearing*, 32(4):498–510, 2011.

[74] Stuart Gatehouse and William Noble. The speech, spatial and qualities of hearing scale (ssq). *International journal of audiology*, 43(2):85–99, 2004.

[75] Thamar EM van Esch, Birger Kollmeier, Matthias Vormann, Johannes Lyzenga, Tammo Houtgast, Mathias Hällgren, Birgitta Larsby, Sheetal P Athalye, Mark E Lutman, and Wouter A Dreschler. Evaluation of the preliminary auditory profile test battery in an international multi-centre study. *International journal of audiology*, 52(5):305–321, 2013.

[76] Stuart Gatehouse and J Gordon. Response times to speech stimuli as measures of benefit from amplification. *British journal of audiology*, 24(1):63–68, 1990.

[77] Rolph Houben, Maaike van Doorn-Bierman, and Wouter A Dreschler. Using response time to speech as a measure for listening effort. *International journal of audiology*, 52(11):753–761, 2013.

[78] Birgitta Larsby, Mathias Hällgren, Björn Lyxell, and Stig Arlinger. Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects desempeño cognitivo y percepción del esfuerzo en tareas de procesamiento del lenguaje: Efectos de las diferentes condiciones de fondo en sujetos normales e hipoacúsicos. *International Journal of Audiology*, 44(3):131–143, 2005.

[79] Conor J Wild, Afiqah Yusuf, Daryl E Wilson, Jonathan E Peelle, Matthew H Davis, and Ingrid S Johnsrude. Effortful listening: the processing of degraded speech depends critically on attention. *Journal of Neuroscience*, 32(40):14010–14021, 2012.

[80] Jonas Obleser and Sonja A Kotz. Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage*, 55(2):713–723, 2011.

[81] Antoine Lutti, Joerg Stadler, Oliver Josephs, Christian Windischberger, Oliver Speck, Johannes Bernarding, Chloe Hutton, and Nikolaus Weiskopf. Robust and fast whole brain mapping of the rf transmit field b1 at 7t. *PloS one*, 7(3):e32379, 2012.

[82] Jonas Obleser, Malte Wöstmann, Nele Hellbernd, Anna Wilsch, and Burkhard Maess. Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience*, 32(36):12376–12383, 2012.

[83] Matthew G Wisniewski, Eric R Thompson, Nandini Iyer, Justin R Estepp, Max N Goder-Reiser, and Sarah C Sullivan. Frontal midline θ power as an index of listening effort. *Neuroreport*, 26(2):94–99, 2015.

[84] Carol L Mackersie and Heather Cones. Subjective and psychophysiological indexes of listening effort in a competing-talker task. *Journal of the American Academy of Audiology*, 22(02):113–122, 2011.

[85] Carol L Mackersie and Natalie Calderon-Moultrie. Autonomic nervous system reactivity during speech repetition tasks: Heart rate variability and skin conductance. *Ear and Hearing*, 37:118S–125S, 2016.

[86] Carol L Mackersie and Natalie Calderon-Moultrie. Autonomic nervous system reactivity during speech repetition tasks: Heart rate variability and skin conductance. *Ear and Hearing*, 37:118S–125S, 2016.

[87] Michael Richter. The moderating effect of success importance on the relationship between listening demand and listening effort. *Ear and Hearing*, 37:111S–117S, 2016.

[88] Thomas Koelewijn, Adriana A Zekveld, Joost M Festen, and Sophia E Kramer. Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, 33 (2):291–300, 2012.

[89] Adriana A Zekveld, Sophia E Kramer, and Joost M Festen. Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and hearing*, 32(4):498–510, 2011.

[90] J Wagner. Augsburg biosignal toolbox (aubt)–user guide. *University of Augsburg*, 2006.

[91] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, 23(10):1175–1191, 2001.

[92] Bruno Herbelin, Patrick Benzaki, Françoise Riquier, Olivier Renault, Helena Grillon, and Daniel Thalmann. Using physiological measures for emotional assessment: a computer-aided tool for cognitive and behavioural therapy. *International Journal on Disability and Human Development*, 4(4):269–278, 2005.

[93] Choubeila Maaoui and Alain Pruski. Emotion recognition through physiological signals for human-machine communication. *Cutting edge robotics*, 2010(317-332):11, 2010.

[94] Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. International affective picture system (iaps): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 1(39-58):3, 1997.

[95] Benedek Kurdi, Shayn Lozano, and Mahzarin R Banaji. Introducing the open affective standardized image set (oasis). *Behavior research methods*, 49(2):457–470, 2017.

