UNIVERSITÀ
DEGLI STUDI
FIRENZE

Università degli Studi di Firenze
Dipartimento di Ingegneria dell'Informazione (DINFO)
Corso di Dottorato in Ingegneria dell'Informazione
Curriculum: Control, Optimization and Complex Systems

Toward Wearable Intelligence: Psychophysiological Responses and Digital Solutions to Climate Change

Candidate
Federico Calà

Supervisor
Prof. Antonio Lanatà

PhD Coordinator
Prof. Stefano Ricci

ciclo XXXVIII, 2022-2025

Università degli Studi di Firenze, Dipartimento di Ingegneria
dell'Informazione (DINFO).

Ai miei nonni

Acknowledgments

Thank you to Prof. Antonio Lanatà for believing in me, for his trust, guidance, and invaluable advice. His curiosity and commitment to these research topics have been a constant source of inspiration throughout this challenging yet rewarding journey.

Thank you to Lorenzo for your scientific guidance, and to Pietro for sharing conference adventures and our unforgettable time in Porto. You have been the best colleagues I could hope for. Thanks also to Sara and to all past and present members of the extended Biomedical Engineering Laboratory.

A heartfelt obrigado to Mariana and Maria, whose support and kindness made my time abroad not only possible, but meaningful. I could not have completed a fundamental part of this thesis without your help. Thanks to prof. António Coelho and all the participants that took part in my research activity in Porto.

To my closest friends - Chiara, Diletta, and Lorenzo - scattered across Italy and Europe, yet always by my side in spirit. Your friendship has been a safe shelter in every moment.

And finally, to my family. Grazie. Words fall short of what you really mean to me. Your love, trust, and dedication have been my first strength and my constant home. I owe everything to you, and I love you more than I could ever say.

Abstract

Global climate change poses profound threats to human psychophysiology. Digital health technologies and artificial intelligence (AI) offer unprecedented opportunities to monitor affective and cognitive mechanisms, support adaptive behaviours, improve the efficacy and reduce the carbon footprint of healthcare services delivery. This thesis has integrated wearable sensing, signal processing, and AI techniques to investigate human responses to climate stressors and to develop digital tools aimed at strengthening clinical practice and promoting sustainable health systems.

A virtual reality-based simulation of a wildfire was designed to address the climate action-value gap. This immersive exposure elicited robust psychophysiological engagement, evaluated with heart rate variability and electrodermal activity, and improved climate knowledge and pro-environmental intentions, particularly in younger male participants. To overcome inherent biases of such self-reported measures, implicit attitudes (i.e., automatic beliefs that are not accessible through conscious introspection) toward climate change were assessed with eye tracking related parameters. Although explicit and implicit attitudes were not correlated, nonlinear ocular dynamics revealed higher cognitive load when participants processed incongruent climate-related associations, suggesting autonomic markers of environmentally motivated decision conflict. This outcome underlines the possibility to better evaluate the efficacy of educational programs and campaigns.

This thesis then covered the development of digital health tools aligned with sustainability goals. Specifically, voice analysis was explored as a low-carbon, scalable biomarker for remote clinical care. AI-based diagnostic models, trained on sustained vowel and speech recordings, demonstrated the feasibility of distinguishing two types of voice disorders. Feature interpretability was also addressed, highlighting gender-specific markers to guide medical decision making. Following this outcome, one study automated post-treatment voice monitoring implementing an analogous strategy, whereas a further investigation adapted the concept of complexity matching in acoustic analysis and introduced a nonlinear, multivariate and multiscale framework to detect incomplete recovery of voice quality. These researches aimed at supporting patients follow-up, to identify individuals in need for additional care, and reducing unnecessary in-person visits. Finally, a smartphone-based voice

analysis was applied to genetic syndromes utterances, achieving promising classification performance and revealing distinctive articulatory and phonatory patterns to support their non-invasive screening.

Collectively, this work advanced multimodal sensing and explainable AI for understanding psychophysiological responses to climate stress and developing voice-based digital health solutions. It underlined how immersive experiences, implicit-cognition paradigms, and accessible technologies may jointly promote sustainable behaviours and healthcare, enhance self-awareness, and empower clinicians in a resource-efficient future.

Contents

Contents **vii**

List of Figures **xi**

List of Tables **xvii**

1 Introduction **1**
1.1 The objective 6

2 Wearable sensing and Psychophysiology monitoring **9**
2.1 Introduction 10
2.2 The nervous system 11
2.2.1 Brain activity measurement 16
2.3 The heart 18
2.3.1 Cardiac activity measurement 20
2.4 The skin 21
2.4.1 Electrodermal activity measurement 23
2.5 Ocular Dynamics 24
2.5.1 Eye movements 24
2.5.2 Pupil dilation 26
2.5.3 Blinking 28
2.5.4 Eye-tracking measurement 28
2.6 Voice and Speech 32
2.6.1 Acoustic measurements 36

3 Virtual Reality, Climate Awareness and Reaction **39**
3.1 Introduction 40
3.2 Materials and Methods 43

3.2.1 Participants . . . 43
3.2.2 Subjective measurement . . . 44
3.2.3 VR equipment and development . . . 46
3.2.4 Experimental procedure . . . 47
3.2.5 The VR experience . . . 48
3.2.6 Physiological data analysis . . . 54
3.2.7 Behavioural measures . . . 59
3.2.8 Statistical analysis . . . 61
3.3 Results . . . 62
3.4 Discussion . . . 72
3.5 Conclusion . . . 78

4 Physiological correlates of environmental implicit attitudes 81
4.1 Introduction . . . 82
4.2 Eye Tracking Technology in Studying Implicit Attitudes . . . 85
4.3 The current study . . . 87
4.4 Materials and Methods. . . . 88
4.4.1 Participants and Experimental Procedure . . . 88
4.4.2 Subjective Measures . . . 88
4.4.3 The Single-Category Implicit Association Test . . . 90
4.4.4 ETT preprocessing and feature extraction . . . 93
4.4.5 Statistical Analysis . . . 99
4.5 Results . . . 100
4.6 Discussion . . . 102
4.7 Conclusion . . . 106

5 Climate Change and Health: Voice Implications and Novel Diagnostic Techniques 107
5.1 Introduction . . . 108
5.2 Materials and Methods . . . 112
5.2.1 Patients . . . 112
5.2.2 Feature Extraction . . . 113
5.2.3 Unsupervised learning: Clustering . . . 114
5.2.4 Supervised learning: Machine Learning . . . 116
5.3 Results . . . 118
5.3.1 Clustering . . . 118
5.3.2 Machine Learning . . . 122
5.4 Discussion . . . 125

5.5 Conclusion 129

6 Climate Change and Health: Insight on Post-Treatment Voice Monitoring 131
6.1 Introduction 132
6.2 Study 1 134
6.2.1 Patients 134
6.2.2 Experimental Procedure 134
6.2.3 Results 135
6.2.4 Discussion 138
6.3 Study 2 140
6.3.1 Patients 142
6.3.2 Signal processing & Nonlinear analysis 142
6.3.3 Statistical analysis 148
6.3.4 Results 148
6.3.5 Discussion 152
6.4 Conclusion 155

7 Climate Change and Health: the Role of Smartphones 157
7.1 Introduction 158
7.2 Acoustic properties of genetic syndromes 159
7.2.1 Cri du chat syndrome 160
7.2.2 Crisponi syndrome 160
7.2.3 Down syndrome 161
7.2.4 Noonan syndrome 161
7.2.5 Costello syndrome 162
7.2.6 Smith-Magenis syndrome 162
7.3 Materials and Methods 163
7.3.1 Patients 163
7.3.2 Feature extraction 164
7.3.3 Statistical analysis & Machine learning 165
7.4 Results 166
7.5 Discussion 168
7.6 Conclusion 171

8 Conclusion 173
8.1 Summary of contribution 175
8.2 Directions for future work 176

A **179**

B **181**
B.1 Guided Tour Scripts 181
B.2 Radio Interview 185
B.2.1 The scripts 185
B.2.2 Images and Graphs 189

C **193**

Bibliography **195**

List of Figures

2.1 Schematic overview of cerebral superficial and deep lobes. . . 14
2.2 EEG measurement principles and technology. 18
2.3 Heart morphology, with a focus on the electrical conduction system. 19
2.4 ECG measurement principles and properties. 21
2.5 Comparative overview of sweat glands found in the human skin. 22
2.6 Overview on the dilation and constriction pathways of pupillary control. 26
2.7 Graphical visualisation of the bright- and dark-pupil methods for gaze tracking. 29
2.8 Wearable eye-trackers in form of glasses (Tobii Pro AB, Danderyd, Sweden). 31
2.9 Overview on voice production physiology. 35

3.1 Nationality distribution across participants. 43
3.2 VR familiarity score distribution. 44
3.3 Teleportation system visual cues to validate movements within the virtual scenario. 47
3.4 Panoramic snapshots taken within the virtual scenario. 49
3.5 Block diagram of the experience mechanics, alongside with the sequence of keyboard keys pressing and console log messages. 50
3.6 The green bubble indicating the position to reach in order to trigger the related audio of the guided tour. 51
3.7 Fire and smoke particle effects intensify with time to give the impression of a realistic wildfire progression. Meanwhile, the background skybox gets darker and the scenario fills with additional smoke. 52

3.8 View of the slideshow appearing over the stand with the radio atop of it. The expression "Incêndios florestais" stands for "Wildfires". 54

3.9 EDA preprocessing and decomposition pipeline. a) 3-steps preprocessing. The close-up highlights the low-pass filtering contribute. b) Phasic Component. c) SMNA Component. d) Tonic Component 56

3.10 ECG pipeline process. a) Identification of R peaks with Pan-Tompkin's algorithm. b) Clean HRV time series. 57

3.11 Distribution of IPQ individual items that presented significant differences according to demographic factors. A (•) represents a difference with $p < 0.05$, (••) $p < 0.01$. A (*) refers to a small Cliff's Δ effect sizes, whereas (**) to a medium one. . . 63

3.12 Distribution plots for statistically significant items Q2, Q4 and Q5 of the CCAS questionnaire. A (•) represents a difference with $p < 0.05$, (••) $p < 0.01$, and (• • •) $p < 0.001$. A (*) refers to a Cliff's Δ small effect. 64

3.13 Distribution plots for statistically significant items Q6, Q8 and Q12 of the CCAS questionnaire. A (•) represents a difference with $p < 0.05$, (••) $p < 0.01$, and (• • •) $p < 0.001$. A (*) refers to a Cliff's Δ small effect. 65

3.14 Distribution plots for statistically significant item Q15 of the CCAS questionnaire. A (•) represents a difference with $p < 0.05$. 66

3.15 NN50 across the baseline and the three stages of the VR experience. A (•) represents a difference with $p < 0.05$, (••) $p < 0.01$. 67

3.16 SCLmean across the baseline and the three stages of the VR experience. A (• • •) represents a difference with $p < 0.001$. . 68

3.17 SCRn across the baseline and the three stages of the VR experience. A (•) represents a difference with $p < 0.05$, and (• • •) $p < 0.001$. 69

3.18 EDAsymp across the baseline and the three stages of the VR experience. A (••) represents a difference with $p < 0.01$, and (• • •) $p < 0.001$. 69

3.19 Distribution of in-VR movements parameters considering demographical factors. A (•) represents a difference with $p < 0.05$, and (••) $p < 0.01$. A (*) refers to a Cohen's d small effect, whereas (**) to a medium effect. 70
3.20 Correlation analysis between IPQ items and in-VR movements parameters. A (•) represents a significant correlation with $p < 0.05$, whereas (••) a $p < 0.01$. 71
3.21 Relationship between the custom wildfire quiz score and the overall yaw rotation range during the audiovisual presentation occuring within the chapel. A (•) represents a significant correlation with $p < 0.05$. 71

4.1 Experimental pipeline. 88
4.2 Schematic representation of the custom web-based application deploying the SC-IAT. 91
4.3 Illustrative example of the preprocessing pipelines of gaze and pupil signals. Data from the whole experiment (T1, T2, T3 and T4, rest and calibration preliminary phases) are shown. . 93
4.4 Data distribution of pupil size sample entropy for the congruent and incongruent associations, considering target words only. A (•) represents a statistical difference $p < 0.05$, whereas (**) indicate a medium Cohen's d effect size. 101
4.5 Significant correlations between the D-score and RQA-derived eye tracking metrics. A (•) refers to a significant relationship with $p < 0.05$. 102
4.6 Correlations between D-score and blinking behaviour metrics. A (•) refers to a significant relationship with $p < 0.05$. 102

5.1 Graphical representation of the data distribution along the principal component axis in the original dataset (right) and the predicted clusters (left). 119
5.2 Barplot of the percentage of misassignments in each cluster found by the aware condition in the three pipelines. 121
5.3 Predictor Importance for the best ensemble models reported in Table 5.7 using Feature set /a/ both for F dataset (a) and M dataset (b). 123

6.1 Predictor Importance for the best ensemble models reported in Table 6.1 regarding the pre- vs post-treatment analysis using feature set /a/+/aiuole/ for the F dataset (a), and feature set /aiuole/ for the M dataset (b). 136

6.2 Distribution of G, R, and B index values across correctly classified and misclassified observations from Table 6.1 best classifiers for the F dataset. 137

6.3 Distribution of B index values across correctly classified and misclassified observations from Table 6.1 best classifiers for the M dataset. 137

6.4 The original time series is centred and integrated, obtaining the orange signal. 143

6.5 Trend computation in each segment across three different scales. A) s = 128, B) s = 256, C) s = 512. 143

6.6 Fluctuation functions for three different scales. The red horizontal line represents the mean of $F(n, s)$. 144

6.7 Fluctuation functions obtained considering a single scale (s = 256) and three different q-order moments (q = 1, q = 2 for reference to the B. panel of Figure 6.6 and q = 3). Since positive q are selected, only the largest variations are emphasised. . . 145

6.8 Hurst Exponent visualisation on a loglog graph for a set of different orders q and scales s. 146

6.9 Complexity Matching between pre- and post-operative audio recordings in FR (solid) and MR (broken) patient groups. LLR = longest range $2^7 - 2^{13}$, LR = long range $2^8 - 2^{13}$, SR = short range $2^9 - 2^{13}$, SSR = shortest range $2^{10} - 2^{13}$. . 149

6.10 Averaged correlation coefficient for each scale range between FR and MR groups. A (***) refers to a significant difference < 0.001. A (• • •) represents a Cohen's d large effect size. . . 150

6.11 Complexity Matching between pre- and post-operative audio recordings in MR patients (solid) and randomised order MR patients (broken). LLR = longest range $2^7 - 2^{13}$, LR = long range $2^8 - 2^{13}$, SR = short range $2^9 - 2^{13}$, SSR = shortest range $2^{10} - 2^{13}$. The abbreviation "rnd" stands for "randomised". . 150

6.12 Trends with standard error bars of nonlinear parameters for FR patients (left) and MR patients (right). Blue and orange lines refer to pre- and post-operative conditions, respectively. A (•) refers to a significant difference between before and after treatment, computed separately for FR and MR groups, < 0.05, (••) to $p < 0.01$, and (•••) to $p < 0.001$. A (*) represents a Cohen's d small effect size, (**) a medium one, (***) a large one. 151

7.1 ROC plots for the best classifier of PS, FA, and MA groups. . 168

B.1 Left panel: Conceptual model of relationships between fire parameters and their drivers. Right panel: Projected fire danger in late XXI century. 190
B.2 Burn hectares in Portugal in 2024. 190
B.3 Left side: Particulate matter and CO_2 emission. Right site: satellite image of 2024 Portugal wildfires. 191
B.4 Home hardening strategies. 191
B.5 Forest thinning techniques. 192
B.6 Campfire management common rules. 192

List of Tables

3.1 CCAS Items. 45
3.2 IPQ Items. 46
3.3 Overview on extracted EDA features 57
3.4 Overview of extracted HRV features 59
3.5 The list of features extracted from in-VR movement trajectories. 60
3.6 List of features extracted from IMU pitch, yaw and roll angles. 61
3.7 IPQ subscale median and IQR (in parentheses). OSP = Overall Sense of Presence; SP = Spacial presence; INV = Involvement; REAL = Experienced Realism. 62
3.8 CCAS differences within groups divided per item between the pre- and post-VR experience. A ↑ represents an increase in score value from pre- to post-VR, vice versa for ↓. A (•) represents a statistical difference with $p < 0.05$, (••) $p < 0.01$, and (•••) $p < 0.001$. PT = Portuguese; ENG = English; P = European Portuguese; I = International; Y = Younger; O = Older; M = Male; F = Female; E = Expert; NE = Not Expert. 66

4.1 BAGCC Items. 89
4.2 SC-IAT stimuli categorisation 90
4.3 BAGCC results. Median values and IQR (in parentheses). . . 100

5.1 Patients distribution divided by age, gender and voice pathology. The acronym y.o. stands for years old. 113
5.2 Acoustic parameters estimated by BioVoice. 114
5.3 Results of k = 2 clusters analysis on the males and females datasets. S = silhouette score, A = accuracy, σ = retained variance percentage. 120

5.4 Results of k = 4 clusters analysis on the females dataset. S = silhouette score, A = accuracy, σ = retained variance percentage. 120

5.5 Results of k = 2 clusters analysis on the males and females datasets, after retaining statistically significant features only. 120

5.6 Results of k = 4 clusters on females dataset, after retaining statistically significant features only. 121

5.7 Results of cross-validation (k-fold=10) for the experiment E1 regarding the binary discrimination between BLVF and UVFP (indicating in the table their True Positive Rate with TPR^{BLVF} and TPR^{UVFP}, respectively). In this experiment, all the subjects were considered without pre-processing according to age. Mean and standard deviation ($\mu \pm \sigma$) statistics regarding the validation performance are reported. The metric used for hyperparameter tuning is highlighted in bold, whereas the feature set leading to the best model is highlighted in grey. . . . 122

5.8 Statistical analysis results for acoustic parameters in E1. . . . 123

5.9 Results of cross-validation (k-fold=10) for the experiment E2 regarding the binary discrimination between BLVF and UVFP on a balanced version of F dataset according to age. Mean (μ) and standard deviation (σ) statistics regarding the validation performance are reported. The metric used for hyperparameter tuning is highlighted in bold, whereas the feature set leading to the best model is highlighted in grey. 124

5.10 Results of cross-validation (k=10) for the experiment E3, on the F dataset regarding the possibility to discriminate among the sub-classes representing BLVF: nodules (acronym=n), polyps (p) and cysts (c). For each class, the AUC values are reported. Mean (μ) and standard deviation (σ) statistics regarding the validation performance are reported. The metric used for hyperparameter tuning is highlighted in bold, whereas the feature set leading to the best model is highlighted in grey. . . . 125

6.1 Cross-validation results related to binary pre vs. post evaluation. In this case, the pre-class merges the observations for the two groups considered (BLVF and UVFP), and the post-class represents the recording related to the post-treatment analysis for both groups. Mean (μ) and standard deviation (σ) statistics regarding the validation performance are reported for F and M datasets. The metric used for hyperparameter tuning is highlighted in bold, whereas the feature set leading to the best model is highlighted in grey. 135
6.2 Statistical analysis results for acoustic parameters. 138

7.1 Mean and standard deviation of age of genetic syndrome patients divided per gender and age. In parentheses, the number of subject per group is reported. 163
7.2 Significant differences in acoustic features, divided per gender and age. A (•) refers to significance values, whereas a (*) to effect size. 166
7.3 Significant differences in acoustic features, divided per gender and age. A (•) refers to significance values, whereas a (*) to effect size. 167
7.4 Performance of the best classifiers for PS, FA and MA groups. Mean value ± std 167

C.1 RQA-related eye-tracking metrics statistical analysis results. The acronyms m.d. and GG stand for mean difference and Green-Gaussler correction, respectively. 193
C.2 Pupil-related metrics statistical analysis results. The acronyms m.d. and GG stand for mean difference and Green-Gaussler correction, respectively. 194

Chapter 1

Introduction

Global climate change represents one of the most profound and multifaceted challenges that humanity needs to confront. It endangers not only food security, water availability, and economic stability, but also the psychophysiological well-being and cognitive resilience of billions of people worldwide [1]. The increasing frequency and severity of extreme weather events, such as heatwaves, floods, and wildfires, are the most visible consequences of climate disruption that are already affecting lives and livelihoods, and are expected to intensify as global temperatures rise further [1, 2]. To limit catastrophic outcomes, major transformations across technological, economic, and behavioural domains are essential. Meeting international climate targets, such as limiting global warming to well below 2°C, will require unprecedented changes in energy systems, consumption patterns, and societal structures at all levels, from individual behaviour to global governance [2]. However, achieving such deep transitions is not solely a matter of technological innovation. Human behaviour, public perception, and psychological engagement are equally critical components to account for planning and communicating effective climate mitigation and adaptation strategies [1, 3, 4]. Despite growing scientific consensus and media attention, climate change is often perceived by the public as psychologically distant, which means that the risks climate change brings are still considered as abstract, uncertain, as well as temporally and geographically remote [3, 5]. This distancing can reduce concern and inhibit actions, a detrimental outcome that contributes to the widening of the gap between knowledge and behaviour [3, 6]. Nevertheless, several works demonstrated that specific conditions, such as local

climate variations and the occurrence of extraordinary events, can alleviate and modulate this psychological phenomenon (e.g., heavy snowfalls have induced people thinking about experiencing harsh winters even if temperatures were not lower than average) [7]. Moreover, they can promote citizen demands for more reliable pro-environmental policies and support political candidates [8].
To understand why such a psychological distance develops, it is important to acknowledge that the attitudes toward climate change are not formed in isolation. On the contrary, social norms, political ideologies, and group identities heavily influence what individuals believe and how they behave in relation to environmental issues [1, 4]. For instance, people may align their expressed beliefs with the dominant views of their political or social ingroup, even when these diverge from scientific evidence [9]. This means that self-reported attitudes might not always reflect true convictions. As a consequence, environmental psychology appealed to the concept of dual-process theory, in which human cognition is considered to operate on two separate levels: one that is deliberate, declarative, and analytical (explicit), and another that is automatic, spontaneous, and emotionally driven (implicit) [10]. Since climate change consists of a politically and socially salient issue, an insight over implicit attitudes is necessary because self-reports may be distorted by psychological constructs, e.g., social desirability [11]. As a matter of fact, implicit measures are less susceptible to such biases and can capture subconscious associations even when individuals are unaware or unwilling to disclose them [11]. Interestingly, it was also discovered that modifications in implicit attitudes may endure longer than explicit ones, making them a more stable target for behavioural interventions [12].
Despite these insights, much of the literature has not directly and exclusively focused on attitudes related to climate change. Hence, there is a need for comprehensive studies that implements a wider perspective on the matter, especially concerning affective responses and attribution tendencies, which are fundamental to shaping relevant actions. Addressing this gap is essential for developing a complete and realistic model of how individuals perceive and respond to the climate crisis, as well as uncovering possible technological solutions that may help reducing the psychological distance, promote awareness and guide individuals in adopting mitigation and adaptation strategies.

Understanding the impact of the climate crisis acquires an even more im-

portant meaning, considering how it has been increasingly recognised as a public health emergency [13–15].
From a mental health perspective, climate-related stressors, such as natural disasters, home damages and loss, have been associated with heightened risks of psychological disorders, including depression, post-traumatic stress, and even cognitive decline in vulnerable populations [1, 16]. As a clear solution to this problem still lacks, a continuous exposure to this particular chronic stress without providing adequate support or resources may have a long-lasting impact especially on younger generations, who indeed will bear a potentially unacceptably high burden in the future [17].
Causes and consequences of climate change have also a direct effect on human health, with both short- and long-term outcomes [13]. Burning fossil fuels to produce energy or move vehicles disperses in the atmosphere a set of hazardous pollutants such as sulphur dioxides or nitrogen oxides. These undergo complex chemical reactions and release the so-called particulate matter (PM). This term refers to microscopic solids or liquid droplets that, in some cases, are small enough to be inhaled (specifically, the PM2.5), leading to potential severe and diversified medical issues. For instance, a persistent exposure to PM2.5 represents one critical pathway that has been consistently associated with an increased incidence and mortality from ischaemic heart disease, stroke, heart failure, and arrhythmias [17, 18]. In fact, these pollutants trigger systemic inflammation, endothelial dysfunction, and oxidative stress, all of which contribute to the formation and destabilisation of atherosclerotic plaques [17, 18]. Moreover, ambient PM is the main risk factor for chronic obstructive pulmonary disease [19] and, concurrently, a potent asthma trigger [13, 19]. An unexpected, yet well-documented, consequence of particulate absorption is also the increased rate of metabolic disorders. It is assumed that pollutants may activate intracellular signalling pathways, resulting in epigenetic alterations [20], cause insulin resistance, favouring the predisposition of developing type 2 diabetes [21] and determine kidneys dysfunction [21, 22]. Industrial activity leads to another well-known detrimental phenomenon which is global warming, a rise in temperature that significantly disrupts weather patterns and ecosystems. Research highlights that heat and cold waves are responsible for more deaths than all other natural disasters combined, with people over 65 years old being at particular risk, especially those with multiple co-morbidities [23, 24]. Indeed, the primary mechanism with which humans dissipate heat, i.e., vasodilation to increase

the superficial blood flow, may be blunted or impeded in the elderly and in subjects with cardiovascular diseases and hypercholesterolaemia, posing these individuals at greater probabilities of occurring a heatstroke, a specific hyperthermic reaction that involves an inflammatory response followed by a potential multi-organ failure [25, 26]. Noteworthily, heatwaves can have disastrous repercussion over the environment as well. Indeed, they can facilitate the ignition and the spread of devastating wildfires, which in turn play a role in diffusing great amount of hazardous gases and PM2.5 [27]. On the other hand, cold snaps seem to exacerbate diabetes symptoms, determining a larger number of emergency department visits, possibly because of an impaired vasoconstriction or behavioural changes that impact glycaemic control (e.g., less physical activity or high caloric intake) [28]. Additionally, together with PM2.5 inhalation, cold snaps may accentuate the occurrence of cardiorespiratory problems [29].

Notably, and paradoxically, the healthcare framework itself contributes to the climate crisis, generating approximately 4-5% of global carbon emissions, making it the most carbon-intensive service sector worldwide [14,15]. Hospitals management, pharmaceuticals and medical technologies production are major contributors [14,15]. As such, there is an urgent imperative not only to protect health in the face of extreme climate events and guarantee services continuity, but also to decarbonise the systems designed to deliver those. Here, digital health technologies (DHTs) and health informatics play a critical dual role as mitigation and adaptation strategies [14,15,30]. By reducing unnecessary in-person visits through virtual care, optimising telehealth platforms, and deploying data-driven decision tools to reduce diagnostic and procedural overuse, DHTs can significantly lower the carbon footprint in the delivery of healthcare services [14, 15].

To work and to enhance their performance, equity, and accessibility, such tools rely on the acquisition of physiological signals, particularly through wearable technologies, to record them in more naturalistic scenarios. These data streams, ranging from heart rate variability to skin conductance, not only can enable and support a real-time, continuous, personalised health management but they also offer powerful insights into human perception, stress response, and cognitive involvement with the climate crisis. Moreover, wearable devices can implement advanced algorithms, also encompassing Artificial Intelligence (AI) techniques, to analyse multimodal information and provide guidance for clinical care decisions and allow the observation of

internal mental states with data that are less susceptible of deliberate manipulation.
Wearable devices represent a valuable bridge between individual-level monitoring and collective action, providing both clinicians and citizens with innovative instruments to track, respond to, and potentially mitigate the psychophysiological effects of climate change. The predictive power of wearable intelligent DHTs may enable the identification of subtle patterns and early warning signs that precede hazardous health conditions.