[96] Ryan A Stevenson and Thomas W James. Affective auditory stimuli: Characterization of the international affective digitized sounds (iads) by discrete emotional categories. *Behavior research methods*, 40(1):315–321, 2008.

[97] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1): 18–31, 2011.

[98] Karan Sharma, Claudio Castellini, Egon L van den Broek, Alin Albu-Schaeffer, and Friedhelm Schwenker. A dataset of continuous affect annotations and physiological signals for emotion analysis. *Scientific data*, 6(1):1–13, 2019.

[99] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, 3(1):42–55, 2011.

[100] Christine Lætitia Lisetti and Fatma Nasoz. Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Advances in Signal Processing*, 2004(11):1–16, 2004.

[101] Lan Li and Ji-hua Chen. Emotion recognition using physiological signals from multiple subjects. In *2006 International Conference on Intelligent Information Hiding and Multimedia*, pages 355–358. IEEE, 2006.

[102] Eun-Hye Jang, Byoung-Jun Park, Sang-Hyeob Kim, and Jin-Hun Sohn. Emotion classification based on physiological signals induced by negative emotions: Discriminantion of negative emotions by machine learning algorithm. In *Proceedings of 2012 9th IEEE International Conference on Networking, Sensing and Control*, pages 283–288. IEEE, 2012.

[103] Andreas Haag, Silke Goronzy, Peter Schaich, and Jason Williams. Emotion recognition using bio-sensors: First steps towards an automatic system. In *Tutorial and research workshop on affective dialogue systems*, pages 36–48. Springer, 2004.

[104] Jacqueline M Kory and Sidney K D'Mello. 27 affect elicitation for affective computing. *The Oxford handbook of affective computing*, page 371, 2014.

[105] Rukshani Somarathna, Tomasz Bednarz, and Gelareh Mohammadi. Virtual reality for emotion elicitation–a review. *IEEE Transactions on Affective Computing*, 2022.

[106] Nazmi Sofian Suhaimi, Chrystalle Tan Bih Yuan, Jason Teo, and James Mountstephens. Modeling the affective space of 360 virtual reality videos based on arousal and valence for wearable eeg-based vr emotion classification. In *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, pages 167–172. IEEE, 2018.

[107] Joe Bardi. What is virtual reality: Definitions, devices, and examples. *Retrieved from Marxentlabs: https://www. marxentlabs. com/what-is-virtual-reality*, 2019.

[108] Giuseppe Riva. Virtual reality in psychotherapy. *Cyberpsychology & behavior*, 8(3):220–230, 2005.

[109] Paul MG Emmelkamp, Mary Bruynzeel, Leonie Drost, and Charles AP G van der Mast. Virtual reality treatment in acrophobia: a comparison with exposure in vivo. *CyberPsychology & Behavior*, 4(3):335–339, 2001.

[110] Azucena Garcia-Palacios, Hunter Hoffman, Albert Carlin, Thomas A Furness III, and Cristina Botella. Virtual reality in the treatment of spider phobia: a controlled study. *Behaviour research and therapy*, 40(9):983–993, 2002.

[111] Barbara Olasov Rothbaum, Larry Hodges, Samantha Smith, Jeong Hwan Lee, and Larry Price. A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of consulting and Clinical Psychology*, 68(6):1020, 2000.

[112] Barbara Olasov Rothbaum, Larry Hodges, Renato Alarcon, David Ready, Fran Shahar, Ken Graap, Jarrel Pair, Philip Hebert, Dave Gotz, Brian Wills, et al. Virtual reality exposure therapy for ptsd vietnam veterans: A case study. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, 12(2):263–271, 1999.

[113] Giuseppe Riva, Monica Bacchetta, Margherita Baruffi, Silvia Rinaldi, and Enrico Molinari. Experiential cognitive therapy: a vr based approach for the assessment and treatment of eating disorders. *Studies in health technology and informatics*, pages 120–135, 1998.

[114] Pietro Cipresso, Giovanni Albani, Silvia Serino, Elisa Pedroli, Federica Pallavicini, Alessandro Mauro, and Giuseppe Riva. Virtual multiple errands test (vmet): a virtual reality-based tool to detect early executive functions deficit in parkinson's disease. *Frontiers in behavioral neuroscience*, 8:405, 2014.