Through this approach, the current thesis proposes to use the concepts of complex system theory, i.e., the idea that a system is composed of many interacting parts such that the collective behaviour of those parts together is more than the sum of their individual ones, to study human attitudes and physiological activity, as well as their possible intertwining. Recent neuroscientific research highlighted that the independence assumption, in which it is presumed that brain areas work without accounting for contextual factors, such as the rest of the body or the surrounding environment, needs to be reconsidered in order to achieve more robust and reliable outcomes concerning brain-behaviour relationships [31]. Thus, mental events (e.g., attention, memory) emerge as a complex ensemble of non-linearly interacting signals coming not only from the brain, but, rather, the whole organism. The hypothesis of the body acting as a complex system was also supported by the discovery of other typical characteristics, e.g., spacial and temporal multiscale behaviour, modularity and many-to-one mappings [31, 32]. As a matter of fact, behind the deep influence neurons exert over each other's activity (even if not in close proximity), evidences suggested that heart rate and respiration are capable of modulating, for example, the functional connectivity between several cerebral regions [33]. In turn, heart electrical activity, despite manifesting intrinsic automaticity, responds to the signalling of the sympathetic and parasympathetic branches of the autonomic nervous system, increasing or decreasing its pacing to meet both conscious or unconscious body requirements [33]. Analogous considerations may be done when studying disease states, where complex system theory can be applied to analyse the interaction between communication pathways (e.g., electrical, chemical), heterarchical interrelations within the same organ or different ones, scale variability and emergence [34].

Therefore, it is important that signal processing pipelines and AI algorithms implemented in wearable intelligent systems are able to capture and describe the complex relationships in pathophysiological states to allow the development of personalised healthcare services. Moreover, these devices can serve as active and passive biofeedback systems to study and control features and mechanics of gamified activities and extended reality experiences. Through them, the latter approaches can actuate dynamic investigations over intricate social and psychological phenomena through a complementary, more precise and objective strategy.

1.1 The objective

The objective of this thesis was two-folded:

1. The analysis of climate change perception through educational experiences, as well as experimental paradigms tapping emotional and psychological involvement.

2. The assessment of adaptation and mitigation strategies to increase health self-awareness, improve the accessibility and resilience of healthcare services (especially in case of isolation following climate change related natural disasters), and reduce the latter's high burden and carbon footprint.

In both scenarios, wearable devices were deployed to obtain a broader perspective over the underlying spontaneous responses of the human body and mind to both climate change stressors and consequences.
To achieve the first objective, two main activities were carried out:

- The development of a virtual reality experience that was specifically designed to reduce both the temporal and physical distance to climate change so that its causes and consequences can be perceived as more urgent and evident. The activity provided a simulation of an increasingly frequent natural disaster, e.g., a wildfire, in a well-known viewpoint located in the northern region of the Porto District, Portugal. Moreover, it featured an interactive explanation of how wildfires originate, spread and can be handled for educational purposes to enhance learning, awareness and self-efficacy. Alongside subjective questionnaires exploring both climate change attitudes and the virtual reality degree

of immersion, physiological signals were acquired with two wearable devices to achieve a deeper and more objective understanding of affective engagement, cognitive mechanisms, and behavioural variables.

- The physiological quantification of climate change-focused implicit attitudes. To do so, a common psychological experimental paradigm, known as the Single-Category Implicit Association Test, was used to study and model the dual theory of human cognition. Data extracted from smart glasses were used to assess their reliability in comparison with traditional time-based measures and detect alterations in cognitive processes. Moreover, the modulating role of eye-tracking parameters over decision making time was explored.

For the second objective, one primary physiological signal, i.e., acoustical recordings of voice and speech, was acquired as its deep interconnection with the respiratory system poses vocal production at relevant risk due to air pollution great sensitivity. This activity was divided in two research branches, mainly depending on the recording device:

- Professional microphones were implemented to record highly disordered voices to get high-quality audio, that were later used to extract acoustic features, to discover potential vocal biomarkers that may be used as features for:

  1. Explainable AI experiments to develop automatic tools that not only may assist otolaryngologists in diagnosing vocal disorders, but also provide them a set of physiologically meaningful metrics that could help planning and organising customised treatments. Furthermore, the agreement with perceptual voice assessment was evaluated.
  2. Assessing the efficiency of surgical treatment by using a nonlinear, multivariate and multiscale framework.

- Smartphone-integrated microphones were used to record the voices of patients diagnosed with rare and ultra-rare genetic syndromes to define the so-called *phonotype*, i.e., to allow the preliminary vocal phenotyping of such diseases, and design tools, also based on AI techniques, that enabled an ubiquitous, contactless and cost-effective solution for screening purposes. Such an approach aimed at reducing additional

exams time and economic related issues, identifying most suitable therapies and improving healthcare systems efficiency.

Chapter 2

Wearable sensing and Psychophysiology monitoring

Wearable sensors are compact, non-invasive devices designed to continuously monitor physiological signals in real-world settings. They are characterised by portability, wireless connectivity, and minimal interference with natural, spontaneous behaviour. In psychophysiological research, wearable technologies enable the measurement of several signals, such as neural activity (EEG), cardiac dynamics (ECG), eye movements, and vocal features, supporting the investigation of cognitive and emotional processes in ecologically valid contexts. Adaptation strategies are critical for accurately capturing dynamic physiological responses outside controlled laboratory environments that also need to consider the specific human-machine interface and interaction. This chapter gives a brief overview of wearable sensors characteristics, specifically in relation to EEG, ECG, eye-tracking and voice measures. It also discusses the physiological origins of the biomedical signals that such devices acquire. [1]

[1]The sections examining eye-tracking and voice physiology and types of measurements has been published as "Wearable technologies for emotion monitoring: measuring physiological and behavioral signals" in *Unlocking the mind and emotions: cutting-edge bioengineering techniques from computational pshysiology to clinical applications* [35].

2.1 Introduction

A wearable sensor is a compact, body-mounted device designed to continuously or periodically monitor physiological, biomechanical, or environmental parameters related to a user's both physical and mental health, movements, or surroundings. These sensors are integrated into accessories, or even clothing, that are worn directly on the body (such as wristbands, smartwatches, glasses, or e-textiles) and capable of real-time data collection, processing, storage, as well as wireless transmission to external systems like smartphones, cloud servers, or medical databases. Wearable devices were traditionally developed and implemented for tracking physical exercise activity, but this has changed rapidly in recent years. Indeed, research is now focused on using them for tackling major challenges in healthcare applications [36]. Their potential utility has become more evident from the increasing rate of newly reported proof-of-concept studies of novel wearable devices. It is hypothesised that smart wearable sensors technology will revolutionise many aspects of human life and interaction in the same way that personal computers have done [37]. In the healthcare context, the fact that wearable devices can provide high accuracy, effective and diversified information about the users' psychophysiological status are enhancing both their reliability and commercial impact. A key feature of such instrumentation is its compact and lightweight structure that allows an ubiquitous and personalised monitoring due to the unmatched portability with respect to traditional clinical equipment. Wearable devices can measure several types of physiological signals, but their architecture presents important common traits:

- The sensing unit. This fundamental component is devoted to the detection of a specific physical, chemical, or biological element. Heart and respiration rate, body temperature are typical examples of tracked activities in physical exercises. On the other hand, biochemical sensors are equipped with specific receptors (e.g., enzymes or antibodies) that recognise a target analyte, for instance glucose concentration in blood. These latter devices commonly present an array of microneedles that puncture the skin to acquire the substance of interest. While their comfortability greatly improved over the years, they still are invasive devices and recent research is attempting to build sensors that can analyse secretions as sweat, saliva and tears to achieve a completely unobtrusive design [36, 38]. Interestingly, even when motion is not the

main objective of investigation, sensing units also integrate accelerometers and gyroscopes in the so-called Inertial Measurement Unit (IMU).

- The data acquisition system (DAQ). The DAQ serves as the bridge between the raw analogue signals captured by the sensors and the digital framework where they will be processed, analysed, and stored. The transducer is responsible for the transformation of the analogue signal into an electrical one. Conditioning circuits aim at several functions, e.g., noise removal and amplification. The analog-to-digital module converts the preprocessed signal into a digital one for further elaboration. Sampling frequency and resolution determine the quality of such conversion. Most DAQs feature a communication system to transmit data through wired or wireless (e.g., USB and Bluetooth, respectively) interfaces.

- The processing unit. It performs computations such as filtering and feature extraction. It may also implement artificial intelligence models, i.e., machine learning, that use the extracted parameters to carry out a classification task (e.g., for the automatic identification of a fall or an accelerated heart rhythm), enabling a wearable intelligence framework.

- The user interface. Modern wearable devices usually present a display where the output of the processing or AI pipeline may be visualised so that the user may be continuously aware about their own psychophysiological state.

Sections 2.2-2.6 discuss the physiological origins of biomedical signals as well as their acquisition modalities through wearable systems.

2.2 The nervous system

The nervous system can be divided into the central nervous system (CNS), composed of the brain and spinal cord, and the peripheral nervous system, consisting of the sensory and motor neurons that spread in whole body.
The brain orchestrates an extraordinary range of the body functions, from perception to memory, reasoning, and emotions. Structurally, it comprises the brainstem, cerebellum, diencephalon, and the two cerebral hemispheres that constitute the cerebrum (also known as telencephalon). It houses approximately 100 billion neurons, interconnected by extensive axonal networks

that enable rapid communication within and beyond the brain.
The brainstem, the most primitive brain region, connects the cerebrum with the spinal cord and oversees essential life functions (e.g., breathing, heart rate), as well as the communication with the rest of the body.
Posteriorly positioned to the brainstem, the cerebellum is critical for fine-tuning motor movements, maintaining posture, and ensuring balance. Recently, its role in contributing to cognitive and affective processing have been demonstrated [39].
The diencephalon is composed of three distinct structures known as thalamus, hypothalamus, and epithalamus. The thalamic nuclei act as a central relay station for sensory and motor information, processing and transmitting these signals from the body to the cerebral cortex and cerebellum for discrimination and interpretation. On the other hand, the hypothalamus is involved in the homeostasis, i.e., the delicate balancing of crucial body functions (e.g., temperature, blood pressure) to maintain the internal equilibrium. Finally, the epithalamus primarily connects with the so-called limbic system, a subcortical group of glands and structures that is deeply related with emotion processing and regulation.
The cerebrum, the largest brain structure, coordinates both sensory and motor functions, and supports higher-order cognitive processes such as vision, language, and decision-making. It is divided into two cerebral hemispheres. These are connected by the corpus callosum, a broad tract of white matter that facilitates interhemispheric communication. The cerebral cortex, a thin, highly folded layer of grey matter covering the cerebrum, is the site of consciousness and complex information processing. Despite its thinness (approximately 2-4mm), the cortex accounts for nearly 40% of brain mass due to its extensive surface area created by folds called gyri, separated by grooves known as sulci (if shallow) and fissures (if deeper). The most prominent fissures demarcate the brain's major lobes, which names derive from the skull bones they are placed under (except for the insula, see Figure 2.1 [40]):

- The frontal lobes form the anterior portion of both hemispheres, lying behind the frontal bone and superior to the orbits. It extends caudally from the forehead to the central sulcus. Inferiorly, it is demarcated by the lateral sulcus, which separates it from the temporal lobe. Functionally, the frontal lobe is central to higher-order cognitive processes and voluntary motor control. It supports abstract thinking, conscious thought, decision-making, planning, motivation, mood regulation, and

social judgment. It also plays a key role in declarative memory and speech production (involving the Broca's area). Another key area is the prefrontal cortex, a highly evolved region that is often referred to as the brain's "executive control centre" or "personality centre", as it integrates information coming from all other cortices.

- The parietal lobe occupies the upper portion of the brain beneath the parietal bone. It is positioned posterior to the central sulcus and extends caudally to the parieto-occipital sulcus. This area plays a central role in sensory integration, spacial orientation and processing. It contributes to visual perception by relaying optic radiations travelling toward the occipital cortex and is critical for constructing spacial maps of the environment.

- The temporal lobe occupies the lateral and inferior portions of each cerebral hemisphere. It is functionally diverse, playing a pivotal role in auditory processing, language comprehension, olfaction, emotion, and various forms of memory. It houses primary and association auditory cortices, as well as Wernicke's area, which is critical for the comprehension of spoken and written language. The medial temporal structures are essential for memory consolidation, the process of forming new long-term memories, and are particularly involved in encoding verbal and visual information.

- The occipital lobe consists of the of most posterior regions of the cerebrum. It is the smallest of the four cortical lobes but plays a critical role as the brain's primary visual processing centre. Indeed, information caught by the retina is transferred by the thalamus to the primary visual cortex. This particular area is responsible for the initial decoding of visual stimuli (e.g., edges, light intensity). Surrounding association areas interpret more complex features of the visual field, performing object recognition and motion tracking, and understanding spacial relationships.

- The insula is a small, folded region of cerebral cortex hidden beneath the lateral sulcus, deep to the frontal, parietal, and temporal regions. Despite its concealed location, the insula plays a central role in the integration of interoceptive signals, such as those related to pain, taste, and visceral sensation. Additionally, it contributes to emotional awareness, empathy, subjective feelings, and consciousness. The insula is

also implicated in autonomic regulation, including the modulation of cardiovascular responses during physical exertion and the maintenance of homeostatic balance. Its connectivity with limbic structures underscores its relevance in affective and motivational processes.

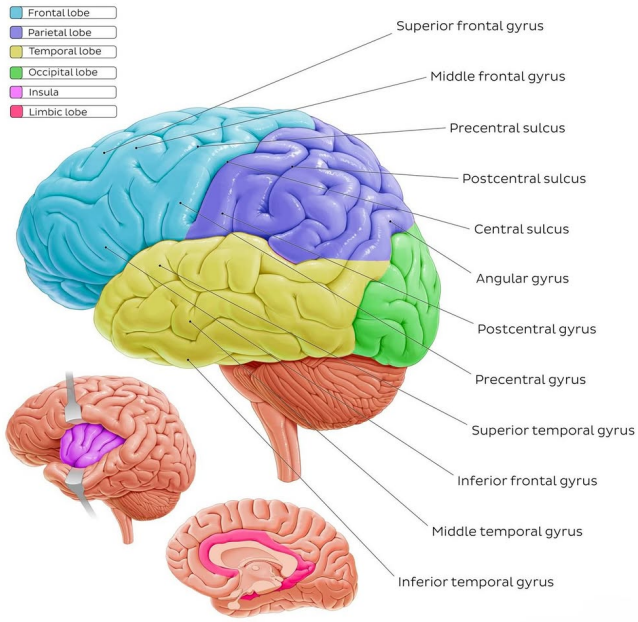


Figure 2.1: Schematic overview of cerebral superficial and deep lobes.

Besides the cortex, the cerebrum also houses the basal nuclei, i.e., a subcortical cluster of grey matter that span over the two hemispheres. The striatum is its largest component and it is composed of two pairs of overlapped structures: the nucleus accumbens and the olfactory tubercule are located in the lower layer, the caudate nucleus and the putamen in the upper one. On the one hand, this cluster coordinates decision-making and reward perception, on the other one, it is largely involved in motor planning and executive function, facilitating smooth voluntary movements [41]. Notably, the striatum receives input from several brain areas, especially the prefrontal cortex, but its output are forwarded to basal nuclei only.
Lateralisation characterises the cerebrum, meaning that certain cognitive

functions tend to be more dominant on one hemisphere than the other. Typically, the left one is specialised for language, analytical thinking, and fine motor control, while the right one is more involved in spacial processing, intuition, and non-verbal communication. Beneath the cortex stands the white matter, consisting of myelinated axons that connect cortical areas within and across hemispheres. This structural connectivity underpins the brain's ability to function as an integrated system, allowing distributed regions to cooperate in complex cognitive tasks. Indeed, although certain brain areas present highly specialised activities, most cognitive functions arise from interconnected neural networks rather than isolated regions. Thus, while distinct lobes and gyri can be anatomically identified, their contributions to perception, action, and thought are inherently collaborative and dynamic.

The brain can also be divided with a functional perspective, rather than merely anatomically. Indeed, the cerebrum is characterised by neural networks that are not localised to any certain regions, but interact over larger distances. Two examples are the limbic system and the reticular formation. The limbic system is a complex neural network that encompasses several cortical and subcortical areas located in the bordering zone between the diencephalon and telencephalon. It collectively handles emotions, motivation and implicit drives, as well as olfaction, thus bridging emotive and cognitive aspects of consciousness [42]. Its function relates to the preservation of the individual and the species, therefore, it has a relevant role in generating responses to challenging situations (e.g., triggering defensive reactions such as fear and faster heart rhythm), and in reproduction context. It also includes pleasure and aversion centres, which balance regulates reward and punishment mechanisms that reflect the neural substrate of motivation and emotional equilibrium. One of the most relevant structures of the limbic system is the amygdala, the primary centre for the processing of memory and emotion. Specifically, it performs the formation and storage of memories that are associated with affective experiences, typically fearful and stressful events. Indeed, the amygdala is deeply involved in the genesis of fearful responses, both behavioural and physical, directly affecting the ANS activity. The amygdala also participates in the modulation of memory consolidation and retention: in fact, emotional conditioning may amplify or dampen the strength of a memorised event itself [43]. The hippocampus, another component of the limbic system, plays a central role in the formation of episodic

and spacial memory. Moreover, it work synergistically with the amygdala to encode the emotional context of a situation.
The reticular formation consists of loose groupings of neurons in the white matter of the brainstem that are specialised in arousal regulation. Indeed, they can fire continued patterns of electrical impulses that affect cortex excitability and, consequently, alter consciousness [44]. At the same time, they filter sensory inputs, selecting the most relevant ones that need further consideration and processing. Moreover, the reticular formation helps controlling visceral motor functions, especially cardiac and respiratory rates.

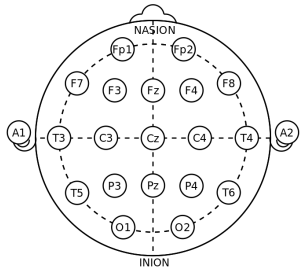
2.2.1 Brain activity measurement

A complete understanding of cognitive functions requires precise non-invasive measurements of brain activity. Electroencephalography (EEG) is a well-established technique for measuring neural activity, has excellent temporal resolution, and is relatively low-cost, especially when comparing it with functional magnetic resonance imaging. Noteworthily, EEG does not capture individual action potentials. Instead, the EEG signal primarily reflects large-scale voltage fluctuations resulting from synchronised post-synaptic activity of neuron clusters. It is believed that neocortical pyramidal neurons, due to their similar orientation and simultaneous firing, are the principal contributors to the EEG [45]. These excitatory cells are densely packed in the cortex and are characteristically aligned perpendicularly to the cortical surface, with cell bodies located in deeper layers and long apical dendrites extending toward the surface. This peculiar arrangement allows for the summation of numerous electrical potentials that creates a detectable field from the scalp. In contrast, subcortical regions contribute minimally to the EEG, largely due to their more irregular cellular organisation that does not favour open field generation.
The voltage differentials created by cortical pyramidal neurons are conducted through the head but due to the different conductive properties of the several layers between cortex and scalp (e.g., grey matter, dura, skull, and skin), the signal is attenuated and spacially blurred. Additionally, EEG recordings are vulnerable to various sources of noise, including muscle activity, skin potentials, and environmental electromagnetic interference. Consequently, high-quality data acquisition requires stringent control of the recording environment and participants' behaviour.
EEG recordings require a minimum of three electrodes: a ground electrode

(used for common mode rejection in case of power line interferences), a reference electrode, and an active electrode. In practical setups, electrodes are placed according to the international 10-20 standard (see Figure 2.2a) to guarantee consistent, repeatable, and anatomically meaningful acquisitions. This name refers to the percent distances between adjacent electrodes relatively to the total length of the skull (both front-back and left-right sides). A typical montage provides 21 electrodes, but extended systems also exist, characterised by up to 256 electrodes to ensure higher resolution. Regarding the type of electrode, the gold standard for clinical studies are the wet ones, that uses conductive gel or paste to lower the impedance between the two interfaces. Recently, dry-electrode technology, as the one implemented in the device shown in Figure 2.2b (Wearable Sensing, San Diego, CA, USA), have been proposed to overcome some well-known limitations of wet electrodes:

- The need for skin preparation (cleaning and, in some cases, abrasion).
- Uncomfortableness, especially in the long term when the conductive gel dries and may cause skin irritation.
- Longer setup time, less durability, more difficult maintenance.

On the other hand, dry electrodes are more suited for long-term wear, offer greater convenience, faster setup, and are ideal for portable and consumer-grade applications, though they often face challenges like higher contact impedance, greater sensitivity to noise and motion artefacts. Nevertheless, advancements in material design allowed to introduce dry electrodes with comparable signal quality metrics to wet ones [46–48].



(a) Electrode arrangement according to the 10-20 International Standard.

(b) A dry-electrode EEG recording helmet.

Figure 2.2: EEG measurement principles and technology.

2.3 The heart

The heart functions as a dual pump where the right side pumps the carbon dioxide-rich blood to the lungs for gas exchange, while the left one circulates the oxygen-rich blood to the body's organs and tissues. Each side is composed of an atrium and a ventricle that work in synchrony. Blood returning from the body enters the right atrium via the large veins, then flows through the tricuspid valve, which opens in response to pressure gradients, into the right ventricle. When the heart contracts and pressure rises, the pulmonary valve opens, allowing blood to be ejected into the pulmonary arteries and transported to the lungs. Once oxygenated, blood returns to the left atrium, passes through the mitral valve, and fills the left ventricle. A subsequent contraction forces the aortic valve to open, propelling fluid into the aorta, which then distributes it toward body's tissues through the arterial network. These rhythmic contractions are governed by a finely tuned conduction system that involves two main interactants: the cardiac tissue, especially the sinoatrial (SA) node located in the right atrium, and both the central and peripheral nervous systems. The SA node acts as the heart's primary pacemaker. The generation of action potentials is due to its unique ability to undergo a spontaneous gradual depolarisation. This automaticity initiates each heartbeat and sets the baseline heart rate. Once generated, this electrical impulse travels from the SA node through the atria, causing their contraction, and then reaches the atrioventricular (AV) node. Here, a slight delay is introduced due to a fibrous barrier of non-excitable cells, allowing the ventricles time to fill with blood before the next contraction. From

the AV node, the impulse descends rapidly through the bundle of His, then branches into two separate pathways, and finally spreads through the Purkinje fibres, leading to a coordinated contraction of the entire myocardium. Figure 2.3 [49] schematically depicts this process.

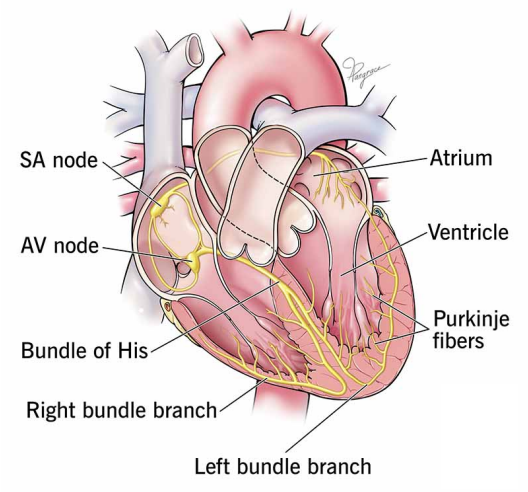


Figure 2.3: Heart morphology, with a focus on the electrical conduction system.

Although the heart can independently generate its own rhythm, its rate and force of contraction are modulated by the autonomic nervous system (ANS). This division of the CNS operates on internal organs, smooth muscle and glands, acting as a largely unconscious control system for bodily functions, such as respiration, digestion, vasomotor and cardiac regulation. The ANS distinguishes the sympathetic and parasympathetic nervous systems (SNS and PNS, respectively). The SNS increases heart rate and contractility by enhancing the rate of depolarisation in the SA node and the excitability of the myocardium. In contrast, the PNS decreases heart rate by hyperpolarising the cells in the SA node through the vagus nerve, and, consequently, slows the conduction toward the AV node. Notably, these functions are performed without any conscious effort and synergistically with each other.
Moreover, the CNS also plays an integrative role in cardiovascular regulation. Indeed, brainstem nuclei receive signals from cortical and limbic areas that influence heart function, particularly during stress, attention, or emo-

tional arousal. For instance, in fearful condition, the amygdala activates a set of physiological responses known as fight-or-flight through the SNS in order to adequately face a threat (e.g., increased heart beating and blood pressure) [50]. In turn, the prefrontal cortex exerts a tonic control over the latter either dampening or amplifying its influence on the heart rhythm [51].

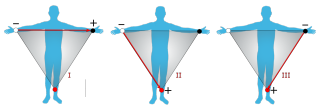
2.3.1 Cardiac activity measurement

The electrical activity generated and propagated within the heart during each cardiac cycle creates measurable voltage changes that can be detected on the surface of the body. These extracellular potentials form the basis of the electrocardiogram (ECG), a non-invasive method for assessing the timing and pattern of cardiac electrical activity. The electrodes are positioned at the vertices of the so-called Einthoven's triangle (depicted in Figure 2.4a), a conceptual figure formed by the extremities of three limbs:

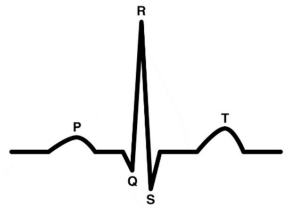
- The first lead, which measures the potential difference between right and left arm (LA-RA derivation).
- The second lead, between left leg and right arm (LL-RA derivation)
- The third lead, between left leg and left arm (LL-LA derivation).

The typical waveform resulting from an ECG consists in a set of components that corresponds to excitatory processes in the heart. As Figure 2.4b shows, the most prominent one is the QRS complex, i.e., three waves that occur in rapid succession reflecting ventricles contraction. This element is particularly important to derive and investigate the heart rate variability (HRV). Indeed, this phenomenon is analysed by identifying and computing the difference in duration of the temporal intervals between consecutive R peaks, that are the upward deflections of the QRS complex. Notably, a healthy heart owns a random variability, which is associated with a balanced control of the SNS and PNS, in order to enhance the adaptation to unexpected changes in the environment [52]. The HRV has been consistently considered as a robust, physiological index to characterise the degree of flexibility of a healthy heart activity [53]. However, research suggested that HRV can also be proposed as a metric with broader meaning and capabilities, also encompassing the evaluation of psychological states. Indeed, it may be implemented to monitor the brain's adaptive regulation over the periphery, thus providing useful

information regarding the overall capacity of the organism to function in a complex environment [53].



(a) The Einthoven Triangle and the three leads.



(b) A schematic representation of an ECG waveform highlighting the P wave, the QRS complex and the T wave.

Figure 2.4: ECG measurement principles and properties.

2.4 The skin

The skin is humans' largest organ and it represents the principal interface between the body and the surrounding environment. It acts as a selective barrier that simultaneously impede the entry of foreign, potentially dangerous matter and facilitate the passage of specific substances from the bloodstream to the body's exterior. It also regulates and controls tissues hydration and the core body temperature by balancing the rate of vasoconstriction or dilation, as well as sweating. To accomplish these tasks, the skin is deeply innervated and constantly receives signals from central and peripheral nervous systems. Extending between the epidermis and lower dermis layers, sweat glands are found (Figure 2.5 [54]).