[115] Elisa Pedroli, Patrizia Padula, Andrea Guala, Maria Teresa Meardi, Giuseppe Riva, and Giovanni Albani. A psychometric tool for a virtual reality rehabilitation approach for dyslexia. *Computational and mathematical methods in medicine*, 2017, 2017.

[116] Ben Meuleman and David Rudrauf. Induction and profiling of strong multi-componential emotions in virtual reality. *IEEE Transactions on Affective Computing*, 12(1):189–202, 2018.

[117] Xiaolan Peng, Jin Huang, Alena Denisova, Hui Chen, Feng Tian, and Hongan Wang. A palette of deepened emotions: exploring emotional challenge in virtual reality games. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13, 2020.

[118] José Gutiérrez-Maldonado, Mar Rus-Calafell, and Joan González-Conde. Creation of a new set of dynamic virtual reality faces for the assessment and training of facial emotion recognition ability. *Virtual Reality*, 18(1):61–71, 2014.

[119] Javier Marín-Morales, Juan Luis Higuera-Trujillo, Alberto Greco, Jaime Guixeres, Carmen Llinares, Enzo Pasquale Scilingo, Mariano Alcañiz, and Gaetano Valenza. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific reports*, 8(1):1–15, 2018.

[120] Marco Granato, Davide Gadia, Dario Maggiorini, and Laura A Ripamonti. An empirical study of players' emotions in vr racing games based on a dataset of physiological data. *Multimedia Tools and Applications*, 79(45):33657–33686, 2020.

[121] Anna Felnhofer, Oswald D Kothgassner, Mareike Schmidt, Anna-Katharina Heinzle, Leon Beutl, Helmut Hlavacs, and Ilse Kryspin-Exner. Is virtual reality emotionally arousing? investigating five emotion inducing virtual park scenarios. *International journal of human-computer studies*, 82:48–56, 2015.

[122] Rosa M Baños, Cristina Botella, Isabel Rubió, Soledad Quero, Azucena García-Palacios, and Mariano Alcañiz. Presence and emotions in virtual environments: The influence of stereoscopy. *CyberPsychology & Behavior*, 11(1):1–8, 2008.

[123] Sven L Mattys, Joanna Brooks, and Martin Cooke. Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive psychology*, 59(3):203–243, 2009.

[124] Martin Cooke, Maria Luisa Garcia Lecumberri, Odette Scharenborg, and Wim A Van Dommelen. Language-independent processing in speech perception: Identification of english intervocalic consonants by speakers of eight european languages. *Speech Communication*, 52 (11-12):954–967, 2010.

[125] Alessia Paglialonga, Ferdinando Grandori, and Gabriella Tognola. Using the speech understanding in noise (sun) test for adult hearing screening. 2013.

[126] Alessia Paglialonga, Gabriella Tognola, and Ferdinando Grandori. A user-operated test of suprathreshold acuity in noise for adult hearing screening: The sun (speech understanding in noise) test. *Computers in biology and medicine*, 52:66–72, 2014.

[127] Nara Vaez, Liliane Desgualdo-Pereira, and Alessia Paglialonga. Development of a test of suprathreshold acuity in noise in brazilian portuguese: a new method for hearing screening and surveillance. *BioMed research international*, 2014, 2014.

[128] M Eberhard David, F Simons Gary, and D Fennig Charles. Ethnologue: Languages of the world. *Twenty-second edition Dallas, Texas: SIL International*, 2019.

[129] Denis Byrne, Harvey Dillon, Khanh Tran, Stig Arlinger, Keith Wilbraham, Robyn Cox, Bjorn Hagerman, Raymond Hetu, Joseph Kei, C Lui, et al. An international comparison of long-term average speech spectra. *The journal of the acoustical society of America*, 96(4):2108–2120, 1994.

[130] Monique CJ Leensen, Jan APM de Laat, Ad FM Snik, and Wouter A Dreschler. Speech-in-noise screening tests by internet, part 2: improving test sensitivity for noise-induced hearing loss. *International journal of audiology*, 50(11):835–848, 2011.

[131] Cas Smits, Theo S Kapteyn, and Tammo Houtgast. Development and validation of an automatic speech-in-noise screening test by telephone. *International journal of audiology*, 43(1): 15–28, 2004.

[132] Bernhard Treutwein. Adaptive psychophysical procedures. *Vision research*, 35(17):2503–2522, 1995.

[133] Cees H Taal, Richard C Hendriks, Richard Heusdens, and Jesper Jensen. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7):2125–2136, 2011.

[134] Jesper Jensen and Cees H Taal. Speech intelligibility prediction based on mutual information. *IEEE/ACM transactions on audio, speech, and language processing*, 22(2):430–440, 2014.

[135] Junfeng Li, Risheng Xia, Dongwen Ying, Yonghong Yan, and Masato Akagi. Investigation of objective measures for intelligibility prediction of noise-reduced speech for chinese, japanese, and english. *The Journal of the Acoustical Society of America*, 136(6):3301–3312, 2014.

[136] David J Finney. Probit analysis, cambridge university press. *Cambridge, UK*, 1971.

[137] HCCH Levitt. Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, 49(2B):467–477, 1971.

[138] Tom N Cornsweet. The staircase-method in psychophysics. *The American journal of psychology*, 75(3):485–491, 1962.

[139] Marjorie R Leek. Adaptive procedures in psychophysical research. *Perception & psychophysics*, 63(8):1279–1292, 2001.

[140] Stanley A Klein. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & psychophysics*, 63(8):1421–1455, 2001.

[141] Cas Smits. Comment on 'international collegium of rehabilitative audiology (icra) recommendations for the construction of multilingual speech tests', by akeroyd et al. *International Journal of Audiology*, 55(4):268–271, 2016.

[142] Sofie Jansen, Heleen Luts, Kirsten Carola Wagener, Bruno Frachet, and Jan Wouters. The french digit triplet test: A hearing screening tool for speech intelligibility in noise. *International journal of audiology*, 49(5):378–387, 2010.

[143] Georg v Békésy. A new audiometer. *Acta Oto-Laryngologica*, 35(5-6):411–422, 1947.

[144] Robert S Schlauch and Richard M Rose. Two-, three-, and four-interval forced-choice staircase procedures: Estimator bias and efficiency. *The Journal of the Acoustical Society of America*, 88(2):732–740, 1990.

[145] BR Shelton and I Scarrow. Two-alternative versus three-alternative procedures for threshold estimation. *Perception & Psychophysics*, 35(4):385–392, 1984.

[146] Miguel A Garcıa-Pérez. Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision research*, 38(12):1861–1881, 1998.

[147] JB MacQuuen. Some methods for classification and analysis of multivariate observation. In *Proceedings of the 5th Berkley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[148] Marco Zanet, Edoardo Maria Polo, Giulia Rocco, Alessia Paglialonga, and Riccardo Barbieri. Development and preliminary evaluation of a novel adaptive staircase procedure for automated speech-in-noise testing. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6991–6994. IEEE, 2019.

[149] MiM Taylor and C Douglas Creelman. Pest: Efficient estimates on probability functions. *The Journal of the Acoustical Society of America*, 41(4A):782–787, 1967.

[150] Cas Smits and Tammo Houtgast. Results from the dutch speech-in-noise screening test by telephone. *Ear and hearing*, 26(1):89–95, 2005.

[151] David McShefferty, William M Whitmer, and Michael A Akeroyd. The just-noticeable difference in speech-to-noise ratio. *Trends in hearing*, 19:2331216515572316, 2015.

[152] Barrie R Cassileth, Andrew J Vickers, and Lucanne A Magill. Music therapy for mood disturbance during hospitalization for autologous stem cell transplantation: a randomized controlled trial. *Cancer*, 98(12):2723–2729, 2003.

[153] Stefan Koelsch, Kristin Offermanns, and Peter Franzke. Music in the treatment of affective disorders: an exploratory investigation of a new method for music-therapeutic research. *Music Perception*, 27(4):307–316, 2010.

[154] Gaetano Valenza, Antonio Lanata, and Enzo Pasquale Scilingo. The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE transactions on affective computing*, 3(2): 237–249, 2011.

[155] Gaetano Valenza, Luca Citi, Antonio Lanatá, Enzo Pasquale Scilingo, and Riccardo Barbieri. Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics. *Scientific reports*, 4(1):1–13, 2014.

[156] Javier Marín-Morales, Juan Luis Higuera-Trujillo, Alberto Greco, Jaime Guixeres, Carmen Llinares, Claudio Gentili, Enzo Pasquale Scilingo, Mariano Alcañiz, and Gaetano Valenza. Real vs. immersive-virtual emotional experience: Analysis of psycho-physiological patterns in a free exploration of an art museum. *PloS one*, 14(10):e0223881, 2019.

[157] Andres Pinilla, Jaime Garcia, William Raffe, Jan-Niklas Voigt-Antons, Robert P Spang, and Sebastian Möller. Affective visualization in virtual reality: An integrative review. *Frontiers in Virtual Reality*, 2:630731, 2021.