They can be divided into two category considering their function and location:

- Apocrine glands are primarily located in the armpit and genital areas.
- Eccrine glands are spread over the whole body but their density becomes larger in palmar and plantar regions (approximately 400/mm^2). They are composed of a coiled, tubular structure that corresponds to the excretory portion and a relatively straight duct that reaches and opens to the surface of the skin as a small pore.

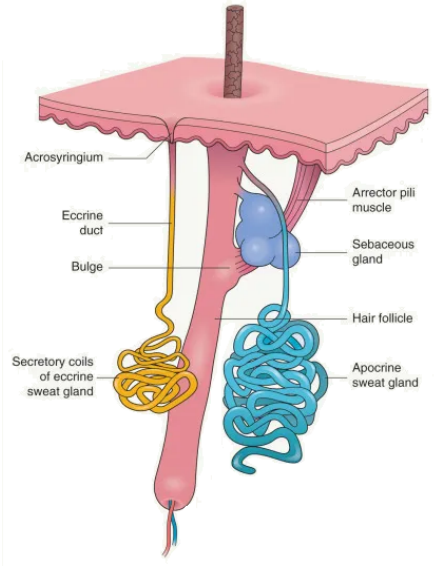


Figure 2.5: Comparative overview of sweat glands found in the human skin.

While the precise functions of the first ones is still under debate [55], the second ones are primarily responsible for thermoregulation. However, it has been hypothesised that palmar and plantar sweat glands are involved in more than just evaporative cooling. Indeed, research suggested that they are more responsive to emotional rather than thermal stimuli. When they are activated, these glands produce sweat, a fluid rich in water, electrolyte and mucin, that fills a number of ducts and up to a certain height depending on the nature and strength of the stimulus. These tubes can be considered as a set of parallel, variable resistors that alter the overall electrical properties of the skin: specifically, the higher the sweat rises, the lower the resistance become. Such a change in the skin's electrical properties is collectively referred as electrodermal activity (EDA). Differently from other physiological systems, eccrine sweat glands are solely innervated by sudomotor fibres originating in the sympathetic branch of the ANS. The SNS receives and integrates the inputs coming from several cortical and subcortical regions of the brain, so that the neural mechanisms involved in the central control of EDA are numerous and complex. Three possible pathways can be distinguished:

- A first level encompasses the influences from the hypothalamus and

the limbic system, with the amygdala generating excitatory and the hippocampus generating inhibitory inputs. Hypothalamic activation primarily relates with thermoregulation mechanisms as such component is considered the control centre of vegetative functions. On the other hand, a limbic activation may reflect affective processes due to its connection with emotional processing.

- A second level involves the activity of both premotor and frontal cortex. In this case, cognitive tasks such as motor planning and preparation, as well as orienting and attention can modulate EDA.

- A third level derive from the activation of the reticular formation. Brainstem-evoked EDA alteration are likely associated with movements and increased muscle tone.

EDA can be decomposed into two distinct components. The tonic component refers to the slow-changing skin conductance level (SCL), which reflects baseline arousal and can drift over time due to changes in general autonomic tone or environmental conditions. The phasic component, typically referred to as the skin conductance response (SCR), is a transient, stimulus-locked change that reflects discrete bursts of sympathetic activity in response to salient events. These include emotionally charged or attention-grabbing stimuli such as loud noises, threatening images, or increasing cognitive load. The typical SCR waveform consists of a rapid rise in conductance followed by a slower return to baseline.

2.4.1 Electrodermal activity measurement

EDA is typically measured by assessing the skin's conductance or resistance to a small, non-invasive electrical current. The most common and reliable configuration involves placing two electrodes on the medial phalanges of the index and middle fingers of the non-dominant hand. This placement is preferred because it minimises movement artefacts and ensures greater consistency in sympathetic response recording, as the dominant hand is more likely to be involved in voluntary motion during tasks and it may also present more surface alterations and thickening due to greater usage [56]. The electrodes are attached using adhesive collars and filled with an isotonic electrolyte gel to enhance signal quality and reduce impedance at the skin-electrode interface. During recording, a constant low voltage (typically around 0.5V) is

applied between the electrodes, and changes in conductance (measured in μS) are continuously monitored.

2.5 Ocular Dynamics

The human eye is a highly specialised sensory organ that enables vision by detecting and processing light. Structurally, it works similarly to a camera: light enters in the cornea, passes through the pupil, and is focused by the lens onto the retina, where photoreceptor cells convert it into electrical signals. These are then transmitted to the brain via the optic nerve. The extrinsic ocular muscles allows the eye to execute precise movements to explore the surrounding environment. Through their coordinated activity, the eyes can rapidly shift focus, track moving objects, and stabilise visual input via gaze shifts. As they determine what we see, attend to, and remember about our surroundings, they are central to executive functions and reflect their possible alterations due to affective and cognitive load.
Notably, eye movements are not solely responsible for directing gaze. Indeed, once it has been locked to a point of interest, other types of movements occurs in order to continuously provide an optimised image to the brain. For instance, the pupil diameter changes to regulate the flow of light within the eye, the curvature of the lens modifies the focus, and even blinking patterns serve to protect the eye and keep it moisturised.
This diversified set of movements provides a valuable insight into underlying neurological and cognitive states.

2.5.1 Eye movements

An eye movement is defined as any shift of its position in the orbit, which is finely controlled by six extraocular muscles. Their particular geometry and insertion location relatively to the eyeball allow this organ to accomplish a wide set of movements that span in all three dimensions. The main aim of eye movements is to enhance visual acuity by keeping images stable on the retina to reduce blur, to guide the fovea, its highest acuity portion, on the targets of interest, and to align both eyes at different focal plane depths, guaranteeing their fixation on the same object [57]. Several cortical and subcortical brain structures are involved in eye movements. The brainstem is largely responsible for the control of both horizontal and vertical gaze, as

the motoneurons of the extraocular muscles are contained in the pons and in the midbrain, respectively. The cerebellum contributes as well to the control of eye movements, especially when they need to fixate on moving targets, in eccentric gaze holding, and when coordinating the necessary adjustments to guarantee clear vision during head motion. Particularly, the vestibulo-ocular and optokinetic reflexes are two types of compensatory eye movements that have evolved to support vision during locomotion so that images do not slip across the retina with each head or body movement. Finally, the primary visual cortex houses the frontal eye field (FEF), for modulating rapid eye movements and their suppression during steady observation, the supplementary eye fields (SEFs), for planning movements sequences in responses to instructional changes, and the dorsolateral prefrontal cortex, important for memory-guided movements [58].
Eye movements can be divided into four types:

- Saccades are rapid, ballistic jumps that shift gaze from one location to another, in order to bring an object of interest to the fovea. Saccades can be elicited involuntarily, as a response to a sudden, unexpected change in the environment (e.g., a flash), and voluntarily, which are more complex and intertwined with the activity of the frontal cortex. During saccades, the information flow into the visual system is reduced.

- A fixation represents the act of maintaining the gaze steadily on a single location or object, allowing detailed visual processing. Ideally, during a fixation, no movement of the eye in its orbit occurs. In fact, in physiological conditions, the eyes are never immobile and small, rapid movements know as microsaccades actually take place. While their physiology is still much debated, they are thought to reduce the fading of the retinal image [59]. Therefore, fixation can only be defined as a period of relative stability between saccades. However, it remains the primary mean humans have for acquiring new information and make it available to the brain.

- Smooth pursuit is activated to track a moving target, in order to keep it centred in the visual field. Functionally, it resembles fixations, in the sense that it maintains the gaze steadily on the object, and mechanically a saccade, as indeed it consists of slow, continuous, voluntary rotation of the eyes.

- Vergence is the only disjunctive eye movement type where the two

eyes move simultaneously in opposite directions to align on objects at different depths, leading to convergence for near objects and divergence to focus far ones.

2.5.2 Pupil dilation

The pupil is a circular opening located at the centre of the eye, whose diameter typically ranges between 2 and 8mm depending on lighting conditions, arousal state, and cognitive load. Light entering the eye passes through the lens via this aperture and is subsequently focused onto the retina, enabling visual perception. Surrounding the pupil is the iris, the pigmented portion of the eye, which width is governed by two antagonistic muscles: the *iris sphincter* and the *iris dilator*. These regulate the diameter of the pupil through complex neuromuscular coordination, making the pupil a dynamic interface between the external world and internal brain states.
Pupillary control is mediated by both the sympathetic and parasympathetic divisions of the autonomic nervous system [60], as Figure 2.6 [61] schematically displays.

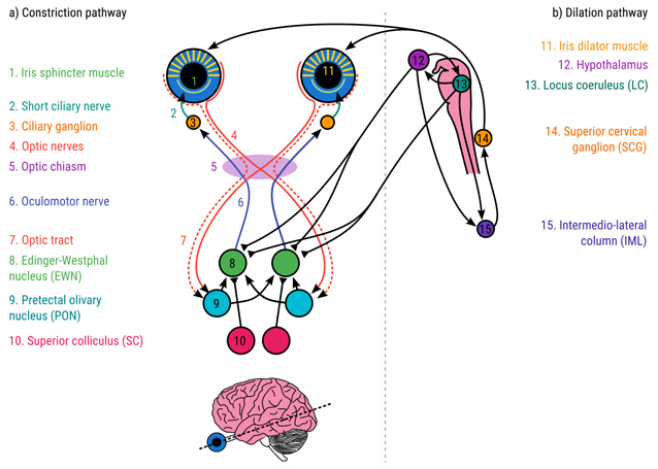


Figure 2.6: Overview on the dilation and constriction pathways of pupillary control.

The sympathetic pathway governs pupil dilation, which is often associated with increased arousal, emotional intensity, cognitive effort, and alert-

ness. While the precise neural circuitry underlying this mechanism is not yet fully understood, numerous studies point to the involvement of the locus coeruleus (LC), a brainstem nucleus located in the pons and a key component of the noradrenergic system [61,62]. The LC modulates cortical activity in response to salient stimuli and plays a central role in adaptive gain modulation and attention.
Conversely, the parasympathetic pathway controls pupil constriction. It involves midbrain structures, which process luminance information and coordinate motor commands via the oculomotor nerve to activate the iris sphincter muscle [61]. This pathway enables the pupillary light reflex, allowing the pupil to constrict in bright environments to protect retinal photoreceptors from overstimulation.
Although these pathways are roughly distinct, interactions have been demonstrated: for instance, the LC can simultaneously trigger the dilation pathway and inhibit the constriction one, actuating a response that have been associated with arousal and mental effort [61].
A change in pupil size is linked to three phenomena:

- The pupil light response (PLR) is the spontaneous mechanism that adjusts pupil diameter to brightness. Constriction is initially driven by the retina photoreceptors, i.e., cones and rods, and later maintained, if the same light condition persists, by specific ganglion cells. The PLR increases visual acuity and depth of field. Interestingly, this reflex is also modulated by cognitive processes, such as visual attention, subjective interpretation, mental imagery

- The pupil near response is the direct consequence of looking at a nearby or distant object, which determines a constriction and a dilation of the pupil size, respectively. It usually occurs together with vergence eye movements and the lens accommodation.

- The psychosensory pupil response is a specific type of dilation that follows an arousing stimuli. It may be triggered also by an unexpected change in the environment or mental effort during the completion of increasingly difficult tasks. In these latter cases, it has been hypothesised that the main responsible for such a response is an activation of the thalamus concurrently with the LC.

Another interesting class of spontaneous pupil size variation is the so-called hippus, i.e., periodic fluctuations, uncoupled from concurrent changes in the

environment, that are also associated with fatigue and cognitive load [63].

2.5.3 Blinking

Blinking is a rapid, coordinated movement involving the closure and reopening of the eyelids. Three muscles govern their movements: the *orbicularis oculi* for the downward motion, the *levator palpebrae superioris* and the superior tarsal muscles for the upward one. These movements protect the eye and maintain the tear film that preserves corneal hydration and clarity. Blinks can be classified into three main types: spontaneous, voluntary, and reflexive [64]. While the seconds are under conscious control and the thirds are triggered by specific external stimuli (e.g., sudden light, acoustic input, or corneal irritation), the first occurs without deliberate intent or obvious external cause.
Spontaneous blink rate varies across individuals and conditions, typically averaging between 10 and 20 blinks per minute. Interestingly, their frequency often exceeds what is strictly necessary for ocular maintenance, suggesting a broader functional role. Indeed, growing evidence indicates that spontaneous blinking reflects central nervous system activity, particularly involving the frontal cortex and dopaminergic systems [65]. Functional neuroimaging has highlighted the involvement of the medial frontal areas, including the supplementary motor area and anterior cingulate cortex, in the modulation of blinking behaviour [66, 67].
Moreover, spontaneous blinking has been associated with cognitive and attentional states. Blink suppression tends to occur during periods of high cognitive demand, such as during focused reading or visually intensive tasks, to minimise visual interruption. Conversely, increased blink frequency is observed during states of reduced vigilance or cognitive load [67]. This modulation is thought to result from a dynamic interaction between the nigrostriatal dopamine system and cortical control networks.

2.5.4 Eye-tracking measurement

The study of eye movements offers the opportunity to understand complex principles underlying executive functions, i.e., high order processes occurring in prefrontal cortex (including attention, memory, dynamic decision making and motor control) that mediate social, emotional and organisational aspects of individuals. Indeed, the relatively low number of degrees of freedom in the

oculomotor plant and the phylogenetically conserved control circuitry have made the eye movement system an ideal model for studying the programming and execution of discrete motor action sequences [68].
Video-based eye tracking has become the predominant method for measuring gaze direction. The advancement of infrared (IR) imaging and real-time image processing solidified pupil and corneal reflection tracking as the gold standard methodology. Most modern systems employ an IR-sensitive camera paired with IR light emitter to detect and extract two key ocular features: the centre of the pupil and the corneal reflection. These elements provide the spacial references that are necessary to determine the point of regard while allowing the system to disambiguate eye rotations from head movements. Indeed, on the one hand, when the eye rotates, the pupil shifts while the corneal reflection remains relatively stationary. On the other one, during head movements, both move in parallel.
There are two primary configurations for illuminating the eye: bright-pupil and dark-pupil methods, as illustrated in Figure 2.7 (Tobii Pro AB, Danderyd, Sweden).

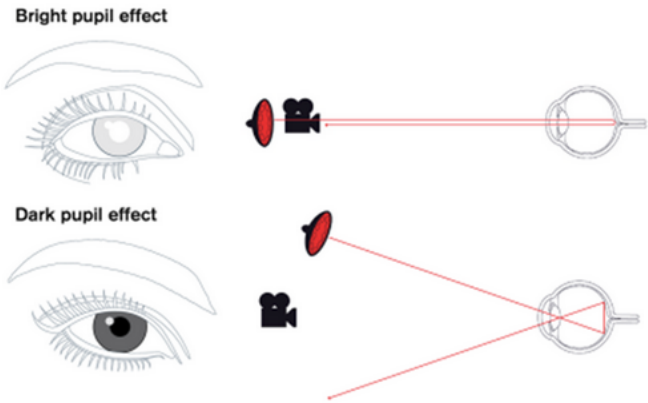


Figure 2.7: Graphical visualisation of the bright- and dark-pupil methods for gaze tracking.

Bright-pupil systems use coaxial IR illumination, i.e, the IR source is placed on the same axis as the camera. This arrangement exploits the retinal reflection, causing the pupil to glow and appear white. The main advantages in using such system is its high sensitivity in low-light environments and high tolerability to cameras' quality. However, their performance may

degrade under high ambient lighting or with smaller pupils. On the contrary, dark-pupil systems use off-axis IR lighting, making the pupil the darkest region in the image. This configuration is more robust under varying lighting conditions and with subjects who have dark irises or small pupils. Notably, the dark-pupil technique generally requires larger devices to be implemented, as more physical distance is needed to allow the spacial separation between the camera and the IR light source.

Eye trackers can be head-stabilised, remote, head-mounted, and embedded, and while their optical principles are similar, the form factor differs:

- Head-stabilised systems utilise a constraining object to avoid the participant's head movements, usually via bite-bars or chin-rests. These devices are used in neurophysiology or vision experiments where comfort is secondary to accuracy and precision. Indeed, this configuration allows to relevantly reduce noise in eye movement signals and to uniform the visual experience across all participants.

- Remote eye-tracker are contactless systems that can automatically alter the camera's field of view to compensate for head movements. They use pupil centre and corneal reflection to track eye position and head orientation. These devices typically set the camera below the stimulus area (e.g., a computer screen), to make the shape of the pupil more visible and less likely to be occluded. These systems always have a functional working area, called a "head box" and map eye movements on a defined calibration plane. If the participant leaves the head box or looks beyond the calibration plane, the tracking will be temporarily interrupted. Even if remote eye-tracker allows for a better interaction with the surrounding environment with respect to head-stabilised systems, the existence of the head box and calibration plane limit the range of exploration, hindering a naturalistic gaze behaviour. Moreover, they can be sensitive to extreme head movements, which cannot be compensated and cause data gaps.

- Head-mounted devices are directly worn by the participant, usually in the form of glasses or a headband (as in Figure 2.8). This system requires a camera or to be positioned in the visual path of one eye (monocular) or both (binocular), and an additional one that records the scene or field of view. Gaze tracking is done relative to the en-

tire field of view, which makes it ideal for real-world experiments, also supported by a wireless and minimally intrusive configuration. Relevant challenges in this type of system is the sensitivity to eccentric eye movements and the absence of an absolute coordinate system. Indeed, gaze data is recorded with respect to the scene camera, which acts like an imaginary screen that moves with the participant's head.

- Embedded eye-trackers are integrated in several type of clinical devices (eye surgery system) or consumer electronics (virtual reality headset). In this last case, rather than only monitoring eye movements, trackers represent also a way to interact with the experience.

Figure 2.8: Wearable eye-trackers in form of glasses (Tobii Pro AB, Danderyd, Sweden).

Regardless of configuration, a calibration procedure is necessary to map the relative positions of the pupil and corneal reflection to points on the stimulus display. This typically involves having the user fixate on a sequence of 5, 9, or 13 calibration points.

The eye tracking process generally comprises three computational stages: image acquisition, analysis, and gaze estimation. In the first phase, images are captured in real-time and forwarded for analysis. During image processing, the system detects the pupil and corneal reflection using either feature-based or model-based methods. Feature-based approaches rely on pixel intensity thresholds, edge detection, and shape heuristics to segment the relevant eye features. However, they can struggle with partial occlusion from eyelids or eyelashes, especially at extreme gaze angles or during blinks.

They are also susceptible to variations in pupil size, which can affect positional accuracy. Model-based methods offer greater robustness by fitting geometric models of the eye to the image data. These often assume that the pupil and iris are ellipsoidal and hierarchically nested within the sclera. By incorporating assumptions about contrast gradients (such as the pupil being darker than the iris, and the iris darker than the sclera), these systems can better localise features even under partial occlusion. However, model-based techniques are computationally demanding and require good initial feature estimates to avoid long recovery times. For this reason, many modern eye trackers use a hybrid methodology, combining the speed of feature-based detection with the resilience of model-based refinement.

While most systems use a single camera and IR source, adding additional cameras or IR sources can significantly improve robustness, especially in setups where the participant's head is free to move. Multi-source systems can relax calibration constraints and improve tracking accuracy across a larger field of view. Some advanced trackers dynamically switch between bright- and dark-pupil modes based on tracking quality, activating or deactivating IR sources as needed. Despite technological improvements, it remains difficult to compare performance across commercial systems due to proprietary algorithms and limited access to raw eye images. Systems that allow researchers to view and adjust the eye image in real-time provide an advantage, enabling better setup and detection of potential problems during data collection.

2.6 Voice and Speech

The human voice emerges from the intricate interaction between the respiratory system, the phonatory apparatus, and the CNS, approximately involving 100 muscles [69]. It is the product of finely coordinated muscular, biomechanical, and neural processes that transform the expiratory airflow into structured acoustic signals capable of conveying linguistic, emotional, and identity-related information.
At the respiratory level, vocal production begins with the lungs and diaphragm generating an airstream that ascends through the trachea and reaches the larynx. This triangle-shaped organ largely consists of cartilages that are attached to one another, and to surrounding structures, by muscles, fibrous and elastic tissue. Muscles are distinguished in extrinsic and

intrinsic:

- Extrinsic muscles support and position the larynx within the neck. Even if they act in this specific region, they originate in another one (e.g., the sternothyroid muscle, while being the principal depressor of the larynx, derives from the upper part of the sternum and insert into the thyroid cartilage).

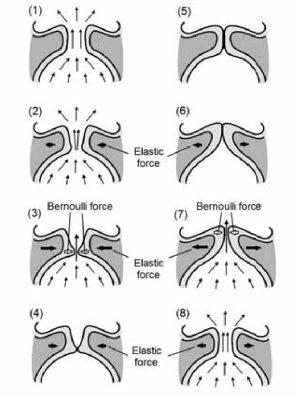
- Intrinsic muscles are the actual responsible for sound production, as they originate and insert in the larynx itself. For instance, the lateral and posterior cricoarytenoids works antagonistically: the posterior ones abduct the vocal folds (VF), while the lateral ones adduct them. Notably, the activity of the intrinsic muscles also allows and controls respiration.

The VF are roughly composed of three layers: an epithelial cover, a ligamentous membrane (that, together with the first constitutes a mucosal layer called *lamina propria*) and a deeper muscular tissue. These layers differ in the macromolecular composition of the extracellular matrix, which underlies the tissues' physical properties (such as shape, thickness and mass), as well as biomechanical characteristics, such as its rigidity and viscosity [70] and, consequently, control the overall vibratory capability of the VF. They are anchored posteriorly to the arytenoid cartilages and anteriorly to the thyroid cartilage. The VF are protected superiorly by the false vocal folds and the epiglottis. The variable opening between the vocal folds is termed the glottis.
During phonation, the diaphragm and thoracic muscles contract to produce subglottal pressure. When the vocal folds are fully adducted by the intrinsic laryngeal muscles, this pressure builds until it exceeds the muscular resistance of the adductors. At this point, the folds begin to part from their inferior margins, initiating a glottal opening through which air is expelled. This leads to a wave-like motion across the surface, that is referred to as the mucosal wave, which propagates in a complex three-dimensional pattern, displayed in Figure 2.9a. As air passes through the glottis into the supraglottic space, pressure drops due to increased velocity (in accordance with Bernoulli's principle), causing the vocal folds to recoil and close. This cyclical process is repeated rapidly, converting steady airflow into a series of pressure pulses. This mechanism is described by the myoelastic-aerodynamic theory of phonation, later refined by oscillation-based models. Therefore, the

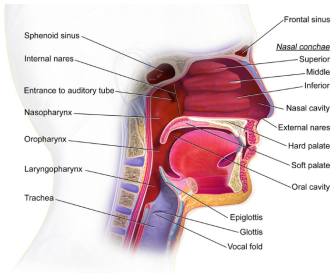
vocal folds act as self-sustained oscillators, translating aerodynamic energy into an acoustic waveform. The rate of this oscillation determines the fundamental frequency (F0) of the voice, perceived as pitch, while variations in subglottal pressure modulate vocal intensity.
Once generated, the sound is further modified as it travels along the vocal tract (VT), which functions as a dynamic filter. The VT comprises anatomical structures (highlighted in Figure 2.9b) that due to their shape and size amplify or attenuate specific frequencies via resonance effects, thereby forming the timbre and partially the intensity of the sound. The overall spectral output of the voice is determined by so-called formants, i.e., clusters of harmonically related frequencies that derive from the resonant characteristics of the vocal tract. Key components of the VT include:

- The pharynx, a muscular tube forming a junction between the respiratory and digestive tracts, capable of modifying its length and diameter.
- The soft palate (or velum), a muscular structure that can elevate or lower to regulate nasal airflow; its lowering permits the articulation of nasal phonemes.
- The nasal cavities, which serve as fixed resonators due to their unchanging volume.
- The paranasal sinuses, contributing minimal resonance due to high damping properties.
- The tongue, a highly mobile organ with intrinsic and extrinsic muscles that can alter its shape and position to shape the oral cavity.
- The teeth, which provide fixed articulation surfaces.
- The mandible (jaw), whose movement influences the configuration of other articulators and thus affects resonance.
- The lips, which participate in the formation of speech sounds and resonance shaping.

The dimensions of the vocal tract, and consequently its resonant characteristics, are influenced by biological sex, developmental stage, and hormonal factors. For instance, puberty induces structural changes such as laryngeal



(a) Schematic visualisation of the mucosal wave [71].



(b) Schematic representation of vocal tract structures [72].

Figure 2.9: Overview on voice production physiology.

descent and elongation of the VT, leading to lower-pitched voices, particularly in males, where it may reach up to 25cm in length compared to approximately 17cm in females. Articulatory precision also evolves with age and practice, improving with language exposure and training, such as in speech therapy or vocal pedagogy. Beyond producing voiced sounds, the vocal tract can actively generate aperiodic acoustic energy by constricting airflow at various points. This results in turbulent airstream or full occlusions, both of which are critical for the articulation of consonant sounds. Thus, the interplay of expiration, tissue dynamics, and vocal tract configuration allows for the full range of speech sounds.

At the neural level, vocal production engages multiple cortical and subcortical areas. The CNS oversees the entire process, from conceptual planning to motor execution. The left inferior frontal gyrus (specifically, the Broca's area) is central to the motor planning of speech and is closely linked with the supplementary motor area, which coordinates the sequences of articulatory movements. Wernicke's area, located in the posterior part of the left superior temporal gyrus, is involved in the comprehension of spoken and written language and it is connected to Broca's area via the arcuate fasciculus, a major white matter tract. Motor commands are relayed to the primary mo-

tor cortex, particularly its inferior and lateral regions, which indeed house the laryngeal motor cortex and thus innervate the articulatory musculature. These muscles are controlled via motor pathways that descend through the brainstem and reach the larynx via the superior and recurrent laryngeal nerves, both branches of the vagus nerve. Interestingly, nuclei located in the brainstem are also responsible for involuntary vocalisations, e.g., newborns' cry or laugh [73], and the coordination between voice production and respiration [74]. Language development, including native language acquisition and the influence of cultural-linguistic environment, shapes the habitual motor patterns of vocal production. These patterns are further modifiable through experience and training, such as in singing or multilingualism. Additionally, the cingulate cortex mediates vocal parameters such as pitch, tempo, and intensity due to emotions and psychological states, highlighting the integrative nature of affective and cognitive systems in voice control [73].

2.6.1 Acoustic measurements

Acoustic recordings are acquired through microphones. Although the European Laryngological Society emphasises the importance of good audio quality, it does not provide detailed technical specifications. Generally, the ideal instrumentation for voice analysis should have a flat frequency response, a noise floor at least 15dB below the quietest phonation, a dynamic range capable of capturing the loudest phonation, and an optimal distance-to-source ratio for maintaining spectral integrity [75,76]. Microphones come in various types, categorised by both their transducer technology and form factor. Dynamic microphones use electromagnetic induction to convert sound into an electrical signal, exploiting the relative movement between a coil attached to a membrane and a fixed magnet. These devices are durable, cost-effective and well resistant to high sound pressure level. However, they are usually bulky and have low sensitivity and frequency response. On the other hand, condenser microphones use a capacitor to actualise such a conversion, they require external power (which can limit portability) but they present higher sensitivity and fidelity. A subtype of condenser microphones are electret microphones, which deploy a dielectric material with quasi-permanent electrical polarisation that provides a constant charge on the capacitor. When miniaturised, these devices can be clipped on clothing, providing high quality signal with a light-weight, wireless technology. However, displacement issues and rustling may negatively affect the recordings, as well as their

omni-directionality acquisition modality. This latter problem can be easily overcome by adopting a cardioid polar pickup, i.e., a peculiar pattern allowing the microphone to be most sensitive to sound coming from the front and less sensitive to sound from the sides and rear. Nevertheless, a proximity effect can arise, overemphasising low frequencies if the speaker is placed too close to the source. Oppositely, headset microphones are more stable and provide better isolation from surrounding noise sources, however, when used for long periods, they might become intrusive and uncomfortable.
More recently, smartphones have emerged as promising alternatives, offering portability, ease-of-use, and cost-efficiency. Indeed, they are equipped with extremely small silicon-based condenser microphones that are robust and power-efficient. Studies reported that such devices yield reliable recordings for acoustic analysis, although caution is needed in uncontrolled environments and with regard to microphone positioning [76]. While certain acoustic parameters (such as the harmonic-to-noise ratio) may be affected, other features appear relatively stable. Although they are not properly wearable devices, smartphones are ubiquitous and remain a compelling tool for longitudinal voice monitoring, especially in naturalistic settings.