[158] Felix Bolinski, Anne Etzelmüller, Nele AJ De Witte, Cecile van Beurden, Glen Debard, Bert Bonroy, Pim Cuijpers, Heleen Riper, and Annet Kleiboer. Physiological and self-reported arousal in virtual reality versus face-to-face emotional activation and cognitive restructuring in university students: A crossover experimental study using wearable monitoring. *Behaviour Research and Therapy*, 142:103877, 2021.

[159] Hooman Sedghamiz. Matlab implementation of pan tompkins ecg qrs detector. *Code Available at the File Exchange Site of MathWorks*, 2014.

[160] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

[161] Wilson Truccolo, Uri T Eden, Matthew R Fellows, John P Donoghue, and Emery N Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of neurophysiology*, 93(2):1074–1089, 2005.

[162] Garrett B Stanley, Kameshwar Poolla, and Ronald A Siegel. Threshold modeling of autonomic control of heart rate variability. *IEEE Transactions on Biomedical Engineering*, 47(9):1147–1153, 2000.

[163] Zhe Chen, Emery N Brown, and Riccardo Barbieri. Assessment of autonomic control and respiratory sinus arrhythmia using point process models of human heart beat dynamics. *IEEE Transactions on Biomedical Engineering*, 56(7):1791–1802, 2009.

[164] Zhe Chen, Patrick L Purdon, Grace Harrell, Eric T Pierce, John Walsh, Emery N Brown, and Riccardo Barbieri. Dynamic assessment of baroreflex control of heart rate during induction of propofol anesthesia using a point process method. *Annals of biomedical engineering*, 39: 260–276, 2011.

[165] Riccardo Barbieri, Eric C Matten, AbdulRasheed A Alabi, and Emery N Brown. A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability. *American Journal of Physiology-Heart and Circulatory Physiology*, 288(1):H424–H435, 2005.

[166] Michael F O'Rourke, Toshio Yaginuma, and Albert P Avolio. Physiological and pathophysiological implications of ventricular/vascular coupling. *Annals of biomedical engineering*, 12: 119–134, 1984.

[167] David Korpas, J Halek, and L Doležal. Parameters describing the pulse wave. *Physiological research*, 58(4), 2009.

[168] Gerardo Tusman, Cecilia M Acosta, Sven Pulletz, Stephan H Böhm, Adriana Scandurra, Jorge Martinez Arca, Matías Madorno, and Fernando Suarez Sipmann. Photoplethysmographic characterization of vascular tone mediated changes in arterial pressure: an observational study. *Journal of clinical monitoring and computing*, 33:815–824, 2019.

[169] Jorn Bakker, Mykola Pechenizkiy, and Natalia Sidorova. What's your current stress level? detection of stress patterns from gsr sensor data. In *2011 IEEE 11th international conference on data mining workshops*, pages 573–580. IEEE, 2011.

[170] Timo Partala and Veikko Surakka. Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, 59(1-2):185–198, 2003.

[171] Milton Pong and Albert F Fuchs. Characteristics of the pupillary light reflex in the macaque monkey: discharge patterns of pretectal neurons. *Journal of neurophysiology*, 84(2):964–974, 2000.

[172] Lawrence R Rabiner and Bernard Gold. Theory and application of digital signal processing. *Englewood Cliffs: Prentice-Hall*, 1975.

[173] Arnaud Delorme, Jason Palmer, Julie Onton, Robert Oostenveld, and Scott Makeig. Independent eeg sources are dipolar. *PloS one*, 7(2):e30135, 2012.

[174] Te-Won Lee, Mark Girolami, and Terrence J Sejnowski. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation*, 11(2):417–441, 1999.

[175] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, 2019.

[176] CM Cassani, S Coelli, A Calcagno, F Temporiti, S Mandaresu, R Gatti, M Galli, and AM Bianchi. Selecting a pre-processing pipeline for the analysis of eeg event-related rhythms modulation. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 4044–4047. IEEE, 2022.

[177] CM Cómez, M Vazquez, E Vaquero, D Lopez-Mendoza, and M^{a}J Cardoso. Frequency analysis of the eeg during spatial selective attention. *International Journal of Neuroscience*, 95(1-2): 17–32, 1998.

[178] Stefania Coelli, Roberta Sclocco, Riccardo Barbieri, Gianluigi Reni, Claudio Zucca, and Anna Maria Bianchi. Eeg-based index for engagement level monitoring during sustained attention. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1512–1515. IEEE, 2015.