Chapter 3

Virtual Reality, Climate Awareness and Reaction

This chapter introduces the design and development of a virtual reality experience to address the well-known action-value gap in climate change research. Such a psychological phenomenon has been identified as one of most critical aspect that hinder people's adoption of mitigation strategies and eco-friendly attitudes. This strategy also implemented a multimodal evaluation approach, encompassing subjective questionnaires, behavioural measures (i.e., in-VR movements), and physiological monitoring (particularly by means of ECG and EDA). The aims were to gain a wide perspective on the feelings of presence and immersion, the attitudinal shift and the spontaneous emotional reactions. By exposing 40 subjects to a VR experience featuring a major natural disaster (i.e., a wildfire), good rates of presence were achieved, also supported by stronger body responses, and both knowledge and intentions toward sustainability positively changed, especially in males, younger and local participants. [1]

[1]The study presented in this chapter was conducted during a research visit at the Faculdade de Engenharia da Universidade do Porto (FEUP), Porto, Portugal. The 3D environment and the movement mechanics were already available at the host institution and served as a conceptual starting point. The overall experimental design, data acquisition protocol, participant management, and data collection procedures were developed and implemented by the doctoral candidate. A preliminary analysis on the assessment over the technical aspects of the developed experience has been published as "Burning Re-

3.1 Introduction

Virtual reality (VR) consists in the use of computer modelling and simulation that enables a person to interact with an artificial 3D visual or multisensory environment. Recent advances in mobile technology and the introduction of consumer-friendly VR headsets have allowed the development of applications for engaging, comfortable user experiences, making VR a growingly dominant media distribution platform in several contexts, from education to healthcare [78]. In the last few years, the global impact of environmental changes has guided research to focus and create a body of work concerning ecological sustainability [79,80]. Learning about climate change represents a complex task, as it requires dealing with abstract and highly interdisciplinary information, as well as trusting them [81, 82]. The lack of engagement that follows such an intricate understanding has been explained with multiple theories, e.g., structural barriers (economic issues) and social factors (political views). Psychological reasons were also identified: environmental threats are indeed perceived as both temporally and spacially distant, and uncertain [3, 83]. Moreover, reported low locus of control and self-efficacy hinder people's adoption and adaptation to sustainable behaviours even if they are genuinely concerned about climatic alterations [84, 85]. VR may effectively address these issues as it has the advantage of building immersive, accurate experiences that may consequently facilitate a better divulgence of climate change topics. Indeed, by reducing the psychological distance, such an approach could create an immediate and interactive visualisation of human responsibilities over ecosystems degradation. On the other hand, it could also show the positive outcome of what sustainable attitudes may bring in the future to encourage climate change awareness and action. Studies regarding the intersection of VR and climate change highlighted that the exposure to exaggerated climatic responses or the virtual depiction of an individual carbon footprint led to more sustainable choices and self-efficacy, which were mediated by the degree of environmental concern [86] and feedback time [87]. Moreover, after the VR exposure, information are retained in the long term, without being constrained by age, VR familiarity or compensation [88, 89]. However, ambiguous outcomes have also arisen: research generally agrees that VR increases the sense of presence and immersion [88, 90–92] and provokes stronger emotional engagement [88, 93] with respect to traditional me-

---

ality: Experiencing Climate Change Through Virtual Reality" in *2025 IEEE 14th Global Conference on Consumer Electronics (GCCE)* [77].

dia means (e.g., videos played on flat screens or paper). Nonetheless, sustainable behaviours did not improve accordingly. This might be caused by inappropriately addressing the spacial distance factor (i.e., using scenarios that are located in remote geographical contexts from participants' habitual place of living) [91, 92] or cybersickness, which may attenuate the positive effects enhanced by VR [92]. Indeed, this specific type of physical discomfort includes symptoms as nausea, headache and dizziness that may relevantly hinder the usability of virtual experiences [94]. Other limitations have been discovered, mostly regarding the designing stages: while the virtual environment realism and human voices' gender do not seem to impact environmentalism measures and learning [81, 84], the cognitive load associated with richer scenarios, the in-experience body movements allowance, and the difference between loss- and gain-frameworks of the VR conveyed message (i.e., observing a tree being destroyed by a natural disaster rather than assisting to its growth) led to inconsistent results [81, 93, 95]. Therefore, great care must be exerted when developing the VR experience so that it taxes the participants' emotional sphere to realise an efficient educational program, without overloading their mental resources.

An additional aspect to consider is the assessment modality of emotion and affective dimensions, which is typically based on subjective measurements, e.g., questionnaires and interviews. Despite their accessibility and ease-of-use, they are also time consuming, especially when used in batch to gain a wider perspective on a multidimensional phenomenon. Furthermore, they rely on an individual's conscious self-introspection. This represents a major limitation as such techniques become susceptible of detrimental psychological effects, e.g., social desirability or recall difficulties, that could consequently bias participants' answers [11]. On the other hand, physiological activity cannot be easily voluntarily altered [96], hence providing a robust and reliable way to investigate whether and how the human body spontaneously reacts to the VR stimuli. Wearable devices are particularly suited for this task due their compact and wireless design, which relevantly diminishes the risk of the break of presence. Electroencephalographic (EEG) headsets are promising candidates, however, their usage can be quite challenging when interacting with the head mounted display providing the VR experience [97, 98]. Despite precautions, they may reduce the sense of presence and immersion (e.g., limiting head movements by laying it on a head rest). Therefore, recent research has also deployed eye-tracking technology [99], electrocardiographic (ECG)

and electrodermal activity (EDA) devices [100].
In the context of climate change themed VR experiences, only a few studies implemented a hybrid approach of both subjective and objective measurements, highlighting the strong need of performing further investigation to gain a comprehensive overview over VR capabilities in promoting environmental sustainability. Nevertheless, these works underlined that the illusions provided in the virtual environment are able to elicit a physical response, i.e., a greater level of arousal as measured with EDA especially in threatening scenarios (e.g., experiencing a fire while being embodied in a tree) [101]. A faster heart rate was also achieved during an experience proposing the fact checking of climate change fake news, that was further enhanced by the humanised speech of an avatar [102]. Finally, longer saccades were associated with a lower rate of task-unrelated thoughts while exploring an endangered coral reef, potentially indicating a higher involvement degree [103]. Such insights into autonomic nervous system activity, along with complimentary psychological evaluations, could help researchers understand why some individuals are indifferent or, on the contrary, relevantly afraid of climate change, therefore objectively characterising their behaviour. Consequently, this approach could help activists and sociologists to plan more effective information campaigns and public service announcements, as well as to guide and protect the population from the spreading of misinformation about climate change.

This chapter describes the development of a VR experience that simulates a wildfire surrounding a well-known viewpoint located in northern Portugal. Concurrently, it introduces a physiological monitoring system, based on ECG and EDA signals, that was used to assess spontaneous body responses to the natural disaster and correlate the latter with behavioural metrics and subjective questionnaires.
The objectives are:

1. To evaluate the sense of immersion and presence enabled by the VR experience and to understand whether the virtual exposure to this natural disaster is able to modify climate change beliefs and attitudes.

2. To investigate if and how several socio-demographic factors (e.g., gender, age, nationality) influence the degree of the above mentioned changes.

3. To assess the relationship between behavioural measures (i.e., in-VR

and head movements) and

- Immersion and presence.
- The performance of a custom wildfire centred quiz.

4. To analyse the physiological responses to the individual phases of the VR experience, uncovering potential differences related to the affective content conveyed by the provided stimuli.

3.2 Materials and Methods

3.2.1 Participants

Forty participants (16 females F, mean age = 32.9±13.5, 24 males M, mean age = 29.9±11.4) were recruited at the Graphics, Interaction and Games (GIG) Laboratory of the University of Porto, Porto, Portugal. Thirty subjects indicated a Portuguese nationality, while the others were international volunteers. Figure 3.1 shows their distribution.

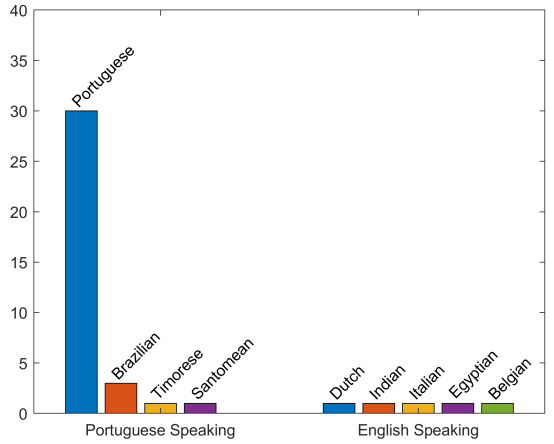


Figure 3.1: Nationality distribution across participants.

The degree of familiarity with VR was assessed with a simple 5-point Likert scale, where 1 = no experience and 5 = expert user. Figure 3.2 represents the score distribution.

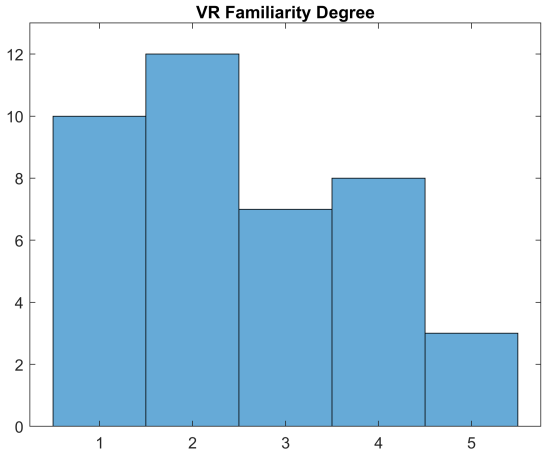


Figure 3.2: VR familiarity score distribution.

Participants were also asked to disclose whether they have visited, or at least knew, the actual place where the VR experience was set at. In total, eight people out of forty (20%) knew such location.
No participant reported having cardiovascular or dermatological diseases, as well as suffering from seizure or diabetes, i.e., the exclusion criteria that were set for wearing physiological sensors and participating the VR experience.

3.2.2 Subjective measurement

The Climate Change Attitude Scale (CCAS, in its original version [104] and in its Portuguese-translated validated version [105]) was administered to investigate climate change beliefs and intentions. It is composed of 15 items that must be answered according to a 5-point Likert scale. Specifically, four of them are reverse coded, meaning that the scores assigned to each response option are inverted, so that the highest score corresponds to the most negative response, and vice versa. This procedure is typically done to reduce response bias (i.e., participants may fall into agreeing with all provided statements) and ensure attention. Table 3.1 displays its components.

On the other hand, the Igroup Presence Questionnaire (IPQ), both in its original [106] and Portuguese-translated validated version [107], was de-

Table 3.1: CCAS Items.

Climate Change Attitude Scale	Beliefs	Intentions
1. I believe our climate is changing.	*	
2. I am concerned about global climate change.	*	
3. I believe there is evidence of global climate change.	*	
4. Global climate change will impact our environment in the next 10 years.	*	
5. Global climate change will impact future generations.	*	
6. The actions of individuals can make a positive difference in global climate change.	*	
7. Human activities cause global climate change.	*	
8. Climate change has a negative effect on our lives.	*	
9. We cannot do anything to stop global climate change.[1]		*
10. I can do my part to make the world a better place for future generations.	*	
11. Knowing about environmental problems and issues is important to me.		*
12. I think most of the concerns about environmental problems have been exaggerated.[1]		*
13. Things I do have no effect on the quality of the environment.[1]		*
14. It is a waste of time to work to solve environmental problems.[1]		*
15. There is not much I can do that will help solve environmental problems.[1]		*

[1] stands for reverse coded items.

ployed to measure the sense of presence felt by participants within the VR environment. The IPQ consists of 14 items with a 5-point Likert scoring scheme and, similarly to the CCAS, can be subdivided into three subscales that individually assess overall sense of presence (OSP), spacial presence (SP), involvement (INV), and perceived realism (REAL). Table 3.2 displays the items.

Table 3.2: IPQ Items.

Igroup Presence Questionnaire	Subscale
1. In the computer generated world I had a sense of "being there".	OSP
2. Somehow I felt that the virtual world surrounded me.	SP
3. I felt like I was just perceiving pictures.[1]	SP
4. I did not feel present in the virtual space.	SP
5. I had a sense of acting in the virtual space, rather than operating something from outside.	SP
6. I felt present in the virtual space.	SP
7. How aware were you of the real world surrounding while navigating in the virtual world?[1]	INV
8. I was not aware of my real environment.	INV
9. I still paid attention to the real environment.[1]	INV
10. I was completely captivated by the virtual world.	INV
11. How real did the virtual world seem to you?[1]	REAL
12. How much did your experience in the virtual environment seem consistent with your real world experience?	REAL
13. How real did the virtual world seem to you?	REAL
14. The virtual world seemed more realistic than the real world.	REAL

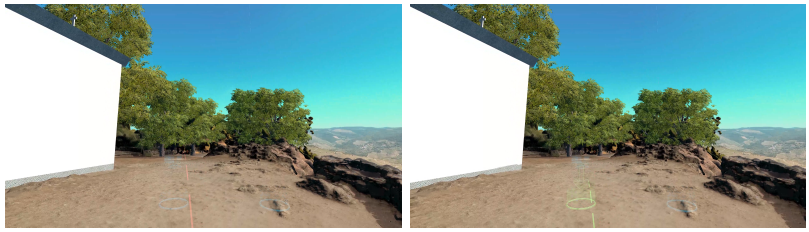
[1] stands for reverse coded items.

3.2.3 VR equipment and development

The experience was built and developed using the Unity game engine (Unity Technologies, San Francisco, CA, USA) and run on a desktop PC equipped with a 11th Gen Intel Core i9-11900F processor and a NVIDIA GeForce RTX 3080 graphics card. It was delivered through the head mounted display HTC Vive Pro 2 (HTC Corporation, Taoyuan, Taiwan). This device presented a 120° field of view, a 120Hz refresh rate, a 2448 x 2448 pixels resolution per eye and integrated headphones.

Participants could move in the real world with small steps and turns, but since the area covered by the actual location is larger than the available space in laboratory settings, in-VR movements were enabled by a set of portals disseminated all around the viewpoint (Fig. 3.3). These can be accessed through a teleportation system, which were developed and validated in previous studies [78,108]. Specifically, after pressing the VIVE controller's

trackpad, participant were able to aim at the spot they want to go to and, once the position was locked (signalled by the portal becoming green and the appearance of a downward arrow, see Fig. 3.3b), they were immediately transported there, when releasing the same button. Portals were made visible only when players pressed the controller's trackpad. This system was implemented also to reduce the occurrence of cybersickness, which is indeed provoked by a mismatch between internal- and intra-sensory inputs coming from visual and somatosensory proprioceptors [109].



(a) Invalid position, represented by a red parabola.

(b) Valid position, represented by a green parabola.

Figure 3.3: Teleportation system visual cues to validate movements within the virtual scenario.

As a novel contribution, for this thesis, the same experience was created in European Portuguese and English to allow foreign volunteers to participate too, therefore broadening its applicability. In the first one, human voices were recorded to acquire a set of audio cues to use in the experience, whereas in the second one, the ElevenLab software [110] was used to perform text-to-speech conversion and create realistic artificial voices, sampled at 44.1kHz and with a 128kbps bitrate.

3.2.4 Experimental procedure

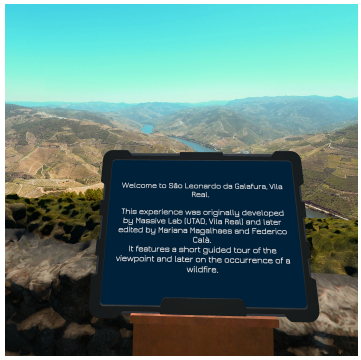
Participants completed a personal data questionnaire and the CCAS before the experiment day.
Upon arrival, they were given the Shimmer 3 wearable sensors (Shimmer Sensing Ltd, Dublin, Ireland) to acquire their electrocardiographic (ECG) and electrodermal (EDA) activity. These devices are equipped with adjustable wrist and chest belts for better usability and portability, and store

the signals on an internal SD card. Once they are removed from the docking station, the acquisition automatically starts.
The synchronisation between the physiological recordings and the VR experience deployment was realised with the pressing of specific keyboard keys. A custom python code ran in the background, logging the time stamps of keyboard pressings on a spreadsheet file. Particularly, when the Shimmers were undocked, the *S* key was pressed.
After applying the wearables, an open eye baseline period followed, where participants looked at a centred white cross on a grey background for 120s in order to acquire their resting state. It started by pressing the spacebar.
Subsequently, the VR experience was deployed, lasting approximately 20 minutes. The *S* key was pressed again to mark the beginning of the experience and synchronise the behavioural signals as well. During the experience, other keys were pressed (see Subsection 3.2.5).
When it finished, participants removed the head-mounted display and the biomedical sensors.
Finally, they completed the IPQ and the CCAS questionnaires, as well as a custom-made quiz about wildfires. It consists of 8 true/false questions that aimed at evaluating the levels of attention and information assimilation during the experience. Items are reported in Appendix A.

3.2.5 The VR experience

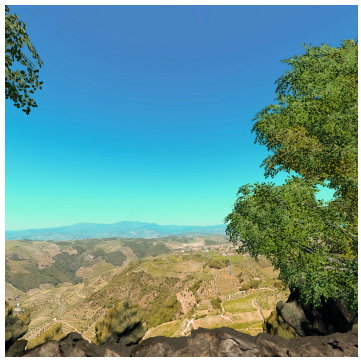
The virtual environment consisted of a replica of the São Leonardo da Galafura viewpoint, located in the Portuguese UNESCO World Heritage Site of the Alto Douro Wine Region (Porto District). The scenario included a panoramic view of the Rio Douro from the summit, its surrounding stone wall and flora, different terrain elements such as a sandy texture and pebbles, and an interactive tablet that greeted participants and briefly summarised the VR experience characteristics and timeline. The environment was originally developed in [78] using photogrammetric techniques. A reproduction of the Santa Barbara Chapel was also already included; however, it was completely redesigned using the ProBuilder Unity package to fulfil the novel research questions of this thesis, so that the building could be accessed internally. Figure 3.4 displays a few snapshots taken within the virtual scenario.

To enhance realism, a so-called wind zone was added. This Unity component recreates the effects of wind, allowing to control several aspect such



(a) Interactive tablet greeting participants.

(b) Panoramic view of the Douro Valley.



(c) Panoramic view of Galafura village.

(d) Reproduction of the Santa Barbara chapel.

Figure 3.4: Panoramic snapshots taken within the virtual scenario.

as the radius of the zone, turbulence, pulse magnitude and frequency. Moreover, a 2D audio cue of birds' chirping was included.

The VR experience featured 3 main events:

- A guided tour of the viewpoint.
- A wildfire.
- A radio interview.

The Unity console log printed custom messages indicating the end of each of these sections to let the experimenter manually trigger the following one, helping to keep track of the ongoing phases. Specifically, the experimenter pressed the "enter" key to start the guided tour, the spacebar to trigger the wildfire, the A to start the radio interview, and the F key to end the key logging process, as the block diagram in Figure 3.5 schematically shows.

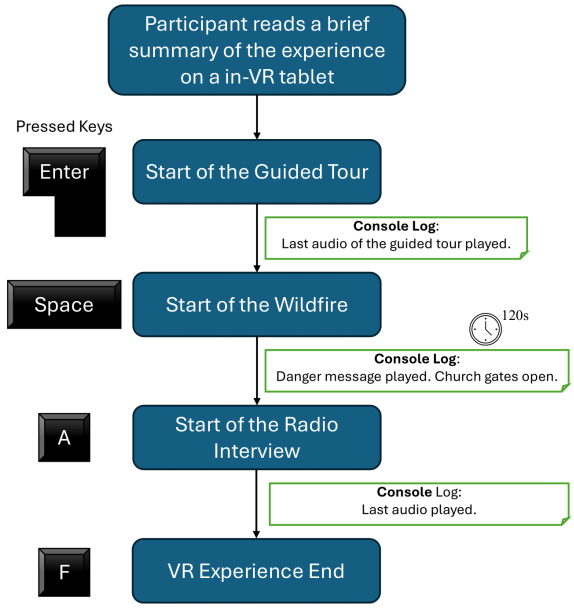


Figure 3.5: Block diagram of the experience mechanics, alongside with the sequence of keyboard keys pressing and console log messages.

Guided tour

This section both served as an in-VR habituation period, in order to get participants acquainted with the experience narrative and mechanics, and as a way to develop a sense of attachment to the place in participants that never visited it. This may enhance an emotional response during the wildfire due to a reduced psychological distance [81]. The tour was divided in 5 stops: a welcoming message, geographical and historical information about

the Douro river and the Rabelo boats that used to navigate it, details about Porto wine tradition and production, an explanation of the flora and, finally, a parting message, that included a note motivating participants into adopting responsible, sustainable behaviours to preserve the beauty of the viewpoint, as well as warning them about the dangers of climate change. All given information were extracted from [108]. Appendix B reports the full scripts in Portuguese and English of this section.
Participants followed green bubbles (as displayed in Figure 3.6) that appeared and vibrated mid-air over portals until the player reached it, causing it to vanish. Then, the corresponding audio was played in the integrated headset headphones. The end of each message hinted the position of the next bubble and made it visible.

Figure 3.6: The green bubble indicating the position to reach in order to trigger the related audio of the guided tour.

Wildfire

A wildfire was chosen to investigate on climate change awareness and perception as it represents a major natural disaster that is recently becoming more frequent and severe also in northern Portugal regions [111]. Moreover, research proved that experiencing extreme weather events strongly relates to increased negative emotions and climate concern [83].

Custom particle effects were attached to each tree and shrub to simulate smoke and flames, which intensified gradually while the wildfire continued. To give the impression of a realistic spreading, the fire started from the farthest flora along the hillside and then it slowly propagated between adjacent branches and advanced towards the summit. The scenario skybox (i.e., a background that gives the impression of a complex scenery at the horizon) gradually changed from sunny to cloudy and additional smoke filled the virtual environment. Figure 3.7 displays the fire sequence events.

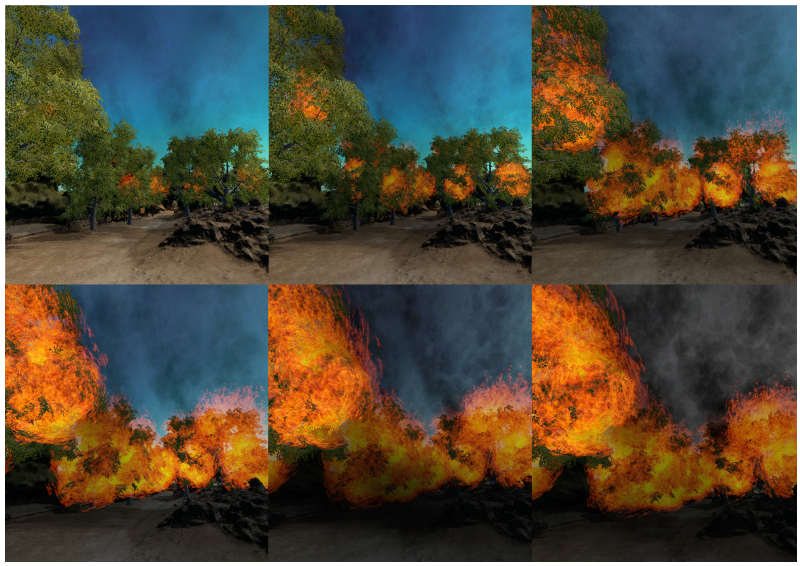


Figure 3.7: Fire and smoke particle effects intensify with time to give the impression of a realistic wildfire progression. Meanwhile, the background skybox gets darker and the scenario fills with additional smoke.

A rabbit escaping the viewpoint was added to further enhance the feeling of danger. Several audio cues were also included: a 2D audio recording of a fire was added to the entire scenario, together with a reproduction of the Porto official fire siren system that automatically started 30s after triggering the wildfire. On the other hand, different spacial audio recordings of fire were embedded in each flora element so that players could hear them louder if they got closer to the burning vegetation.
During the wildfire, participants were not given any specific instruction,

thus, they were free either to stand still or to further explore the environment. However, after 120s, an automated message was played to warn them to seek shelter within the chapel. Although this longer duration may have hindered realism, the window length was chosen to gain a minimally appropriate interval where to investigate participants' physical responses. When they reached the portal right before the chapel gate, this latter opened allowing them to enter within the building, and then automatically closed.

Radio interview

This section was designed according to the hypothesis that the presentation of useful information about risk mitigation when experiencing stressful situations can better support their integration and make them more effective. Indeed, it is supposed to help players believe they can manage and overcome a dramatic situation [88]. Therefore, it may promote awareness and self-efficacy strategies.
The chapel indoor environment was not an exact replica of the actual one. Nevertheless, it was provided with coherent elements such as a cross, a bookrest, three rows of wooden benches, candles, and a small stand with a radio atop of it. This object was included to deliver the simulated radio interview between a host and two specialists that explains:

- How wildfires originate and their relation with the worsening climate crisis.
- The severe consequences of wildfires on the environment, as well as on both human physical and mental health.
- A few advices regarding personal protection in case of wildfires, and precaution methods to prevent their occurrence.

To better involve participants, a slideshow appeared over the radio to display images and graphics, supporting the recordings content (see Figure 3.8).

The audio cue of the outside fire was maintained but its loudness was reduced. Then it progressively decreased until fading out completely so that it could match the final message of the interview, where the host notified that the fire was extinguished.
Appendix B reports all images and graphs presented in the slideshow, and the scripts for the interview in both languages.

Figure 3.8: View of the slideshow appearing over the stand with the radio atop of it. The expression "Incêndios florestais" stands for "Wildfires".

3.2.6 Physiological data analysis

Both wearable devices used a 128Hz sampling frequency. After acquisition, a code loaded and read the key logger file, extracted the time stamps and converted them in order to automatically segment the physiological signals into four sections: baseline, guided tour, wildfire and church.

Signal pre-processing

EDA signals were preprocessed with a low-pass FIR filter of order 500, with a cut-off frequency 5 Hz, and z-scored. Figure 3.9a displays these steps. The separation between the tonic and phasic components was performed applying the cvxEDA mathematical modelling approach, based on convex optimisation problem [112]. It is based on Bayesian statistical principles and models the observed skin conductance (SC) signal using a physiologically meaningful framework. Specifically, its application allows to decompose and obtain a slowly varying tonic component (Figure 3.9d), a phasic component (Figure 3.9b), and sudomotor nerve activity signal (Figure 3.9c). Notably, changes in the phasic component can be spontaneous or can arise in response to an external stimulus with a latency of 3-5 seconds. Each SCR is the result

of a neural burst in the SMNA, which elicits the sweating from the glands under the skin surface of the hand, as explained in Chapter 2.