[179] Roberto Sassi, Sergio Cerutti, Federico Lombardi, Marek Malik, Heikki V Huikuri, Chung-Kang Peng, Georg Schmidt, Yoshiharu Yamamoto, Document Reviewers:, Bulent Gorenek, et al. Advances in heart rate variability signal analysis: joint position statement by the e-cardiology esc working group and the european heart rhythm association co-endorsed by the asia pacific heart rhythm society. *Ep Europace*, 17(9):1341–1353, 2015.

[180] Barbara Ohlenforst, Dorothea Wendt, Sophia E Kramer, Graham Naylor, Adriana A Zekveld, and Thomas Lunner. Impact of snr, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing research*, 365:90–99, 2018.

[181] Chiara Visentin, Chiara Valzolgher, Matteo Pellegatti, Paola Potente, Francesco Pavani, and Nicola Prodi. A comparison of simultaneously-obtained measures of listening effort: Pupil dilation, verbal response time and self-rating. *International Journal of Audiology*, 61(7):561–573, 2022.

[182] Sara Alhanbali, Piers Dawes, Rebecca E Millman, and Kevin J Munro. Measures of listening effort are multidimensional. *Ear and Hearing*, 40(5):1084, 2019.

[183] J Andrew Taylor, Deborah L Carr, Christopher W Myers, and Dwain L Eckberg. Mechanisms underlying very-low-frequency rr-interval oscillations in humans. *Circulation*, 98(6):547–555, 1998.

[184] Gary G Berntson, J Thomas Bigger Jr, Dwain L Eckberg, Paul Grossman, Peter G Kaufmann, Marek Malik, Haikady N Nagaraja, Stephen W Porges, J Philip Saul, Peter H Stone, et al. Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology*, 34(6): 623–648, 1997.

[185] Walton T Roth, A Ehlers, C Barr Taylor, J Margraf, and WS Agras. Skin conductance habituation in panic disorder patients. *Biological psychiatry*, 27(11):1231–1243, 1990.

[186] JA Grey and N McNaughton. The neuropsychology of anxiety. an enquiry into the functions of the septo-hippocampal system, 2000.

[187] Cong Zong and Mohamed Chetouani. Hilbert-huang transform based physiological signals analysis for emotion recognition. In *2009 IEEE international symposium on signal processing and information technology (ISSPIT)*, pages 334–339. IEEE, 2009.

[188] Bo Cheng and Guangyuan Liu. Emotion recognition from surface emg signal using wavelet transform and neural network. In *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*, pages 1363–1366. IEEE, 2008.

[189] Florian Hönig, Johannes Wagner, Anton Batliner, and Elmar Nöth. Classification of user states with physiological signals: On-line generic features vs. specialized feature sets. In *2009 17th European Signal Processing Conference*, pages 2357–2361. IEEE, 2009.

[190] Ping Gong, Heather T Ma, and Yutong Wang. Emotion recognition based on the multiple physiological signals. In *2016 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pages 140–143. IEEE, 2016.

[191] Jenni Anttonen and Veikko Surakka. Emotions and heart rate while sitting on a chair. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 491–499, 2005.

[192] Jane Lessiter, Jonathan Freeman, Edmund Keogh, and Jules Davidoff. A cross-media presence questionnaire: The itc-sense of presence inventory. *Presence: Teleoperators & Virtual Environments*, 10(3):282–297, 2001.

Acknowledgements

This thesis is the result of great dedication and hard work, which wouldn't have been possible without the people by my side. I begin by thanking Professor Riccardo Barbieri, my supervisor who has been with me since my master's thesis and has become a friend over the years. Thank you, Professor, for your teachings and the trust you have placed in me on many occasions. Working with you has been a delight, and I look forward to the possibility of collaborating with you again in the future. A super special thanks goes to Dr. Alessia Paglialonga, who has always been available to help me with any problems. Thank you, Alessia, for the time you have dedicated to me over the years and for the teachings provided. You have been essential during my doctoral program, both professionally and emotionally. I really hope to work with you again in the future. I owe thanks to all my colleagues in the B3 Lab, who have accompanied me on this beautiful three-year journey, making it more enjoyable, fun, and light. A special thanks to my colleague Max, who has always been a source of inspiration and has helped me in every way during the three years. Last but not least, I thank my family, who has always been by my side and supported me in all the choices I have made and Davide, a sweet discovery that has given me great strength from the very beginning. Ad maiora semper.