The ECG signal was preprocessed with Pan-Tompkin's algorithm [113, 114] to extract the tachogram, i.e., the time series of the beat-to-beat intervals over time. After mean removal and bandpass filtering with a 3rd-order Butterworth filter (between 5-15Hz), the signal is differentiated to better highlight the steep slopes of the QRS complex. The resulting series is squared point-wise to make all values positive, emphasise larger differences and suppress small noise components. Then, a moving average smooths the signal to represent the energy of the QRS complex, approximating the area under its curve. Finally, two thresholds are dynamically updated: one for signal peaks and one for noise. In the first case, if a peak exceeds the signal threshold and occurs within a plausible time frame since the last detected R-peak (based on physiological limits), it is marked as an R-peak, whereas missed beats (e.g., due to arrhythmia) are recovered using search-back logic. The heart rate variability (HRV) signal was obtained by subtracting successive R peaks locations. Possible outliers were detected (i.e., the elements that deviated more than five scaled MAD from the median) and substituted using a shape-preserving piecewise cubic spline interpolation. Figure 3.10a displays the identification of the R-peaks, whereas Figure 3.10b the computation of the HRV sequence.

Feature Extraction

The feature extraction pipeline for the EDA signal was developed in according to [115] to analyse the SCRs.
Two parameters were derived from the tonic component by applying a 20s-long non-overlapping window: skin conductance level (SCL) mean and standard deviation.
From the phasic component, the SCRs were analysed with 5s-long non-overlapping windows. For each time segment the mean value of the SCR (SCRmean), the number of SCR (SCRn), the amplitude of the maximum SCR peak (SCRPeak), and the sum of all peaks amplitude within the time windows (SCRAmpsum) were extracted.
Furthermore, by dividing the overall cleaned EDA signal into 30s-long non-overlapping windows, the spectral power in the range 0.04Hz-0.15Hz was computed (EDAsymp).
The extracted features within all-time segments were then averaged. Table

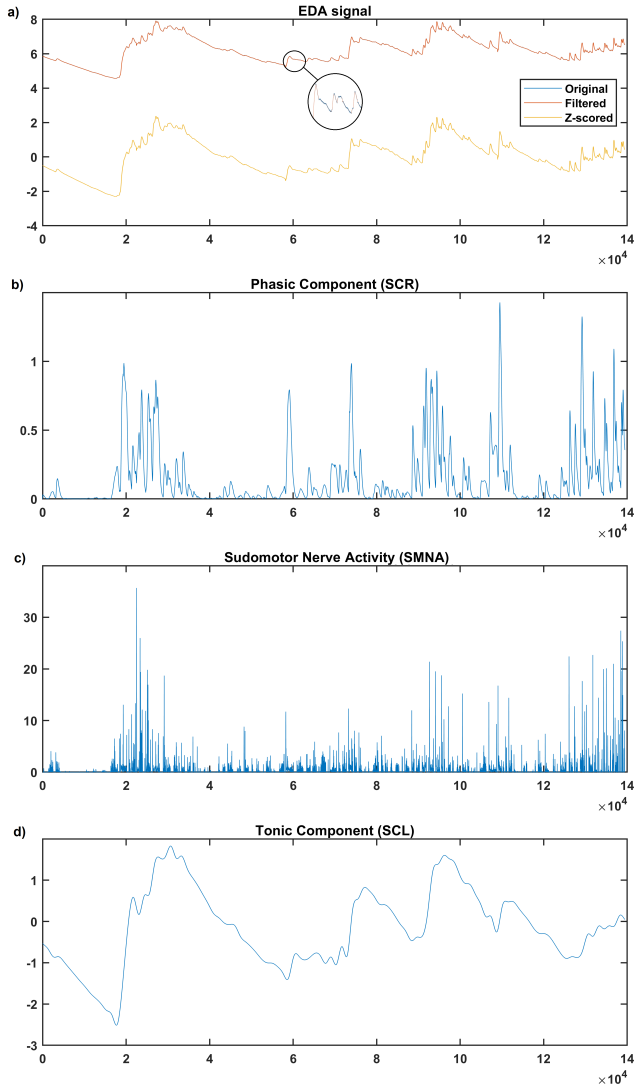


Figure 3.9: EDA preprocessing and decomposition pipeline. a) 3-steps preprocessing. The close-up highlights the low-pass filtering contribute. b) Phasic Component. c) SMNA Component. d) Tonic Component

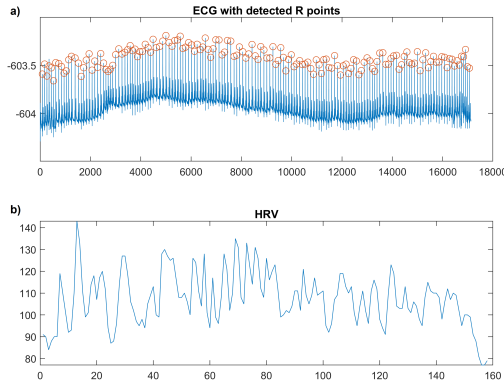


Figure 3.10: ECG pipeline process. a) Identification of R peaks with Pan-Tompkin's algorithm. b) Clean HRV time series.

3.3 reports them along with a brief description.

Table 3.3: Overview on extracted EDA features

Feature	**Description**
SCLmean	Mean value of the tonic component
SCLstd	Standard deviation of the tonic component
SCRmean	Mean value of the NSSCR peaks amplitude
SCRn	Total number of NSSCR peaks
SCRPeak	Maximum value of NSSCR peaks amplitude
SCRAmpsum	Summation of all NSSCR peaks amplitudes
EDAsymp	EDA spectral power in the 0.04Hz-0.15Hz band

Concerning HRV, a set of features from both temporal and frequency domains were included. Notably, as phases have different durations, temporal metrics were computed within 30s-long windows without overlap. Firstly, the meanRR (i.e., the average value of the consecutive RR intervals) was calculated, from which the beats-per-minute (BPM) was also derived.
Then, the standard deviation of the inter beat intervals (IBIs) was computed. This metric, typically indicated with the acronym SDRR, is largely mediated by the respiratory sinus arrhythmia, a parasympathetic phenomenon that generates physiological fluctuations due to breathing.

The NN50 represents the number of adjacent NN intervals that differ from each other by more than 50ms. Since this value can be relevantly biased by the length of recordings, the percentage of adjacent NN intervals that differ by ore than 50ms was proposed. Both metrics are related to the activity of the PNS [52].
The root mean square of successive differences between normal heartbeats (RMSSD) is obtained according to Equation 3.1:

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2} \tag{3.1}$$

where N is the number of RR intervals.
The RMSSD reflects the primary time-domain measure used to estimate the vagally mediated changes of HRV.
The SDRM, computed as the ratio between SDRR and RMSSD, was introduced as a potential surrogate of a frequency measure, i.e., the ratio between low and high frequency of the HRV. It is supposed to detect the balance between long- and short-term variability, regardless of the underlying mechanism. Moreover, the simplicity of calculation and the absence of limiting assumptions on the data make the use SDRM appealing [116].
The Triangular Index is calculated by dividing the total number of RR intervals by the number of RR intervals in the modal bin (i.e., the maximum of the density distribution) of a histogram of RR intervals. From the same graphical representation, the TINN can be extracted as the baseline width of the histogram displaying NN intervals.
Similarly to EEG analysis, the HRV can be separated in four components that operate within different frequency ranges. Specifically, the ultra-low frequency band is defined under 0.003Hz and to be properly computed, an HRV recording of 24h should be available. Therefore, this rhythm was not included in the analysis. The very-low frequency (VLF) band is defined between 0.0033-0.04Hz and it can be influenced by physical activity, thermoregulation and parasympathetic blockade [52]. The low frequency (LF) band (0.04-0.15Hz) mainly reflects baroreceptor activity (controlled by both SNS and PNS). Finally, the high frequency (HF) band (0.15-0.40Hz) is conventionally associated with parasympathetic activity and mirrors heart rate variations caused by the respiratory cycle. The total power (TP) was computed as the sum of the VLF, LF and HF rhythms. Then, each band was used to extract the peak, the percentage over TP, and the power expressed in

normalised units. Finally, the LF/HF ratio was derived to estimate the balance between SNS and PNS activity. Usually, a low ratio suggests parasympathetic dominance. This is seen when engaging in tend-and-befriend behaviours. In contrast, a high value indicates sympathetic dominance, which occurs in fight-or-flight condition or parasympathetic withdrawal [52]. Table 3.4 summarises the extracted HRV metrics, alongside with a brief description.

Table 3.4: Overview of extracted HRV features

Feature	**Description**
BPM	Beats-Per-Minutes
MeanRR	Average duration of RR intervals
SDRR	Standard deviation of RR intervals
NN50	Number of successive RR intervals that differ by more than 50ms
pNN50	Percentage of successive RR intervals that differ by more than 50ms
RMSSD	Root mean square of successive RR interval differences
SDRM	Ratio between SDNN and RMSSD
TrInd	Integral of the density of the RR interval histogram divided by its height
TINN	Baseline width of the RR interval histogram
TP	HRV total power
VLF	Absolute power of the very-low-frequency band (0.0033-0.04Hz)
LF	Absolute power of the low-frequency band (0.04-0.15Hz)
HF	Absolute power of the high-frequency band (0.15-0.4Hz)
VLFpeak	Peak frequency of the VLF
LFpeak	Peak frequency of the LF
HFpeak	Peak frequency of the HF
VLFprc	Relative power of the VLF band
LFprc	Relative power of the LF band
HFprc	Relative power of the HF band
nLF	Relative power of the LF band in normal units
nHF	Relative power of the HF band in normal units
LFHFratio	Ratio of LF-to-HF power

3.2.7 Behavioural measures

Two custom C# codes were developed to sample at 60Hz and log in a spreadsheet file:

- The in-VR movement trajectories of participants, i.e., the x- and z-coordinates of the fixed location visited during the whole experience.

- The Inertial Measurement Unit (IMU) data extracted directly from VIVE headset. Specifically, the x-, y- and z-coordinates as well as yaw, pitch and roll angles.

Movement trajectories were analysed throughout the experience, in order to find possible relationship between participants' physical behaviour and IPQ responses (which indeed underlie the whole VR scenario). A set of 9 parameters is extracted. Table 3.5 displays them, along with a short description.

Table 3.5: The list of features extracted from in-VR movement trajectories.

Feature	**Description**
N_{stop}	The total number of visited spots
T_{mean}	The average time spent in each spot
T_{std}	The standard deviation of the time spent in each spot
T_{max}	The maximum time spent in a spot
$Pos_{T_{max}}$	The spot where the maximum visiting time occurs
$Dist$	The total travelled distance
V_{mean}	Mean velocity
V_{std}	The standard deviation of velocity
V_{max}	Maximum velocity
T_{w2c}	The time spent before reaching the chapel gates from the warning
V_{w2c}	The velocity of the player to reach the chapel's gates from the warning
T_{fire}	The total time spent experiencing the wildfire

Notably, the last three metrics were computed for the wildfire stage only and T_{w2c} and V_{w2c} were normalised over the number of travelled spots from the warning to the chapel.

Regarding IMU data, due to the teleportation system, only the head's rotations carried relevant information about attentional focus and thus were considered for further investigation.
Pitch, yaw and roll angles were preprocessed separately with the same pipeline: firstly they were filtered with a 4-order low-pass Butterworth filter with a 6Hz cut-off frequency, then they were resampled at 10Hz. A set of 7 features was extracted per angle time series. Table 3.6 reports them alongside a brief explanation.

Table 3.6: List of features extracted from IMU pitch, yaw and roll angles.

Feature	**Description**
Vel_{mean}	Mean angular velocity
Vel_{std}	Standard deviation of angular velocity
Range	Angular range
powerST	Spectral power in the 0-0.3Hz bandwidth (for slow, tracking movements)
powerFDG	Spectral power in the 0.3-1Hz bandwidth (for fast, fidgety movements)
N_{peaks}	Number of peaks in angular velocity
Tot_{angle}	Overall angular displacement

3.2.8 Statistical analysis

Data normality was checked with a Shapiro-Wilk test, setting the α-level of significance at 0.05.
This assumption was verified for all physiological parameters, therefore, data from the VR experience was analysed through a repeated-measures ANOVA (RANOVA) framework, specifying the within-subject factor as the four separate phases of the experiment. Notably, the p-value calculations in the RANOVA test are accurate if the theoretical distribution of the response variables has compound symmetry. This reflects a specific pattern of correlation where all variances, as well as covariances, are equal. If this assumption is false, then the degrees of freedom for the RANOVA test must be adjusted by a factor ϵ, and the p-value must be computed using the adjusted values. Mauchly's test of sphericity tests the null hypothesis that the sphericity assumption is true for the response variable ($\alpha = 0.05$) and returns the ϵ adjustment computed with the Greenhouse-Geisser and Huynh-Feldt techniques. Generally, when $\epsilon < 0.75$ the Greenhouse-Geisser correction should be considered, and vice versa. Moreover, if the p-value (whether corrected or not) is significant, multi comparisons were also performed, applying the Tukey-Kramer correction.
Regarding behavioural data, a correlation analysis was carried out between the in-VR movement parameters (as extracted from the entire experience) and the IPQ items using the Kendall τ. Moreover, it was explored whether demographic factors influenced participants' attitude while performing the VR experience. Specifically, the sample was divided in two population according to the spoken Language (European Portuguese vs English), nationality (Portuguese vs international citizens), age (the overall mean age was

used as a threshold, not considering gender), gender, VR familiarity (which considered expert users participants with a Likert-scale value ≥ 3), previous visits at São Leonardo de Galafura. Again, after checking for behavioural data normality, each metric underwent either a t-test or a Mann-Whitney U test ($\alpha = 0.05$) to discover possible differences.
CCAS questionnaire items are ordinal data, therefore, the potential attitude shift after being exposed to the VR experience was investigated with a Wilcoxon sign test ($\alpha = 0.05$). Moreover, the influence of the above mentioned demographic factors was also investigated. After sample split, CCAS answers were explored within group, in order to remove possible biases due to the counterpart (i.e., to analyse the shift only in the male or the female population). A Wilcoxon sign test was performed between pre- and post-VR conditions ($\alpha = 0.05$). Finally, regarding IMU data, a correlation analysis was performed between each angular metric and the quiz score, using Spearman technique to compute the correlation coefficient.

3.3 Results

The narrative and mechanics of the developed VR experience were assessed through the IPQ answers, grouped according to the different subscales illustrated in Table 3.2. Their median values and interquartile ranges (IQR) are displayed in Table 3.7.

Table 3.7: IPQ subscale median and IQR (in parentheses). OSP = Overall Sense of Presence; SP = Spacial presence; INV = Involvement; REAL = Experienced Realism.

OSP	**SP**	**INV**	**REAL**
4 (1)	4(1)	3.25 (1.625)	3 (1)

The influence of demographic factors over individual IPQ items showed that previous visits at the real viewpoint, gender and language/nationality determined statistically significant differences regarding spacial presence (Q5), involvement (Q7 and Q8) and experienced realism (Q12), as the boxplots in Figure 3.11 illustrate.

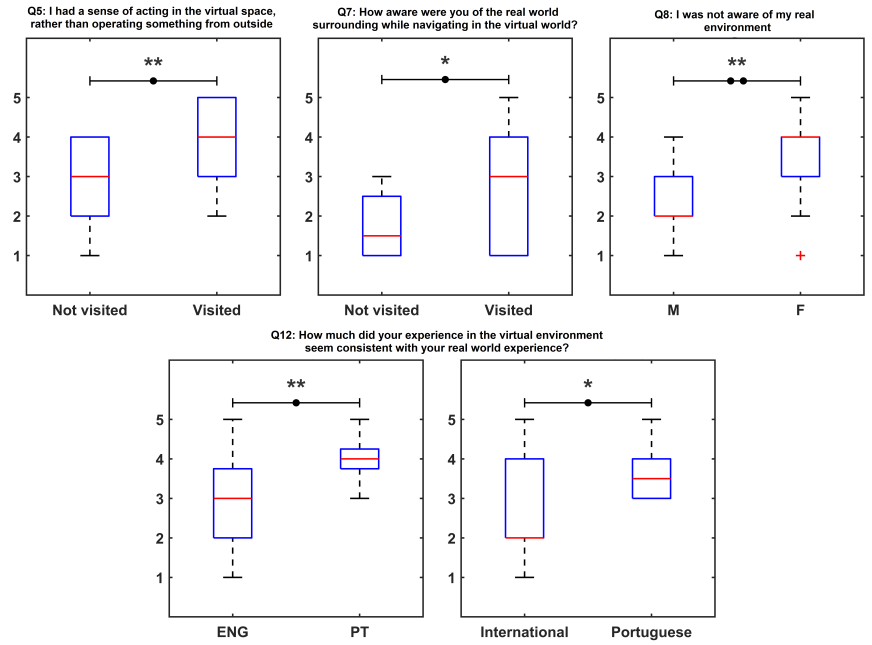


Figure 3.11: Distribution of IPQ individual items that presented significant differences according to demographic factors. A (•) represents a difference with $p < 0.05$, (••) $p < 0.01$. A (*) refers to a small Cliff's Δ effect sizes, whereas (**) to a medium one.

The statistical analysis on the CCAS answers showed that seven items, encompassing both beliefs (5) and intentions (2) dimensions, were significant between the pre- and post-VR condition. Figures 3.12-3.14 display their distribution rendered with boxplots on the left side and histograms on the right side, to better highlight the shift in the scoring result (especially for Q12 in Figure 3.14).

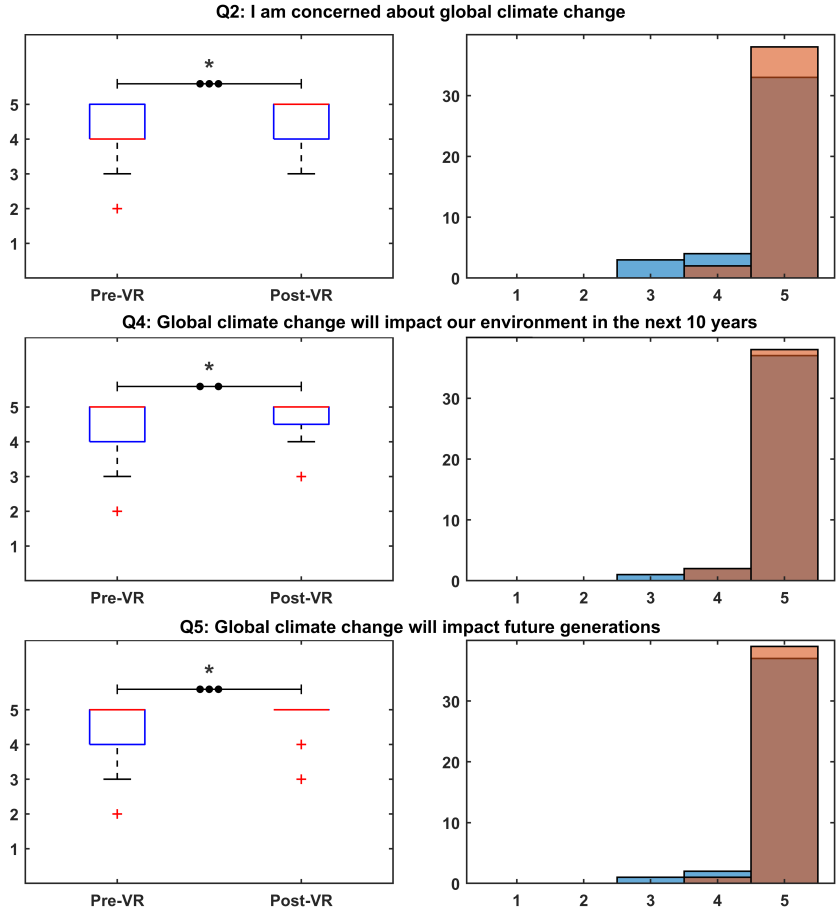


Figure 3.12: Distribution plots for statistically significant items Q2, Q4 and Q5 of the CCAS questionnaire. A (•) represents a difference with $p < 0.05$, (••) $p < 0.01$, and (• • •) $p < 0.001$. A (*) refers to a Cliff's Δ small effect.

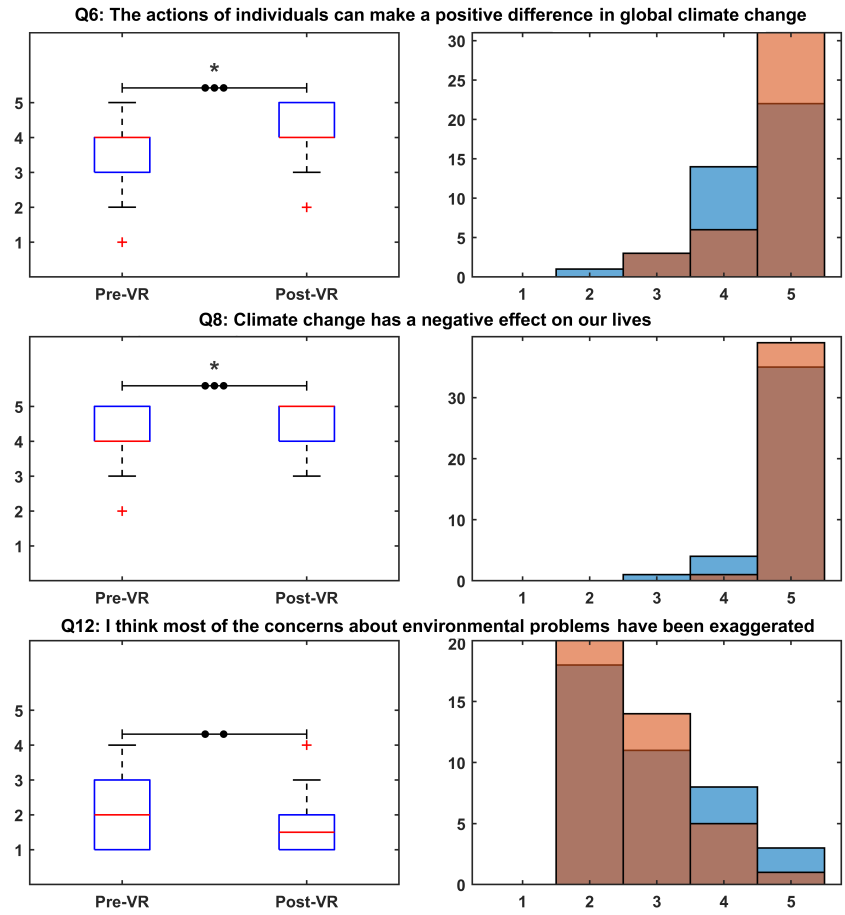


Figure 3.13: Distribution plots for statistically significant items Q6, Q8 and Q12 of the CCAS questionnaire. A (•) represents a difference with $p < 0.05$, (••) $p < 0.01$, and (• • •) $p < 0.001$. A (*) refers to a Cliff's Δ small effect.

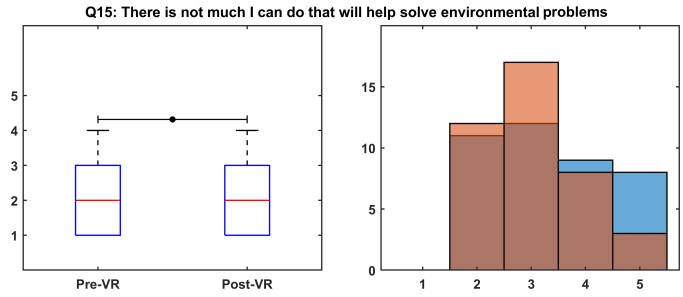


Figure 3.14: Distribution plots for statistically significant item Q15 of the CCAS questionnaire. A (•) represents a difference with $p < 0.05$.

The first sub-analysis on the CCAS scores highlighted that even if item responses did not change globally, when dividing the sample according to the demographic factors, additional statistical differences were detected within groups. For instance, CCAS's Q1 was not significant between pre- and post-VR condition ($p = 0.07$), but when considering the younger population only, the difference became significant ($p = 0.03$). Table 3.8 reports the result for such exploration.

Table 3.8: CCAS differences within groups divided per item between the pre- and post-VR experience. A ↑ represents an increase in score value from pre- to post-VR, vice versa for ↓. A (•) represents a statistical difference with $p < 0.05$, (••) $p < 0.01$, and (• • •) $p < 0.001$. PT = Portuguese; ENG = English; P = European Portuguese; I = International; Y = Younger; O = Older; M = Male; F = Female; E = Expert; NE = Not Expert.

	Language		**Citizenship**		**Age**		**Gender**		**VR**		**Visit**	
Item	PT	ENG	P	I	Y	O	M	F	E	NE	Yes	No
Q1					↑•							
Q2	↑•••		↑••		↑••		↑••	↑•		↑••		↑•••
Q4	↑•		↑•		↑••		↑••					↑•
Q5	↑•••		↑••		↑••		↑••			↑•		↑••
Q6	↑•••		↑•••		↑••	↑•	↑•••		↑••	↑•		↑•••
Q8	↑•••		↑•		↑••		↑••			↑•		↑•••
Q9	↓••		↓•		↓••							
Q12	↓•		↓•		↓•		↓•					↓••
Q15						↓•	↓•					

Regarding the physiological data, the RANOVA analysis highlighted statistically significant differences across the VR stages for one HRV metric, i.e., the NN50. Subscript GG refers to the Greenhouse-Geisser adjustment.

- NN50: $p_{GG} = 0.04$. As supported by Figure 3.15, the post-hoc comparison highlighted difference between BSL and the phases of the VR experience: GT ($p^{BSL \leftrightarrow GT} = 0.002$, $d = -0.64$), WF ($p^{BSL \leftrightarrow WF} = 0.04$, $d = -0.43$), and CRC ($p^{BSL \leftrightarrow CRC} = 0.02$, $d = -0.49$).

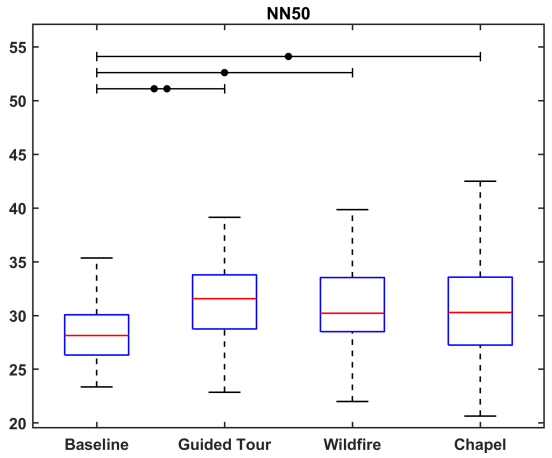


Figure 3.15: NN50 across the baseline and the three stages of the VR experience. A (•) represents a difference with $p < 0.05$, (••) $p < 0.01$.

Moreover, other 3 parameters extracted from EDA proved to be statistically significant across the VR experience stages. Specifically:

- SCLmean: $p_{GG} < 0.001$. The multiple comparison analysis further discovered significant differences between BSL and GT ($p^{BSL \leftrightarrow GT} < 0.001$ and $d = -1.57$), WF ($p^{BSL \leftrightarrow WF} < 0.001$ and $d = -2.55$), and CRC ($p^{BSL \leftrightarrow CRC} < 0.001$ and $d = -2.50$). Moreover, SCLmean differed between GT and both WF and CRC ($p^{GT \leftrightarrow WF} < 0.001$ and $d = -1.56$, $p^{GT \leftrightarrow CRC} < 0.001$ and $d = -1.07$). Figure 3.16 displays the SCLmean data distribution.

- SCRn: $p < 0.001$. The post-hoc analysis, supported by Figure 3.17, highlighted significant differences between BSL and each VR stage

($p^{BSL\leftrightarrow GT} < 0.001$ and $d = -1.50$, $p^{BSL\leftrightarrow WF} < 0.001$ and $d = -1.19$, $p^{BSL\leftrightarrow CRC} < 0.001$ and $d = -1.06$), as well as between WF and CRC ($p^{BSL\leftrightarrow WF} = 0.044$, $d = 0.44$).

- EDAsymp: $p_{GG} < 0.001$. The post-hoc comparisons discovered significant differences between BSL and each VR stage ($p^{BSL\leftrightarrow GT} < 0.001$ and $d = 1.91$, $p^{BSL\leftrightarrow WF} < 0.001$ and $d = 0.57$,$p^{BSL\leftrightarrow CRC} < 0.001$ and $d1.57$), between GT and both WF and CRC ($p^{GT\leftrightarrow WF} < 0.001$ and $d = -0.99$, $p^{GT\leftrightarrow CRC} < 0.001$ and $d = -0.22$), and also between WF and CRC ($p^{WF\leftrightarrow CRC} = 0.007$ and $d = 0.80$). Figure 3.18 shows the boxplots for this parameter.

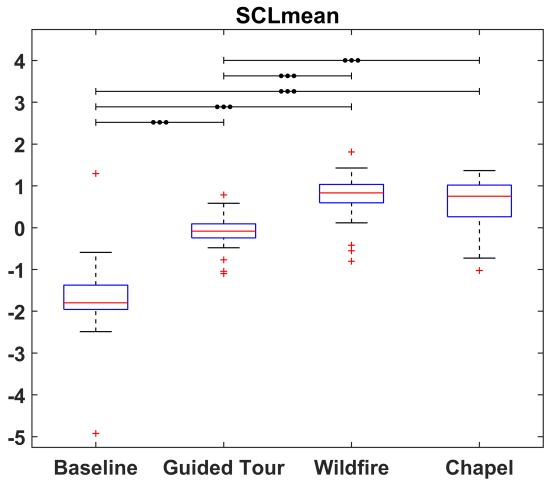


Figure 3.16: SCLmean across the baseline and the three stages of the VR experience. A (• • •) represents a difference with $p < 0.001$.

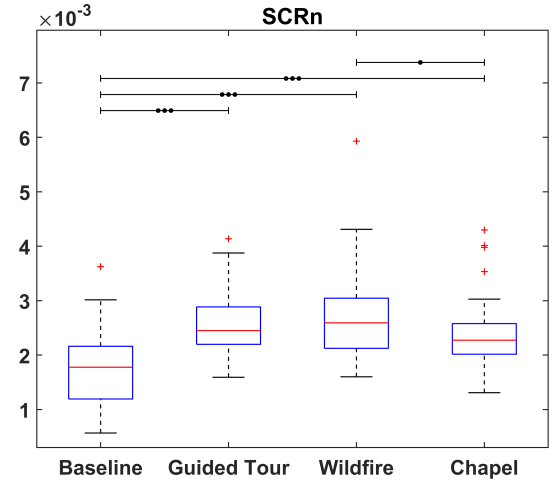


Figure 3.17: SCRn across the baseline and the three stages of the VR experience. A (•) represents a difference with $p < 0.05$, and (• • •) $p < 0.001$.

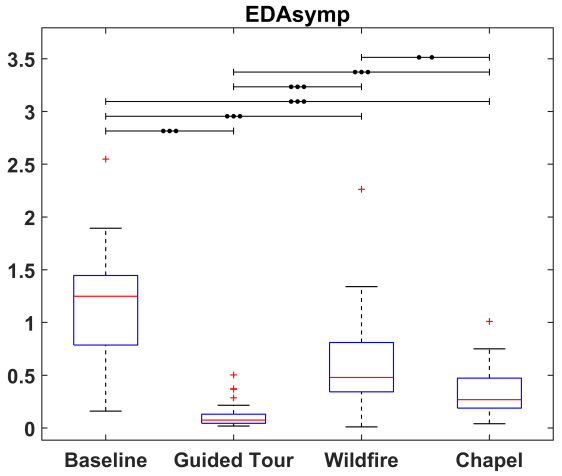


Figure 3.18: EDAsymp across the baseline and the three stages of the VR experience. A (••) represents a difference with $p < 0.01$, and (•••) $p < 0.001$.

The demographical factors also affected the in-VR movements, leading to statistically significant differences between groups. This is showed in Figure 3.19.

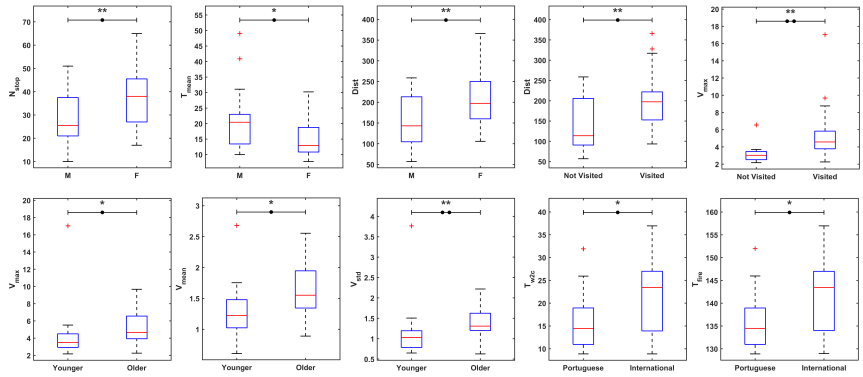


Figure 3.19: Distribution of in-VR movements parameters considering demographical factors. A (•) represents a difference with $p < 0.05$, and (••) $p < 0.01$. A (*) refers to a Cohen's d small effect, whereas (**) to a medium effect.

The correlation analysis between the IPQ items and virtual movement data also highlighted significant relationships, as shown by the scatter plots in Figure 3.20.

Finally, regarding the other behavioural measure, i.e., head rotations, only one significant relationship was found between the quiz score and yaw range. The scatterplot is displayed in Figure 3.21.

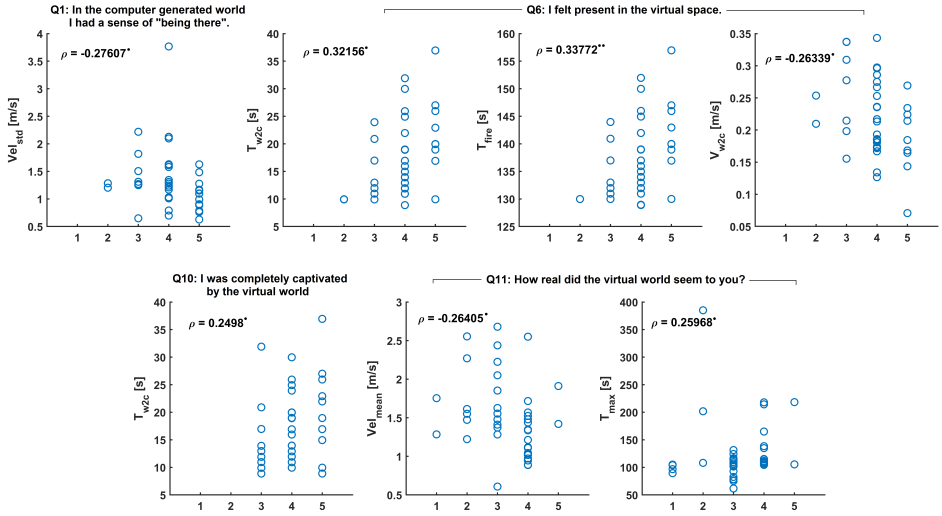


Figure 3.20: Correlation analysis between IPQ items and in-VR movements parameters. A (•) represents a significant correlation with $p < 0.05$, whereas (••) a $p < 0.01$.

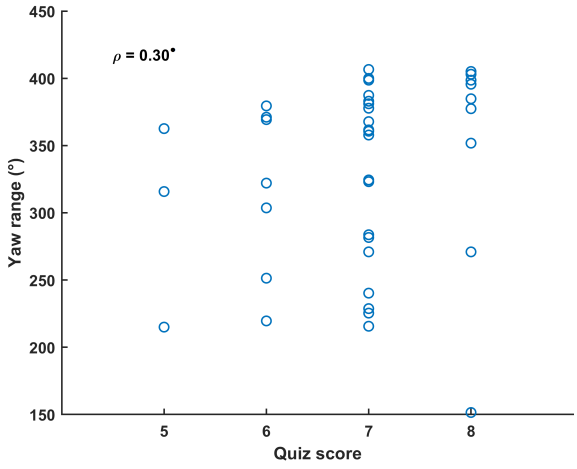


Figure 3.21: Relationship between the custom wildfire quiz score and the overall yaw rotation range during the audiovisual presentation occuring within the chapel. A (•) represents a significant correlation with $p < 0.05$.

3.4 Discussion

Immersion, Presence and Climate Change Attitudes

The development of an effective VR experiences presents a unique set of challenges, particularly when the goal is to promote a behavioural change, such as fostering better attitudes toward climate change. A successful VR environment must balance three key elements: immersion, involvement, and realism. Immersion requires not only the right technological setup to isolate participants from external distractions but also the careful selection of stimuli that emulate a coherent and compelling virtual world. Involvement depends on a narrative that is both emotionally engaging and grounded in real-world issues, ensuring that users feel personally connected to the unfolding experience. Finally, realism demands that audiovisual cues, and if possible, other sensory modalities, closely mirror those encountered in real life, thus maintaining a sense of authenticity.
These aspects are not merely aesthetic. They are instrumental in ensuring that the intended impact of the VR scenario is achieved. Here, the VR experience aims at reducing the value-action gap, i.e., the disparity between individuals' expressed beliefs and their actual behaviours in sustainability contexts. If the narrative feels contrived, the mechanics disjointed, or the environment unconvincing, the user's suspension of disbelief may falter, ultimately weakening the intervention's capacity to drive a meaningful behavioural shift.
The IPQ is a valuable and well-established survey that is commonly used to evaluate the sense of presence, i.e., the combination of immersion and engagement. The IPQ subscales demonstrated that the developed VR experience presented a good sense of presence (Table 3.7), confirming the results from a preliminary study [77]. This last result is particularly relevant because, even if participants were not allowed to move freely in the virtual scenario, the teleportation system did not hinder the sense of "being there". On the other hand, considering the *INV* subscale, a lower median score of 3.25 was achieved, highlighting that some characteristics should be addressed to improve such an aspect. Indeed, the lower values related to how the participants were aware about the surrounding real world (item 7 and 8 from Table 3.2, where the median values were 3 for both responses). This outcome may have been determined by the fact that the experiments took place in laboratory settings without the proper soundproof barriers. Therefore, not

inherent acoustic stimuli may have been perceived. Moreover, the experience provided audiovisual cues only: the inclusion of olfactory or haptic stimuli (e.g., the smell of smoke, temperature) could have improved the level of engagement. However, these precautions might have been insufficient, taking into account that the HTC VIVE Pro 2 presented a cabled connection. Indeed, the wire linking the headset with the desktop computer intermittently touched participants' body and arms, as well as forcing them to take unspontaneous turns to reach a desired location in the virtual scenario, contributing to continue perceiving the surrounding real world. The *REAL* scored the lowest, with a median value of 3. While participants seemed satisfied with the degree of realism of the experience, item 12 and 14 highlighted additional issues. As mentioned earlier, a multisensory environment could have helped participants to feel more immersed, especially when living the simulation of the wildfire. Nevertheless, item 12 answers might be related to the duration of the natural disaster. This phase was designed to last 2min in order to have physiological time series of acceptable length. Yet, since participants cannot escape such an event, this may go against their instinctual response and determine a lower consistency value between virtual and real world behaviours. Interestingly, some of the IPQ items seemed to be mediated by demographical factors, as supported by Figure 3.11. Indeed, previous visits at the São Leonardo de Galafura viewpoint determined a stronger sense of acting but lower engagement. This may be explained in the sense that, participants that already went seeing the place were curious to understand whether the scenario matched their acquired knowledge and memories. Simultaneously, they might have felt less engaged because they already knew the information provided in the guided tour. Also, female participants were much more involved than male ones, agreeing with several evidences pointing out their greater concern about climate change [117]. Finally, geographical proximity seemed to have triggered a higher coherence degree between virtual and real environment. Although not all Portuguese participants visited the location (only 20% did), they may have recognised a familiar landscape where they would assumed similar attitudes and actions to those proposed in the VR scenario. This result also strongly highlights the importance of addressing the spacial distance factor in order to create a meaningful and effective experience [91, 92].

Despite its limitations, the generally positive evaluation of the experience indicates its potentiality in addressing the action-value gap. The CCAS, the

questionnaire administered to collect the opinions and, most importantly, the intention toward climate change, proved to be a successful tool to investigate the attitudinal shift before and after the VR exposure. Indeed, as hypothesised by [5, 83], experiencing a natural disaster (even if simulated) led to both increased awareness and worrying. Moreover, being informed about the preventive measures that can be individually taken for protection, positively affected future actions toward eco-friendliness, aligning with [88]. Interestingly, items Q4 and Q5 specifically mention the impact of climate change on future generation, this is why it was decided to analyse the influence of demographical factors also concerning the CCAS.
As Table 3.8 shows, the experience was particularly efficient for younger subjects, helping them to better acknowledge the occurring change in climatic patterns and its dangerous consequences (Q1, Q8), that will especially fall upon them (Q4, Q5), thus raising their concern about the topic (Q2, Q12). Furthermore, it seems that the experience also supported them to recognise their role and the importance of their actions when adopting sustainable behaviours that can mitigate the negative effects of climate change (Q6, Q9). The analysis of the CCAS values within groups also suggested that the experience was more effective for Portuguese citizens. This result should not be interpret as the fact that such a population is sceptical or unaware of climate change (on the contrary, a recent survey sponsored by the European Union proved the opposite [118]), but, rather, they were more impacted by the destruction of a nearby, popular setting. Interestingly, this could agree with the heightened worry of people who has not visited São Leonardo de Galafura, probably because they have realised that, with wildfires becoming more severe and threatening, they might not be able to visit the place in the future. Moreover, the VR experience helped male participants to become more aware of the issue, matching their female counterpart in multiple aspects. Finally, Table 3.8 also highlights that VR familiarity seems not to downgrade the positive attitudinal shift, possibly underlining the broader applicability of such an approach.

Real and virtual movements analysis

Subjective measures like the IPQ or the CCAS provide immediate and accessible evaluations of the VR experience, but they also present major drawbacks. First of all, they rely on individuals' self-introspection, which may be

biased by psychological phenomena or might not be easily performed for certain topics [11]. Thus, the need for alternative assessment tools determined an increase of VR studies that simultaneously acquired behavioural patterns and physiological activity to examine and validate the proposed experience. Here, behavioural data consisted in body and head movements within the VR experience. In-VR trajectories were not segmented but considered as a whole: indeed, although the path users chose to reach the four pre-fixed locations may slightly differ (e.g., some may decide to take several smaller steps rather than trying making fewer and longer "jumps"), their contribution could be marginal. A similar consideration can be done for the chapel part. Indeed, inside the building, only one portal is provided, thus interrupting VR-movements. Hence, the latter were analysed in relation with demographical factors and the IPQ items. The first study highlighted that gender influenced the degree of the scenario exploration, with female participants being more curious than males (higher number of stops, less time spent per stop and greater travelled distance, Figure 3.11). Exploration behaviour was understandably affected by previous visits of the viewpoint. As a matter of fact, people who already visited the setting explored more and moved faster, possibly to see how the virtual environment matched their acquired memories. Interestingly, younger people had a steadier and calmer inspection of the scenario with respect to older participants, possibly to better examine its details and evaluate its realism. This behaviour could also be explained by the fact that most of the younger players were students of Informatics Engineering and had greater knowledge of VR. Another relevant result concerns the interaction between citizenship and the time people were immersed in the wildfire. Since Portuguese participants took more time to reach the chapel with respect to foreign ones, this might mean that the first were more involved and possibly overwhelmed by the natural disaster, leading them to a slower response to the danger message. This again supports the idea that a VR experience to be successful should be set in environments that are closely located to participants nationality, also encompassing congruent events that may endanger them. Moreover, the correlation analysis between in-VR movements and IPQ highlighted that participants that felt more present and were more captivated by the environment presented higher dwelling times and slower velocity while reaching the chapel's gates (Q6 and Q10, Figure 3.20). The *OSP* was also significantly related to the standard deviation of velocity: players moved with a more stable behaviour, poten-

tially helping them to delve deeper into the experience and match a pace that they would adopt to explore an unknown place in real life. Regarding *REAL*, a slower average velocity, also supported by longer idling times, helped participants to appreciate the details of the experience.
The other behavioural measure was head rotations, divided in pitch, yaw and roll angles. They were related to the quiz score, hypothesising that attention and focus could benefit from participants' slower and steadier movements. However, correlation analysis only highlighted a moderate relationship between quiz performance and overall yaw range, suggesting that a wider scanning of the presentation slides guided players to gather more information and better follow the audio contents, thus achieving a higher score. This outcome may indicate that head rotations might be unsuited for such an analysis and that finer movements, as the ones extracted from eye-tracking devices, are required to associate the physiological activity underlying cognitive mechanisms and human performance [119, 120].

Physiological Introspective

In-VR movements offer a more objective framework to study participants reactions, the parameters might be too coarse-grained to efficiently describe them. However, biomedical signals provide the possibility to extract valuable information that underlie and describe the sequence of events embedded in the experience with better precision. They can grasp the spontaneous responses related to individual stimuli, which can be extremely more challenging when using questionnaires or behavioural measures. Notably, physiological metrics provide access to covert cognitive and emotional processes that are not necessarily expressed through movement. Indeed, a participant might appear passive while listening to the radio interview, yet maintain elevated EDA or HRV levels, which are indicative of sustained arousal. Literature showed that a strikingly small set of works included a deeper investigation of climate change centred VR experience combined with physiological monitoring. Therefore, in this study, ECG and EDA were recorded during the whole duration of the trial and used to parametrise the physical responses. Four metrics brought evident changes in physiological arousal while living the different stages of the VR experiment with respect to the baseline, out of VR, condition (as Figures 3.15-3.18 illustrate). Specifically, regarding HRV, the NN50 showed a significant increase from the baseline to each section of

the experience. The greater difference stands between the resting state and the GT phase, suggesting an elevated parasympathetic activity and possibly a sense of interest and safety. Therefore, even if the guided tour did not present any particular stimuli that may trigger strong emotional responses, the audiovisual rich environment determined a significant alteration from the baseline, supporting the outcome from the IPQ. Although not significant, Figure 3.15 shows a slight decrease from GT and WF, which may represent a complex emotional state combining alertness with situational awareness. Interestingly, a similar median NN50 value was found between WF and CRC phases. This could mean that, despite virtually safe from the fire, participants' psychophysiological state remained altered, reflecting an increased cognitive workload potentially associated with information processing. EDA parameters provided further differentiation, also across VR phases. SCLmean exhibited an increase from baseline to all subsequent phases, with statistically significant contrasts across nearly all transitions. This trend suggests a clear elevation in sympathetic arousal, consistent with the escalating narrative intensity. Importantly, the WF phase elicited the highest arousal, reflecting a stronger activation that is coherent with possible stress, fear or alertness provoked by the simulated flames. The CRC phase maintained an elevated SCLmean, marking sustained engagement during the post-crisis sheltering period. an incomplete recovery. This was not an unexpected outcome as indeed participants heard about stress-inducing information (and still listened to background fire noises). Moreover, their cognitive load could have been taxed by learning additional coping strategies. Therefore, SCLmean appears to be a promising parameter that can detect both the transition between out- and in-VR conditions, as well as being sensitive to the different arousal levels across phases. Figure 3.17 showed that SCRn followed a similar trajectory to that of SCLmean, significantly rising from baseline through WF, with a slight, yet significant decline in the CRC. This dynamic may indicate an acute phasic activation during the progressive unfolding of the wildfire, potentially built up by the fast sequence of the selected stimuli, followed by partial recovery of the autonomic system. The virtual safety provided by the closed environment and the inclusion of details about mitigation strategies could have supported emotion regulation, also illustrated by the maintenance of high NN50 levels. Finally, EDAsymp displayed a pronounced value during the baseline period, where subjects are supposed to be relaxed. Such an unexpected result could be a direct response

to the activity of the baroreceptors, given that during baseline, participants were seated, whereas in the VR condition, they stood and moved. However, NN50 also showed a lower value in the same section, both features could indicate that the sample may not have been truly at rest. Indeed, they might have been excited or anxious about experiencing VR, wearing the headset, or using biomedical devices. These biases are especially common in studies involving immersive technology, where pre-task apprehension can lead to elevated sympathetic tone. Nonetheless, similarly to SCLmean, EDAsymp proved to be consistently and coherently able to separate the individual sections. The WF was characterised by a high spectral power, reflecting increased sympathetic arousal, which is congruent with fear responses. This finding could be an encouraging index of the immersive realism of developed VR experiences. Finally, the CRC-related EDA power significantly decreased from WF and, simultaneously, was larger than that of GT, highlighting again only a partial recovery due to a continuous arousal tapping. This indicates a successful design of the CRC section, as it was indeed intended to be a cognitively complex moment involving both stress processing and regulatory engagement, with subjects being not passive viewers but, rather, engaged, attentive listeners (which was also supported by the median quiz score of 7/8).

Overall, multisource data (subjective, behavioural, and physiological) offered a comprehensive framework for evaluating the robustness and the impact of an immersive climate-related VR experience.

3.5 Conclusion

This chapter investigated the potential of a VR experience to reduce the value-action gap in the context of climate change. The VR environment, designed to simulate a wildfire near a culturally and geographically significant location in northern Portugal, aimed to foster deeper cognitive and emotional engagement through a carefully constructed combination of immersion, involvement, and realism. The evaluation of the experience using the IPQ revealed a generally strong sense of presence, with particularly positive results in spacial immersion, despite the limitations posed by laboratory settings, lack of multisensory cues, and cabled VR equipment. However, the *INV* and *REAL* subscales presented areas for improvement, highlighting the

need for reducing real-world distractions and enhancing sensory feedback. Demographic factors played a significant role: female participants exhibited greater involvement, while Portuguese users and those familiar with the São Leonardo de Galafura viewpoint demonstrated more consistent and place-based reactions, emphasising the importance of geographical and cultural proximity in VR design procedure.

The CCAS, administered pre- and post-experience, confirmed the effectiveness of the VR exposure in promoting attitudinal shifts. The simulated disaster increased awareness, concern, and perceived personal agency, especially among younger participants. Portuguese users, possibly due to their proximity to the threatened location, and male participants, who initially showed lower engagement, also exhibited significant gains. These results affirmed previous findings that VR can serve as a transformative tool in environmental communication. Notably, prior VR experience did not diminish the intervention's impact, suggesting wide applicability across different user profiles.

Beyond self-report measures, behavioural and physiological data provided objective insight into participants' engagement. In-VR movement patterns correlated with subjective presence scores and demographic characteristics, reinforcing the validity of spacial behaviour as a proxy for immersion. Furthermore, physiological monitoring through ECG and EDA detected clear differences between the VR experience's phases. Particularly, sympathetic nervous system activation increased during the wildfire simulation, reflecting emotional arousal, and remained elevated during the subsequent coping content phase, indicating prolonged cognitive processing.

Chapter 4

Physiological correlates of environmental implicit attitudes

As explicit questionnaires are exposed to several psychological biases, an investigation over implicit attitudes that does not require conscious introspection is necessary to evaluate the action-value gap from an alternative point of view. In this chapter, the Single Category Implicit Association Test was used to analyse the decoupling between implicit and explicit attitudes about climate change in 20 participants. Moreover, since traditional methods have been long criticised, an experimental paradigm that also implemented eye tracking technology was developed. While no correlation was found between implicit and explicit measures, linear and nonlinear parameters of eye movements dynamics (gaze, pupillometry and blinking) highlighted that making incongruent associations that goes against an individual's automatic beliefs determined an increase in cognitive load. Moreover, it revealed that repeated visual patterns, longer fixations and reduced blinking activity may be physiological modulating processes that slow down the decision-making process. [1]

[1] A preliminary analysis has been published as "Eye tracking correlates of the Implicit Association Test" in the *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* [121]. The development of the platform used to deliver the experiment and the data collection campaign occurred before the beginning of this research activity, and the doctoral candidate was involved only at the latest stages of its deployment [122], as well as for data processing and analysis.

4.1 Introduction

The public perception of climate change is an ambiguous topic, and it has experienced strong fluctuations over time, possibly due to political opinion polarisation and social media misrepresentation [123]. In their review, Capstick *et al.* [123] stated that after an increasing awareness of climate change between the 80s and the 90s, which encouraged policymakers to make relevant environmental decisions (e.g., CFC banning), survey responses in 2010 years underlined more scepticism and a decline in climate change concern, despite the growing scientific consensus about its anthropogenic causes. Environmental psychology has tried to explain this issue, nevertheless, their results were equivocal [124]. More recently, protests organised by younger generations have revived interest in this topic [125], and several studies have supported this positive trend [126, 127], highlighting a new change in public opinion.
To understand the degree of engagement in adopting environmentally-friendly behaviours, it is essential to analyse how people perceive climate change as real, believe it is human-caused and has negative consequences [1]. Behaviour is shaped by attitudes, and the latter can be defined as a summary of evaluations, driven by emotions, experiences, and acquired knowledge of social objects (e.g., people or policies) along a dimension ranging from negative to positive [128]. They are typically distinguished in:

- Explicit attitudes consist of conscious concepts that an individual is fully aware of and willing to report. They are assessed with questionnaires and surveys. In the context of environmental sustainability, several scales have been proposed, such as the New Ecological Paradigm [129] or the Beliefs about Global Climate Change (BAGCC) [130].

- Implicit attitudes refer instead to automatic evaluations that are not accessible through conscious introspection. Therefore, indirect strategies have been developed for their quantification. Studying implicit attitudes is critical, mainly when socially sensitive topics such as climate change are investigated, to overcome the issue that various psychological effects can alter openly expressed points of view, e.g., social desirability or recall biases [11]. Hence, scale scores may not reflect the full complexity of a person's attitude. These phenomena also negatively affect possible relationships between implicit and explicit attitudes, which is actualised in typical low correlation values or divergent

results [131].

The Implicit Association Test (IAT) represents one of the most common computer-based tools for assessing automatic concept-attribute associations underlying implicit biases, prejudices, and beliefs [11]. The core idea behind the IAT resides in measuring differences in decision-making reaction times between "congruent" and "incongruent" associations, namely the IAT effect. Concept congruency is defined based on the type of subjects to test and the attributes are adjectives with positive (e.g., good) and negative (e.g., bad) valence [11]. Even if the reasons for this delay remain unclear, two hypotheses can be made: either participants perform additional mental processes to inhibit their own automatic, stereotypical evaluations, or they perform the same mental processes but require more time to be completed (e.g., for selecting a motor response) [132]. The IAT has been applied to study various topics, such as physical as well as cognitive disability, sexism and racism [133–135]. These studies highlighted negative implicit attitudes toward determined groups of people, which helped to increase the awareness about implicit prejudices and the need to counteract them, therefore fostering the development of diversity training programs and campaigns to promote integration and improve quality of life in work environments and healthcare.
Implicit attitudes regarding climate change perception and environmental sustainability have also been assessed. In a study investigating the relationships between plastic practicality and pollution, participants tended to implicitly rate plastic waste more negative and riskier than plastic packaging and microplastics, but this finding did not agree with explicit attitudes [136]. Similarly, Songa *et al.* found that, despite good implicit attitudes toward recyclability, no significant correlation was found with explicit attitude regarding the same behaviour [137]. This ambiguity characterises several other works that attempted to correlate the outcome of the IAT with explicit answers. Indeed, these two dimensions were not significantly related in company executives [138] and low-carbon footprint shoppers [139], but they were when the public opinion toward an innovative drainage system was probed in a specific population where such implant was about to be installed [140]. Notably, the decoupling between implicit and explicit attitudes represents one of the major drawbacks of implementing this experimental paradigm. Multiple sources of variability have been tested to clarify this point: the specific topic under exam, or the properties of the explicit and implicit mea-

sures (e.g., the influence of the compatible/incompatible blocks order) [131]. Moreover, research demonstrated the IAT effect's malleability, which can be faked and strategically modified [141, 142]. In a study, participants could successfully fake their personality trait in an extraversion IAT [143]. As this outcome was achieved without clear instructions regarding how to alter response times, great care toward problem contextualisation should be used when recruiting subjects. Additionally, other confounding factors have been recognised to facilitate or inhibit the reaction times that are used to compute the IAT effect and may negatively affect it: task-switching costs, processing speed (due to fatigue or stress, or even emotions) and/or cross-category associations [144].
These limitations suggest that a measure solely based on reaction times or reaction time-derived parameters such as the D-score [11] cannot detect and explain implicit attitudes. Further information must be integrated to obtain a complete insight into their origin and functioning.

Therefore, several studies proposed the implementation of biomedical devices to monitor physiological activity while completing an IAT to understand which cognitive mechanisms might account for the IAT effect.
Although non-invasive neuroimaging techniques such as EEG and functional magnetic resonance imaging (fMRI) have been fundamental, they are not without limitations. These methods often require specialised and expensive equipment, controlled laboratory settings, and can cause discomfort or restricted movement due to their immobile and sometimes intrusive setup [145]. Moreover, both EEG and fMRI can yield ambiguous measures of neural activity, also due to the various types of stimuli that can be used in experimental paradigms. For instance, EEG offers high temporal but limited spacial resolution, vice versa for fMRI, that relies on slow haemodynamic responses that may not be adequate to capture fast neural dynamics [146]. These constraints have spurred the exploration of alternative or complementary methodologies to study cognitive and neural mechanisms. One promising strategy is represented by eye tracking technology (ETT), that in fact offers a cost-effective, non-invasive, and highly portable solution for inferring cognitive states through oculomotor behaviour. Here, the eye movements are a means to access integrative visual and cognitive functions. Indeed, eye movements are tightly linked to attentional and decision-making processes, and their analysis can yield valuable insights into the underlying neural functions

of implicit attitudes.

4.2 Eye Tracking Technology in Studying Implicit Attitudes

Eye tracking has been widely implemented in implicit attitudes studies. When developing a VR application to rehabilitate heavy drinkers, it was found that implicit attitudes decreased after the VR exposure. Still, this change was not mirrored by different gaze patterns over alcohol-related visual cues [147]. A similar result was obtained when analysing how people with persistent low back pain perceive and react to triggering images of acts that may cause back injury (e.g., bending). By measuring latency and magnitude of the blinking reflex, no statistical difference was found when watching target pictures with respect to neutral, pleasant and unpleasant ones. Moreover, focusing on a subgroup with higher level of pain fear did not alter such perception, despite implicit attitudes confirmed a stronger association with target images and a sense of danger [148]. It was also investigated whether students' gender affects gaze behaviour of pre-service teachers: independently of adults' sex, both explicit and implicit attitudes were more positive toward girls. Even if this general outcome may have been driven by female teachers due to psychological biases (e.g., self-categorisation) and different sample sizes, gaze patterns actually highlighted that, without the mediation of participants' gender, the number of fixations was higher on girls than on boys and correlated significantly with the IAT D-score [149]. Overall, these findings suggest that more research is needed in order to prove the existence of a relationship between behavioural and objective metrics. Additionally, such studies advise for the implementation of biomedical devices in order to gain a wider perspective on it.
Regarding climate change-related topics, it was discovered that implicit attitudes predict spontaneous emotional reaction as measured with a larger pupil dilation. Interestingly, a higher fixation time over recyclability logos also mediates positive implicit attitudes and positive facial expression toward commercial goods, determining an increase in the correlation coefficient between the two variables [137]. Moreover, participants with higher implicit attitudes required less time to focus on the area of interest (AOI) containing the logo. It was also discovered that, when presented simultaneously with other neutral and positive images, people tend to look at climate

change-related negative pictures less than 40% of the visual exploration time. However, participants with better implicit attitudes about carbon footprint showed a longer dwelling time over negative images of climate change, suggesting a correspondence between attitude and attentional focus [150]. Furthermore, it has been speculated that if the unconscious focus of attention leads to positive attitudes, and not the other way around, this may suggest a powerful opportunity: by increasing awareness of climate change damages, it could be possible to foster more positive implicit attitudes towards the environment [150]. Notably, this same behaviour was not found out when considering explicit attitudes.

However, these works have used the IAT outcome as a criterion to separate their samples and subsequently compare the gaze and pupil dynamics of the two populations rather than implementing ETT to study and explain implicit attitudes per se.

Only a few studies have applied ETT while participants were performing an IAT to account for its effect and provide more robust, physiologically explainable correlates that may characterise implicit attitudes. In a racism-based IAT, results highlighted that fixation duration, indicating participants' attention, was larger in trials that disagreed with their implicit negative preference (i.e., black/good) [151]. Additionally, the number of fixations demonstrated that associations confirming, or rejecting, one's implicit attitudes toward an out-group ethnic category are those that attract visual attention. However, this metric was significantly biased by an eye lateralisation effect (i.e., participants are incline to look first at the upper left part of a screen). The autobiographical IAT, an alternative version of the original paradigm typically used in forensic contexts, also proved that eye and pupil movements are altered when psychological attributes are incongruent with the individuals' automatic biases. It was discovered that the proportion of fixations in specific AOIs was higher in the incompatible block for the innocent group as the association did not match their memory trace [152], and vice versa for the guilty group. Notably, the difference in gaze parameters did not concern the overall scanning of the screen, but only the specific AOI containing the boxes with the category label to support the categorisation of the current IAT block. This behaviour may be interpreted as a response monitoring phenomenon, i.e., an effort to enhance response accuracy by repeatedly verifying the concept/attribution pair between the category label and the corresponding key to press. The same paradigm was deployed in

combination with pupil diameter measurements, which proved to be larger in the incompatible block compared to the compatible one [153]. This suggests a higher cognitive effort when making associations that go against their prior knowledge or previous actions. The efficiency of the pupil size difference was comparable to the traditional D-score in identifying innocents and guilty groups. This partially agrees with [152], where a surrogate measure based on fixation determined an increase of the sensitivity to detect guilty participants. This relevant finding advises for a multivariate approach to analyse and understand implicit attitudes. Objective measures derived from physiological reactions could represent a robust and reliable support to psychology research. Furthermore, such an outcome fosters the need to delve deeper into the psychophysiology of implicit attitudes by applying nonlinear algorithms and pipelines in order to better highlight the complex relationship underlying behaviour and information processing.

4.3 The current study

The objective of this study is four fold:

- To understand the degree of agreement between implicit and explicit attitudes toward climate change.

- To investigate the sensitivity of oculometric features, which were extracted from recordings acquired during a SC-IAT, to the paradigm switch. Specifically, the study aims at highlighting higher cognitive load and/or emotional engagement while making incongruent associations than making congruent ones.

- To evaluate modulatory phenomena that may explain the reason why making incongruent associations typically determines longer reaction times.

- To assess the feasibility of a physiologically-based, rather than a reaction time-based, metric that could help better analysing implicit attitudes.

4.4 Materials and Methods.

4.4.1 Participants and Experimental Procedure

Twenty healthy volunteers (13 females mean age = 23.9± 3.0 years, 7 males, mean age = 25.2± 1.9 years) were recruited for this study. Participants were seated in front of a 24-inch full HD monitor screen and a keyboard at a fixed distance of 30cm, where they first filled a form that collected demographical data (e.g., age, gender, highest achieved education degree). Then, they worn the Tobii Pro Glasses 3 (Tobii Pro AB, Danderyd, Sweden) for eye tracking measurement.
The smart glasses perform a video-based gaze reconstruction through eight infrared illuminators and two cameras per lens, with an accuracy of 0.6°. They implement the dark pupil technique in combination with corneal reflection detection, acquiring data at 50Hz.
An optical sensor-based triggering system was used to monitor start and end time instants for both the experimental paradigm macro-sections and the individual stimuli within each block. This system was wirelessly connected to the biomedical instrumentation to allow for the synchronisation of the neurophysiological signals.
After devices calibration, participants performed a climate change-centred Single-Category Implicit Association Test, which divided in four phases as described in Section 4.4.3.
To assess climate change explicit attitudes, the BAGCC questionnaire was administered.
Figure 4.1 summarises this experimental pipeline.

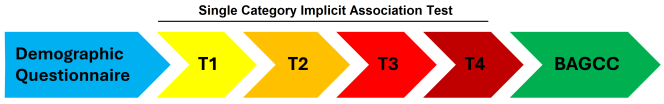


Figure 4.1: Experimental pipeline.

4.4.2 Subjective Measures

The BAGCC is composed of 23 items that must be answered according to a 5-point Likert scale. It is divided in 5 subscales: occurrence (OCC), causes

(CAU), consequences (CONS), self-efficacy (SELF), intentions (INT) and knowledge (KNOW). Table 4.1 displays its questions.

Table 4.1: BAGCC Items.

Beliefs about Global Climate Change	Subscale
1. How likely do you think it is that global warming is occurring now?	OCC
2. I have already noticed some signs of global warming.	OCC
3. It seems to me that temperature is warmer now than in years before.	OCC
4. It seems to me that weather patterns have changed compared to when I was a child.	OCC
5. I am quite sure that global warming is occurring now.	OCC
6. Article reading[1,2]	OCC
7. Global warming is mainly due to natural causes, not human activity[1].	CAU
8. The main causes of global warming are human activities.	CAU
9. Global warming is merely a natural fluctuation, not caused by human activity[1].	CAU
10. I am quite sure that human activities are to be blamed for global warming.	CAU
11. Unlike what most scientists say, there will be some[1]. positive consequences of global warming for the environment.	CONS
12. The consequences of global warming will be harmful. for the environment.	CONS
13. Global warming will bring about some serious negative consequences.	CONS
14. The consequences of global warming will be more positive than negative overall[1].	CONS
15. There are simple things that I can do that will have a meaningful effect to alleviate the negative effects of global warming.	SELF
16. I believe that little things I can do will make a difference to alleviate the negative effects of global warming.	SELF
17. Even if I try to do something about global warming, I doubt if it will make any difference[1].	SELF
18. There is very little I can do to mitigate the negative effect of global warming[1].	SELF
19. I plan to take some actions to stop global warming.	INT
20. I personally do not intend to do much to stop global warming[1].	INT
21. I will make some efforts to mitigate the negative effects of global warming.	INT
22. I intend to take concrete steps to do something to mitigate the negative effects of global warming.	INT
23. I would say my technical knowledge about global climate change is...[3]	KNOW

[1] stands for reverse coded items.
[2] The article reads: North Pole free of ice for first time in 50 million years. The North Pole is melting. The thick ice that has for ages covered the Arctic Ocean at the pole has turned to water, recent visitors there reported this week-end. At least for the time being, an ice-free patch of ocean about $1\frac{1}{2}$kilometers wide has opened at the very top of the world, something that has never before been seen by human beings and is more evidence that global warming may be real and already affecting climate.
[3] The item must be completed with a ruler between 1 = minimal to 5 = professional.

4.4.3 The Single-Category Implicit Association Test

The Single Category Implicit Association Test (SC-IAT) was designed after the same architecture in [154]: it comprises two stages to measure evaluative associations with a single category or attitude object, in this case, climate change. In the first stage, target words (i.e., related to climate change) are coupled with credible (positive valence) words in one single response key, whereas dubious (negative valence) words are categorised on a different key. In the second stage, the paradigm is switched. Table 4.2 lists used terms.

Table 4.2: SC-IAT stimuli categorisation

CATEGORY	WORDS
Target	*Global warming, Greenhouse effect, Ozone depletion, CO2 emissions, Sea level rise, Climate change, Ice melting, Air pollution*
Credible	*True, Verified, Certain, Real, Proved, Valid, Evident, Tested*
Dubious	*Wrong, False, Erroneous, Incorrect, Unfounded, Discredited, Unlikely, Irrational*

Specifically, when participants are in the first stage and see on the screen words like "Global warming" or "Real" they need to press the M key and when they see words like "Incorrect" they need to press the Z key. Vice versa for the second stage, where participants are required to press the M key when the screen presents words like "Incorrect" or "Ice melting". Each stage is divided into a practice part of 24 trials (words) and a test part of 72 trials. T1 and T2 refer to the practice and testing sections of the first stage, respectively, whereas T3 and T4 refer to the training and testing sections of the second stage. A custom web-based application, introduced in [122], presents words in randomised order and in a balanced way. After each decision, the interface displays the following word after 1500ms. Figure 4.2 schematically depicts this experimental framework.

The D-score was computed following the pseudocode explained in Algorithm 1.

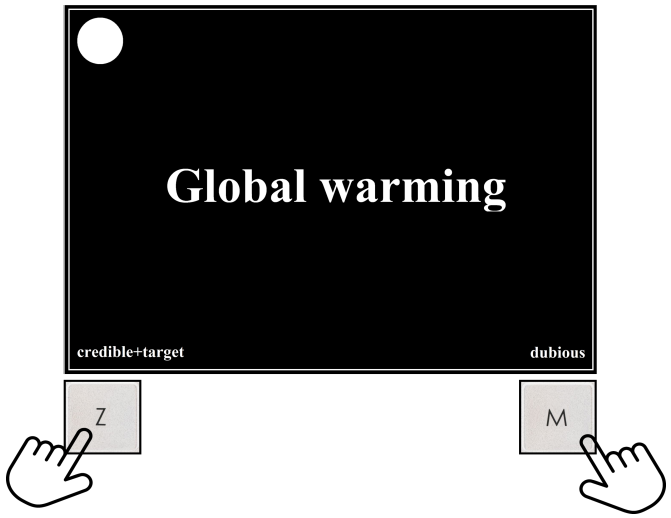


Figure 4.2: Schematic representation of the custom web-based application deploying the SC-IAT.

Algorithm 1 Compute D-score for Single Category IAT

procedure ComputeDScore(RT_Block2, RT_Block4)
    // **Step 1: Clean response times**
    Remove all response times < 350 ms from RT_Block2 and RT_Block4
    Remove all nonresponses from RT_Block2 and RT_Block4
    Let $ErrorPenalty \leftarrow 400$
    for all error responses in RT_Block2 **do**
        Replace with (Block2 Mean of correct responses)$+ErrorPenalty$
    end for
    for all error responses in RT_Block4 **do**
        Replace with (Block4 Mean of correct responses)$+ErrorPenalty$
    end for
    // **Step 2: Compute average response times**
    $\mu_2 \leftarrow$ Mean of RT_Block2
    $\mu_4 \leftarrow$ Mean of RT_Block4
    // **Step 3: Compute standard deviation**
    $SD_{pooled} \leftarrow$ Standard deviation of all correct responses in RT_Block2 and RT_Block4
    // **Step 4: Compute D-score**
    $D \leftarrow \dfrac{\mu_4 - \mu_2}{SD_{pooled}}$
    return D
end procedure

4.4.4 ETT preprocessing and feature extraction

Gaze and pupil data were preprocessed following the pipeline proposed by [155]. Specifically, it was implemented separately for the x- and y-gaze coordinates, as well as the left and right pupil, to remove artefacts (e.g., head movements, calibration loss due to sudden head movements) and blinks:

- Thresholding: after extracting the median and the median absolute deviation (MAD) from the signal, data that were found below and above the median and five times the MAD were removed and their locations stored [155].
- The missing points were filled with a shape-preserving piecewise cubic interpolation.
- Finally, the reconstructed signal was smoothed with moving average filtering, removing possible outliers.

Figure 4.3 shows an example of gaze and right pupil signals cleaning: the left panels display the raw data, the middle panels illustrate the three preprocessing steps, and the right panels show the filtered signals.

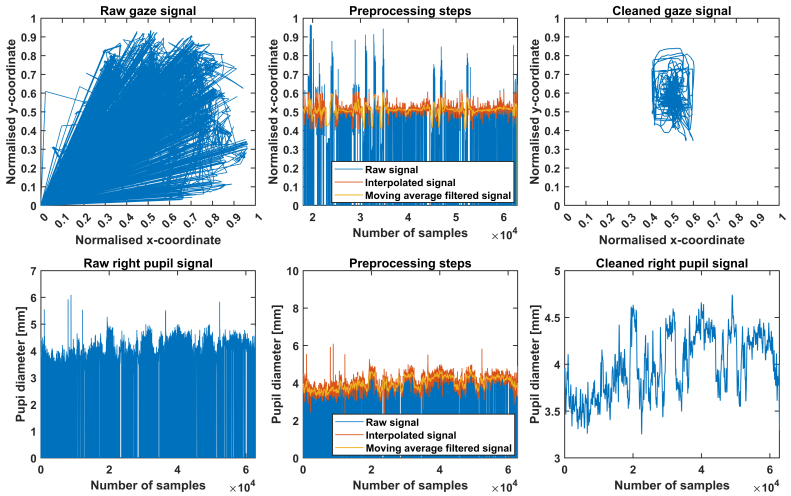


Figure 4.3: Illustrative example of the preprocessing pipelines of gaze and pupil signals. Data from the whole experiment (T1, T2, T3 and T4, rest and calibration preliminary phases) are shown.

Concerning eye movements, the 2D gaze time series was converted into a monodimensional vector by calculating the Point-to-Point Instantaneous Gaze Direction (PPIGD) to facilitate the description of its nonlinear dynamics [156]. It is computed considering the arctangent of the spacial distance for each pair of successive gaze points, whose coordinates are $g_i = (g_{x,i}, g_{y,i})$. Specifically, considering two consecutive gaze points referred to as the horizontal ($\Delta g_{x,i}$) and vertical ($\Delta g_{y,i}$) decomposition, the PPIGD is expressed in Equation 4.1:

$$\theta_i(t) = \arctan \frac{\Delta g_{x,i}(t)}{\Delta g_{y,i}(t)} \tag{4.1}$$

On the other hand, since it has been demonstrated the high correlation between left and right pupil data [157], the two signals were merged in a single vector as the mean between the two.
Finally, an additional method was designed to process the raw pupil data to obtain a suitable signal to analyse the blinking behaviour. Firstly, blinks were considered as sample points where left and right pupil dilation was below 2.5mm [158] or equal to 0. Then, a temporal threshold was applied to consider effective blinks as a continuous interval between 50ms and 1500ms [159, 160].

The triggering signal was used to automatically segment gaze and pupil signals in windows containing the physiological response for the individual stimulus. On the other hand, since blinking, in physiological conditions, rarely goes beyond 22 blinks/min [161], data were windowed considering the beginning and end of each SC-IAT section.

Gaze nonlinear features

Nonlinear analysis of the PPIGD signal was performed to highlight the complex dynamics that underlies the activation of brain areas related to visual scanning. This approach first requires the reconstruction of the dynamical system state space from which it is possible to extract features characterising its nonlinear behaviour. Since it is usually unfeasible, especially for physiological signals, to know and measure every variable that generates and controls a biological process, Taken's theorem was exploited. It allows the reconstruction of the phase space with equivalent dynamics related to a specific signal by using delayed copies of the available time series [156]. In the univariate case, this means that the embedding vector will be in the

form expressed in Equation 4.2:

$$x_n = (x_n, x_{n-\tau}, ..., x_{n-(m-1)-\tau}) \tag{4.2}$$

where $\{x_n\}_{n=1}^{N}$, n=1, ..., N represents the measured time series, m is the embedding dimension and τ is the time delay.
Several methods exist to compute m and τ. In this study, a custom MATLAB code was designed to iteratively calculate these metrics for each window as the first minimum of the false nearest neighbour function and the first minimum of the mutual information profile, respectively [121, 156].
After reconstruction, the properties of the state vectors were evaluated using the Recurrence Quantification Analysis (RQA). This method, proposed by Zbilut and Webber [162], allows for revealing the time points where the phase space trajectory visits roughly the same area in the phase space to build the so-called Recurrence Plot (RP), which is generated following Equation 4.3:

$$RP_{i,j} = \Theta(\epsilon - ||x_i - x_j||), i, j = 1, ..., N \tag{4.3}$$

where N is the number of points, ϵ is a threshold distance, $\| * \|$ is a norm (here, the Euclidean norm), and $\Theta(x)$ is the Heaviside function. Specifically, when the distance between x_i and x_j is lower than ϵ then $RP_{i,j} = 1$, otherwise is 0.
Similarly to m and τ, there is still no accordance in the optimal method to calculate the threshold ϵ. In the current study, it was calculated as the 10% of the phase space size [163].

Once the RP was obtained, the Cross Recurrence Plot methodology developed by Marwan *et al.* [163] was implemented to quantify the matrix structures and extract RQA features for each signal window:

- Recurrence Rate (RR): which calculates the density of recurrence points of the whole RP. It is expressed in Equation 4.4:

$$RR = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} RP_{p_i, p_j} \tag{4.4}$$

  where N is the number of measured points.

- Determinism (DET): which computes the percentage of recurrence points which form diagonal lines. It is expressed in Equation 4.5:

$$DET = \frac{\sum_{l=l_{min}^{D}}^{N} lP_D(l)}{\sum_{l=1}^{N} lP_D(l)} \tag{4.5}$$

where $P_D(l)$ represents the histogram of diagonal lines of length l. A length threshold l^D_{min} is set to exclude the diagonal lines related to tangential motion in the phase space.

- Average Diagonal Line (ADL)
- Length of the Longest Diagonal Line (LLDL)
- Entropy of Diagonal Length (EDL). It is expressed in Equation 4.6;

$$EDL = -\sum_{l=l_{min}}^{N} p(l)\ln(p(l)) \tag{4.6}$$

- Laminarity (LAM): that accounts for the percentage of recurrence points forming vertical lines. It is expressed in Equation 4.7

$$LAM = \frac{\sum_{v=v_{min}}^{N} vP_V(v)}{\sum_{v=1}^{N} vP_V(v)} \tag{4.7}$$

where $P_V(v)$ is the total number of vertical lines of the length v. A length threshold v_{min} is set to exclude vertical lines related to tangential motion in the phase space.

- Trapping Time (TT): the analogous of the ADL for vertical lines.
- Length of the Longest Vertical Line (LLVL).
- Recurrence Time of the First Type (RT1T): which computed the average distance between each point p of the RP and its neighbours. It is expressed in Equation 4.8:

$$RT1T = |p_i - p_j| : r_{p_i}, r_{p_j} \in R_{p_i} \tag{4.8}$$

where R_{p_i} is a set of recurrence points belonging to state r_{p_i}.

- Recurrence Time of the Second Type (RT2T): a correction to RT1T in order not to consider sequential states. It is expressed in Equation 4.9:

$$RT2T = |p_i - p_j| : r_{p_i}, r_{p_j} \in R_{p_i}; r_{p_{j-1}} \notin R_{p_i} \tag{4.9}$$

- Recurrence Period Density Entropy (RPDE): which computes the normalised entropy of recurrence times distribution. It is expressed in Equation 4.10:

$$RPDE = -(\ln T_{\max})^{-1} \sum_{t=1}^{T_{\max}} p(t) \ln p(t) \tag{4.10}$$

  where T_{max} is the maximum recurrence time and p(t) is the probability density function of the recursion times.

- Clustering Coefficient (CC): that represents the tendency to create from a cluster node. It is expressed in Equation 4.11

$$CC = \frac{2L_i}{k_i(k_i - 1)} \tag{4.11}$$

  where k_i and L_i are the numbers of linked nodes and edges corresponding to the i-th nodes, respectively.

- Transitivity (TRAN): that defines the probability that an edge connects two nodes. It is expressed in Equation 4.12:

$$TRAN = \frac{\sum_i \sum_j \sum_k A_i A_j A_k}{\sum_j \sum_k A_{j,k} A_{k,i}} \tag{4.12}$$

  where A represents the adjacency matrix computed as $A_{ij} = RP_{ij} - \delta_{ij}$ and δ_{ij} is the Kronecker delta.

Pupillometric features

A set of 10 linear and nonlinear parameters were extracted for each word. The diameter's mean (D_{mean}), standard deviation (D_{std}), minimum (D_{min}) and maximum (D_{max}) were computed. Then, to account for single stimulus processing times, the latency of D_{min} ($Lat_{D_{min}}$) and D_{max} ($Lat_{D_{max}}$) were also included [164]. Moreover, pupil dilation velocity (V_{dil} and $V_{dil_{minmax}}$) was calculated to obtain a sensitive metric to the rate of how mental resources are quickly allocated to complete a task [165].
To overcome the issues related to baseline corrected measures and to account for possible nonlinear behaviour of the pupil dynamics, three other parameters were considered: the Index of Pupillary Activity (IPA), Sample Entropy (SampEn) and the Largest Lyapunov Exponent (LLE).

The IPA, proposed by Duchowski *et al.* [166], consists of a wavelet decomposition of the pupillometric signal that, through a hard thresholding procedure, reflects the number of pupil diameter fluctuations related to the so-called pupillary hippus. This physiological phenomenon is known to be related to cognitive efforts [61, 167]. Specifically, a MATLAB code was developed to replicate the authors' description of the procedure:

- Obtain 2-level Discrete Wavelet Transform coefficient with sym-4 waveform.
- Detect local maxima in coefficient vector.
- Hard thresholding local maxima with a universal threshold ($\lambda_{univ} = \sigma * \sqrt{2 \log n}$, where σ represents local maxima standard deviation and n the number of local maxima).
- Counting remaining peaks NP.
- IPA computation as $IPA = \frac{NP}{TT}$, where TT is the length of the pupillometric signal.

Similarly, the SampEn represents a common measure to describe the degree of fluctuations in a time series. The embedding dimension was chosen iteratively for each stimulus analogously to the procedure illustrated in subsection 4.4.4, whereas the threshold r was always set to 0.2, aligning with [168].
The computation of the LLE followed the same procedure in section 4.4.4 to reconstruct the signal trajectory in the phase space. Depending on its characteristics, more attractors can be present in the state space. To quantify the sensitivity of this physiological system to initial conditions, the Lyapunov theory was included to assess its chaoticity. Specifically, Lyapunov exponents describe the evolution of trajectories in each dimension of the attractor, and its largest value (i.e., the LLE) represents a simple measure of how two initial nearby trajectories rapidly diverge or converge in the phase space. A $LLE > 0$ represents divergence, and vice versa.

Blinking features

Blinking metrics were computed section-wise (e.g., for the whole T2 stage). This study considered the number of blinks (N_{blinks}), mean, STD, minimum

and maximum of the blink durations (B_{mean}, B_{std}, B_{min}, B_{max}, respectively), the total time the eyes were closed (TTEC, by summing the duration of each blink event [169]), blinking rate ($r_{blinking}$), and the latency of the first blink (Lat_{1blink}).

4.4.5 Statistical Analysis

A correlation analysis was performed between BAGCC subscales scores and the SC-IAT D-score to identify a potential relationship between explicit and implicit association. Since BAGCC data are ordinal, the Kendall τ method was implemented to compute the correlation coefficient.
A RANOVA was conducted to evaluate the effects of Block Type (congruent vs. incongruent), Word Category (credible, dubious, and target stimuli), and their interaction with oculometric features, adapting the analysis framework introduced in 3.2.8 to control for data sphericity and p-value adjustments. Post-hoc tests were planned to follow up on any significant interaction effects only between Block Type and Word Category, applying the Tukey-Kramer correction. Multiple comparisons for the Block Type and Word Category conditions separately were not considered because it was out of scope to understand how participants' gaze varies while looking at words belonging to different groups without knowing its location within the experimental paradigm, and vice versa. Moreover, since the information related to implicit attitudes of climate change is inherent solely to target stimuli, the post-hoc analysis will report the comparison of target words between the congruent and incongruent blocks.
For blinking, since features were extracted section-wise, the Condition term dropped and depending on the outcome of Shapiro-Wilk tests for data normality (with α-level of significance set at 0.05), either a paired sample t-test or Wilcoxon sign test ($\alpha = 0.05$) was performed to detect significant differences between the congruent and incongruent associations.
Moreover, two further correlation analyses were performed: the first one aimed at investigating possible modulation phenomena of physiological behaviour to the D-score (thus, reaction times and decision making processes), the latter aimed at finding a prospective, physiology-based metric that could be used as a robust, not reaction time-based support in psychological research to better evaluate and highlight implicit attitudes. Coefficients were calculated with Pearson technique.

4.5 Results

The explicit attitudes toward climate change were assessed through the BAGCC scale, which median values and the related IQR are reported in Table 4.3.

Table 4.3: BAGCC results. Median values and IQR (in parentheses).

OCC	**CAU**	**CONS**	**SELF**	**INT**	**KNOW**
5 (0.5)	4 (1)	5 (1)	4 (0.5)	4 (0.5)	3 (1)

Implicit attitudes expressed by the SC-IAT modified D-score highlighted a wide distribution characterised by a negative median value ($M_{D-score} = -0.025$, IQR = 0.53, min-max range: [-0.66 1.1]).

For the first hypothesis, explicit and implicit attitudes presented non-significant correlation ($\rho = 0.10$, p = 0.52).

The second hypothesis highlighted that all RQA-related eye tracking and pupillometric measures presented significant differences for the Block Type x Word Category interaction, suggesting that the effect of block congruency on recurrence patterns was contingent upon the specific category of which the target stimuli were embedded. Appendix C reports the tables containing the RANOVA tests results. However, no significant alteration was found when investigating the comparison between congruent and incogruent associations of target words only, with the exception of pupil size sample entropy. Figure 4.4 shows that such a parameter was higher in the incongruent condition than the congruent one.

Finally, no blinking-related feature was significantly different between congruent and incongruent association blocks.

As for the third hypothesis, only the parameters presenting a significant correlation in compatible and incompatible blocks are reported to focus on possible physiological modulation processes that may relate to the slower motor response selection. For gaze, DET, LLDL, LAM, TT, LLVL presented significant relationships in both sections. Specifically, TT and LLVL are characterised by a stronger link in the case of congruent association, vice versa for DET, LLDL and LAM. Figure 4.5 displays the scatterplots for

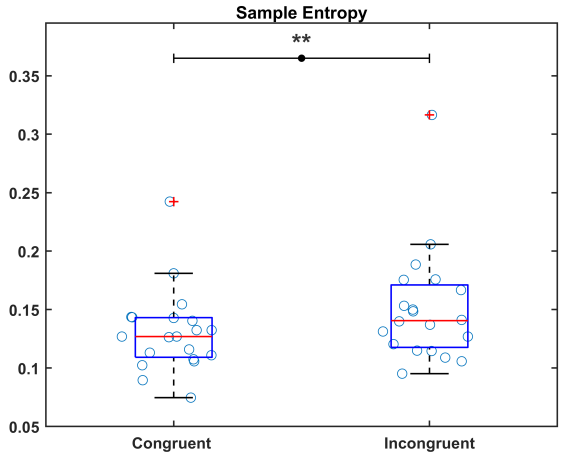


Figure 4.4: Data distribution of pupil size sample entropy for the congruent and incongruent associations, considering target words only. A (•) represents a statistical difference $p < 0.05$, whereas (**) indicate a medium Cohen's d effect size.

these parameters.

In pupillometry, no significant correlation was found. On the other hand, for blinking behaviour, significant relationships were found between D-score and the number of blinks, the TTEC, and the blinking rate (Figure 4.6).

A correlation analysis was also performed to evaluate possible correlations between the D-score and the differences in oculometric features from T4 and T2 for the fourth hypothesis. However, no significant relationship was found in this analysis.

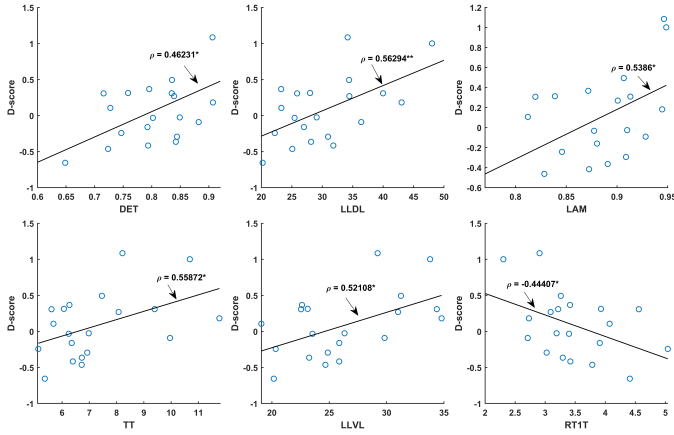


Figure 4.5: Significant correlations between the D-score and RQA-derived eye tracking metrics. A (•) refers to a significant relationship with $p < 0.05$.

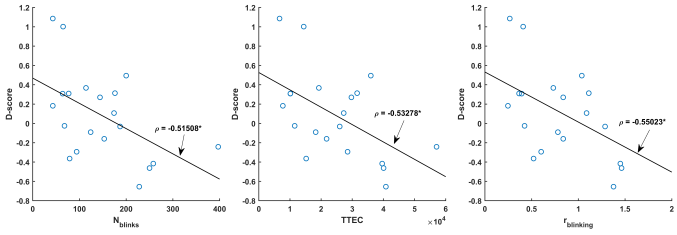


Figure 4.6: Correlations between D-score and blinking behaviour metrics. A (•) refers to a significant relationship with $p < 0.05$.

4.6 Discussion

This study investigated whether and how oculometric features are sensitive to making associations that go against an individual's internalised opinions toward climate change. Moreover, this study aimed to discover the physiological processes that provoke the delayed motor response of incongruent associations. It also proposed physiological variables to be used concurrently with psychological scores in order to provide alternative, objective measures to assess the IAT effect.

Based on the median BAGCC subscales scores, the recruited population was considerably aware about the ongoing climatic crisis, also showing a good knowledge about its causes and consequences. However, the subscales concerning *SELF* and *INT* highlight that they seemed not to be confident in its ability to act meaningfully on climate issues and, thus, had also weak behavioural intentions to engage in pro-environmental decisions. These results somewhat anticipated the low implicit attitudes measured by the D-score. However, this outcome did not mean that the current sample was sceptical or hostile to the topic (on the contrary, as Table 4.3 shows), but that it might feel unable or ineffective, causing their implicit attitudes to become more negative. This could be driven by frustration, helplessness, or defensive detachment. As a consequence, subjects may downregulate emotional engagement and avoid thinking about the problem. Moreover, as other studies obtained [170], no significant correlation was found between explicit and implicit attitudes, hence suggesting that a multivariate processing approach is needed.

The integration of ETT while performing a SC-IAT allowed to get a more complete perspective on the sensitivity of oculometric features to the making of incongruent associations. Indeed, it was hypothesised that such activity would increase participants' cognitive load due to an internal conflict caused by the previous knowledge of climate change issues. Interestingly, the repeated-measures ANOVA revealed robust and statistically significant Block Type x Word Category interaction across all RQA parameters (all $p < .001$). On the other hand, post-hoc comparisons between T2 and T4 target words did not yield any significant differences (Table C.1). This apparent discrepancy can be meaningfully interpreted in light of the effect sizes, which highlight subtle variations in attentional dynamics that did not reach statistical significance but could suggest potentially meaningful cognitive-emotional trends. Among the metrics, Recurrence Rate (RR) showed a small effect sizes ($dz = -0.23$), suggesting a tendency toward a less temporally structured and a more variable gaze behaviour in response to T4 targets compared to T2. This finding could be caused by an increased saccadic activity due to a perceivably higher visual complexity in the incongruent section [169]. Thus, it could underline more frequent eye switches between the locations of the words and keyboard commands. Indeed, the latter were always displayed on the screen as a memo for participants. Although these findings warranted

caution due to their lack of statistical significance, the consistent pattern of small effect sizes across several RQA measures suggested that the temporal and structural properties of visual attention were modulated by target word semantics, which may interact with participants' underlying attitudes toward climate change. These micro-dynamics underscored the value of eye movement analysis as a sensitive index of implicit processing in socially charged contexts.
A similar consideration can be done for pupillometric features (Table C.2). The Block Type x Word Category interaction was significant for each parameter ($p < .001$), but the multiple post-hoc tests found a significant difference for sample entropy. As Figure 4.4 displays, the lower regularity and increased complexity of the pupil signal under the T4 section may index elevated autonomic variability or internal state disruption, often associated with ambiguity or conflict in affective decision-making contexts [169, 171]. Taken together, these findings may indicate that oculometric features are not adequate to study implicit attitudes. However, the lack of significant differences between target T2 and target T4 could have been biased by the small sample size. Nonetheless, Tables C.1 and C.2 suggest that evaluative incongruency still elicits subtle but consistent changes in both visual attention and autonomic arousal, particularly in response to climate-related targets that clash with implicit structures. These dynamics, as captured through both linear and nonlinear features, underscore the affective and cognitive sensitivity of ETT in uncovering implicit attitudinal conflict, even in the absence of overt behavioural responses.

As far as potential modulating physiological processes are concerned, correlation analyses highlighted a promising outcome. The DET and LLDL metrics, both underlying sequential gaze structure, indicated that more patterned and predictable visual engagement aligned with stronger pro-climate attitudes. Indeed, DET quantifies the predictability in a participant's gaze behaviour. A higher DET value is typically linked to longer sequences of repetitive or stable visual patterns, as well as a suppression of chaotic or random saccades. A higher DET associated with larger D-score may reflect a possible stronger cognitive effort and focused attention while making associations that goes against an individual's internalised opinions. Therefore, a more structured scanning path could represent a compensatory mechanism to resolve the automatic conflict that might affect negatively the reaction

time, thus, the decision making process. Similarly, the LLDL is connected to the similarity degree of the longest scanning paths. Hence, the steeper regression line in the incongruent section seemed to suggest more coherent gaze trajectories that can again reflect conflict monitoring to maintain high task performance. On the other hand, T2 was less cognitively demanding, so gaze stability may naturally follow from alignment between attitude and task demands. This outcome was also supported by RQA parameters that are related to vertical structures, i.e., LAM, TT and LLVL, which underline how long the eyes are locked in a specific area of interest. Therefore, higher values highlighted longer fixations and, consequently, attentional freezing or focused dwelling. The stronger gaze stasis that occurred in the incongruent block seemed to emphasise a longer time to resolve the internal conflict that may physiologically translate into longer fixation, possibly on the category labels that were always shown in the screen as reminders for participants, slowing down reaction times. Finally, the negative correlation between D-score and RT1T also suggested that participants with better implicit attitudes not only retraced the same scanning path as DET and LLDL indicated, but they did it at a progressively slower rate. This could be a direct explanation of the delayed motor response that participants experienced in the T4. Moreover, a shift in attention that recalled more cognitive resources was also supported by the blinking activity. Figure 4.6 shows that blinks were relevantly reduced for participants with better implicit attitudes. This pattern could represent a visually focused attentional state, with a more stable oculomotor control and a suppression of momentary disruptions in vision. Indeed, blinking acts as a natural interrupter of visual input, and its inhibition has long been associated with increased concentration, cognitive load, and heightened sensory vigilance [172, 173]. In this context, less blinks may reflect a strategic effort to avoid fragmenting critical input during rapid categorisation of emotionally or semantically relevant information, particularly when target stimuli carry moral salience. Thus, participants with stronger pro-environmental attitudes may experience the evaluative task as more difficult, engaging or personally meaningful, leading to reduced blink rates as a form of motivated attention.

Finally, the correlation analysis between the D-score and the difference in ETT metrics between incongruent and congruent associations did not identify any significant relationships. This finding seemed to advise for a more

robust computation of a "physiological" D-score, potentially replicating its original algorithm (1), in order to propose in psychological research an alternative metric to the traditional reaction time-based ones.

4.7 Conclusion

This study examined the capacity of oculometric parameters to investigate implicit attitudes toward climate change during a SC-IAT. Despite participants demonstrating high awareness and knowledge about climate change, many reported low confidence in their individual efficacy and weak behavioural intentions-factors, that likely influenced the low implicit pro-environmental attitudes. The significant alteration in pupil size sample entropy suggested a shift in cognitive engagement during incongruent trials due to an increased conflict. Correlation analyses further highlighted how more structured and predictable gaze behaviours, reduced blink rates, and slower re-scanning trajectories were associated with stronger implicit pro-environmental attitudes. These patterns may reflect heightened attentional control and mental effort to resolve internal conflicts during implicit evaluations. These findings may indicate the use of oculometric indicators as valuable, objective complements to traditional reaction time-based measures. Notably, the study did not find direct correlations between physiological shifts and D-score. The small sample size requires caution when interpreting such findings. However, they could lay the groundwork for developing a refined, physiologically derived index of implicit attitudes, offering new possibilities for psychological research in climate-related decision-making and in assessing behavioural interventions.

Chapter 5

Climate Change and Health: Voice Implications and Novel Diagnostic Techniques

The voice production system is sensitive to air pollution, one of the major causes of climate change. This may determine relevant alterations in pitch, loudness and rhythm collectively known as dysphonia. This chapter proposes a feasibility study of an AI-based tool to support the differential diagnosis of two vocal disorders: benign lesions of the vocal folds (BLVF) and unilateral vocal fold paralysis. A set of 287 patients was required to utter a sustained vowel and an Italian word from which acoustic features were extracted to train unsupervised and supervised explainable models. The first confirmed the existence of unique vocal phenotypes of dysphonia. The latter refined this outcome, pointing out that the combination of vocal tasks improved the diagnostic accuracy, which was also not dependent on age. Relevant parameters for the distinction of vocal pathologies were different across genders. The sub-distinction in the BLVF patients (i.e., nodules, polyps and cysts) will require a larger sample size to obtain better performance. [1]

[1]The part concerning unsupervised learning has been published as "Unveiling Vocal Phenotypes of Dysphonia with Unsupervised Learning" in *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOSIG-*

5.1 Introduction

Environmental pollution refers to the introduction of harmful substances into the environment. Pollutants (e.g., heavy metals, particulate matter, microplastics) in air, soil and water are a crucial issue that has drawn great public concern and scientific attention [176]. Specifically, air pollution can be defined as the presence of any liquid, solid, or gas compounds in the atmosphere at such concentration values that can directly or indirectly affect humans, animals, and/or plants health. It originates from certain domestic and industrial activities. When inhaled, these hazardous substances can provoke systemic responses of the human body. Recently, research highlighted how air pollutants are toxic also to the voice production organs, mainly the lungs and the larynx [177,178], determining highly variable voice alterations that are collectively known with the term dysphonia. This is a medical condition characterised by irregular and usually acute pitch, lower vocal quality and loudness with respect to gender- and age-matched normophonic subjects [179]. It represents one of the most evident markers of voice disorders, with a prevalence in the adult population estimated between 3% and 10% [180–182]. The etiopathogenesis can be strongly associated with climate change triggers. The absorption of air pollutants by the vocal folds' epithelium superficial layer induces changes in ions transport system and in the synthesis of specific proteins that are essential to the mucosal wave [183] and to keep this tissue hydrated [184]. It can potentially lead to inflammation (e.g., greater prevalence of chronic laryngitis were found in deeply industrialised areas [185]) and oedemas, as well as pulmonary excessive efforts to put the vocal folds in motion due to altered mechanical properties. Moreover, a compromised lung functionality may determine compensatory adjustments resulting in voice fatigue, nodule and polyps formation and even neuromuscular problems (e.g., muscle tension dysphonia) [177].

The gold standard method for the diagnosis of dysphonia is based on the visual inspection of the vocal folds by laryngoscopy. Throughout the decades,

NALS, 2025 [174]. Part of the study regarding explainable AI has been published in "Towards an explainable Artificial intelligence system for voice pathology identification and post-treatment characterisation" in *Biomedical Signal Processing and Control, v. 104, p. 107530, 2025* [175]. The supervised artificial intelligence experiments were developed in collaboration with Prof. Eng. Lorenzo Frassineti and the unsupervised experiments with Dr. Francesco Correnti.

complementary assessments were also proposed to overcome some relevant limitations, such as the lack of high-resolution endoscopy in decentralised ambulatories or primary care units, invasiveness and low tolerability [186, 187]. Moreover, laryngoscopy requires being physically present in hospitals, which can be demanding for fragile subjects, and in case of worldwide health emergencies or natural disasters [188, 189].
Perceptual voice evaluation is an indirect, well-established method typically performed along with laryngoscopy to assess voice disorders. Despite its advantages (e.g., cost-effectiveness, time efficiency), it suffers from inter-rater variability and physicians' experience [190]. Another objective practice combines the acoustic analysis, i.e., the parametrisation of voice and speech production through linear and nonlinear signal processing pipelines, and AI techniques. This approach could allow the development of reliable automatic tools to support screening, diagnosis, and monitoring of vocal properties after surgical or logopaedic treatments. Additionally, it is compatible with current healthcare facilities objectives in reducing their high carbon footprint. Indeed, this sector contributes up to 10% of the overall greenhouse gases emission, highlighting the need to design novel strategies for improving rooms thermal insulation, limiting waste and physical admissions for minor medical issues by means of telehealth solutions [14, 15].
Automatic systems of voice analysis need careful planning, starting from the choice of the input utterance, which indeed plays an essential role in feature extraction, and it is strongly related to the type of pathologies of interest [179]. To fully grasp the multidimensionality and complexity of voice production, considering a sustained vowel only (typically the /a/) is not sufficient. Therefore, other tasks have been proposed, such as:

- Running speech. It contains important information about voice breaks, onset, and offset. It can be crucial for pathology detection and identification. Ali *et al.* . [188] extracted linear prediction coefficients from spontaneous speech and achieved high performances with a Gaussian Mixture Model in distinguishing healthy and disordered speech. Godino-Llorente *et al.* [191] have obtained similar results using features extracted from the standardised "Rainbow Passage" and implementing a multilayer perceptron.

- Number listing task for the evaluation of coarticulation. Frassineti *et al.* [192] demonstrated that acoustic parameters and entropy indexes can discriminate vocal phenotypes of patients diagnosed with genetic

syndromes. In contrast, Ali *et al.* [193] computed voice contour measures and successfully used them as features in a support vector machine (SVM) to separate pathological and healthy subjects.

From these, feature extraction is carried out. Conventional acoustic metrics (e.g., fundamental frequency, jitter, shimmer, harmonic-to-noise ratio), mel-frequency cepstrum coefficients (MFCCs) and other parameters (as cepstral peak prominence CPP, high-order statistics, nonlinear measures) are then typically used in diversified classification problems.

The latter can be performed with two main alternatives: unsupervised and supervised learning. The main difference relies on the type of data used. Supervised learning uses labelled data, which means that models, when trained, are provided with a baseline understanding of what the correct output should be. On the contrary, unsupervised learning or clustering works with unlabelled data. Therefore, it analyses the inherent structure of data without any specific instruction to identify hidden groups by considering proximity or dissimilarity rules [194,195]. In acoustic analysis, clustering has been used firstly as a feature weighting technique, i.e., to estimate and rank the relevance of extracted features to better highlight an underlying patterns and remove noise in data and improve supervised learning performance [196]. Moreover, clustering is particularly indicated in case of highly heterogeneous diseases that are characterised by complex and wide pathophysiology. Indeed, by addressing data quality, it can refine the phenotype of a given disease, potentially helping to plan better precision medicine procedures. For instance, the k-means algorithm was implemented to characterise the degree of speech impairments and find speech subtypes in Huntington [197] and Parkinson's disease [198], where no widely accepted guideline exist to define such subgroups.

On the other hand, supervised learning has been extensively used in the acoustic context to develop automatic tools that can distinguish vocal properties of healthy and pathological subjects [199–202], achieving high performance up to 95% accuracy [202]. Recently, research started to focus on a distinct problem, i.e., designing AI-based technologies to support the differential diagnosis of voice disorders. In this context, a SVM using MFCCs was able to discriminate nodules, polyps, and cysts (collectively referred to as a single class), unilateral vocal fold paralysis (UVFP), and glottis neoplasm on a small cohort of patients with a 97% sensitivity [203] and 64% accuracy [204]. These promising results led to more detailed research: Harar *et*

al. [205] extracted conventional acoustic features, spectrogram parameters, and MFCCs from the sustained /a/ of patients affected by several voice disorders contained in the Saarbrücken Voice Database [206]. They obtained an F1-score of 73% in separating them into different classes, with an XGBoost Tree. Moreover, Marchese *et al.* [207] performed an exploratory analysis by implementing an SVM classifier to separate nodules, polyps, cysts, and Reinke's oedema that obtained 55% accuracy for female patients.
Notably, these works mainly concerned methodological approaches. Indeed, they focused on feature extraction and selection methods and designing machine or deep learning frameworks to increase performance only. Hence, their results, while encouraging, have been achieved to the detriment of explainability, which has rarely been addressed. This represents a critical issue since a technique that is not directly interpretable and trustworthy can pose a major threat to the usability of AI-based methods [208]. As a matter of fact, it is now necessary to move toward making black-box models' output more understandable, since clinicians need more information to make a decision, and a simple binary prediction is insufficient to be accepted as clinically relevant. Therefore, to be considered practical tools and to avoid limiting their effectiveness, explainable AI (XAI) has been proposed to allow field experts to understand and manage AI model development and results. Two different types of XAI approaches can be defined [208, 209]:

- Post-hoc techniques that provide text, visual or local explanations to allow model interpretability. In voice analysis, permutation feature importance and Shapley values were used to understand which acoustic features were considered relevant by machine learning models to separate patients diagnosed with polyps and UVFP [210]. Shapley values were also computed to find acoustic parameters able to detect Parkinson's disease in patients' standardised speech [211]. Furthermore, the Local-Interpretable Model-agnostic Explanation method was implemented to evaluate the relationships between perceptual and objective assessment of spasmodic dysphonia [212].

- AI frameworks that are intrinsically interpretable. Classifiers such as K-Nearest Neighbors or Decision Trees are commonly used for voice pathology detection [179]. Still, their explainability is overshadowed by a usually lower classification performance with respect to more complex but less transparent models such as Support Vector Machine or Ensemble Trees classifiers [209].

Moreover, even when interpretability is addressed, explainability relies also on extracted features, which should be easily related to physiological processes [175]. Specifically, it would be important for otolaryngologists to understand which acoustic features were considered most relevant for a classifier for the distinction of two or more classes and that such metrics could have a reliable physiological meaning (e.g., the first formant is related to the degree of the pharynx constriction [213]). Such an approach would make it easier for healthcare professionals to plan intervention and rehabilitation programs.

This chapter presents two separate analyses on vocal properties that were used as features both in an unsupervised and a supervised learning framework. The aim is to study patients diagnosed with nodules, polyps, and cysts (considered separately and jointly as the class of benign lesions of the vocal folds, BLVF) and unilateral vocal fold paralysis (UVFP). Specifically, clustering is used as an exploratory technique to understand whether dysphonia presents unique vocal phenotypes and validate the feasibility of the supervised classification experiment.
The major contributions of this investigation are:

- To evaluate the implicit role of age in the voice identification problem. Indeed, it is well-known that ageing affects vocal quality. However, models performing this task have never addressed its potential bias.
- To assess the capabilities of AI-based solutions in the BLVF differential diagnosis task, which still constitutes a poorly studied research question.

5.2 Materials and Methods

5.2.1 Patients

This study recruited 287 patients (of which 183 were female F, mean age = 44.6±4.6 years, 104 were male M, mean age = 42.6±9.4 years) at the Ospedale Maggiore Policlinico Milano (Milan, Italy). All patients underwent both perceptual evaluations of their voice and video-laryngostroboscopic assessments. Patients were diagnosed with nodules (15 F, 2 M), polyps (56 F, 38 M), cysts (34 F, 6M), and unilateral vocal fold paralysis (78 F, 56 M).
Voice samples were recorded using a C1000S microphone (AKG, Vienna, Austria) with a sampling frequency of 44.1kHz and at a fixed distance of

5cm from the patient's mouth during the phonation, at comfortable pitch and loudness of the sustained cardinal vowel /a/ and the Italian word /aiuole/, which contains all the basic vowels of the Italian language..

Commonly, UVFP patients present a higher mean age than BLVF ones [214]. In this database, female patients diagnosed with UVFP had a mean age of 44 years (minimum-maximum range: 19-68), whereas the ones diagnosed with UVFP presented a mean age of 50 years (min-max range: 21-72). For male patients, the mean age for BLVF was 42 years (min-max range: 19-78), whereas the mean age of UVFP subjects was 51 years (min-max range: 26-80). After checking for normality of data distribution, a t-test was performed with a α-level of significance set at 0.05 to understand whether the age difference between BLVF and UVFP cohorts was statistically significant. Firstly, the test was carried out considering all available data. Then, in case of a significant difference, an age subrange between 18 and 60 years was selected to get a homogeneous group (according to [205]). The distribution of patients, for both age inhomogeneous and homogeneous groups, is displayed in Table 5.1.

Table 5.1: Patients distribution divided by age, gender and voice pathology. The acronym y.o. stands for years old.

	Female patients		**Male patients**	
Pathology	Whole database	18-60 y.o. age group	Whole database	18-60 y.o. age group
Nodules	15	7	2	0
Polyps	56	36	38	20
Cysts	34	18	6	4
UVFP	78	43	58	18

5.2.2 Feature Extraction

After manual segmentation, audio files were processed using the open-source BioVoice software [215]. This tool provides a user-friendly interface that allows selecting the proper frequency range for adults (also distinguishing between male and female), children, and newborns to perform acoustic analysis.

Specifically, it computes several parameters in both the time and frequency domain. Table 5.2 shows those used in this work and their description. Two features, namely T0 (F0 min) and T0 (F0 max), were further processed to make them independent from the total duration of the audio recording. Therefore, they were normalised with respect to the time duration of the recording. It allowed obtaining a reliable estimation of the time instant at which the minimum or maximum of the F0 occurs, i.e., at the beginning, in the middle, or at the end of each recording. These metrics were chosen due to their direct relationship with physiological processes.

Table 5.2: Acoustic parameters estimated by BioVoice.

Feature	**Description**
F0 mean [Hz]	Mean fundamental frequency
F0 median [Hz]	Median fundamental frequency
F0 std [Hz]	Standard deviation of fundamental frequency
F0 min [Hz]	Minimum value of the fundamental frequency
T0 (F0 min) [s]	Time instant at which the minimum of F0 occurs
F0 max [Hz]	Maximum value of the fundamental frequency
T0 (F0 max) [s]	Time instant at which the maximum of F0 occurs
Jitter [%]	Irregularity of F0 in the frequency domain
NNE [dB]	Normalized Noise Energy
F1 mean [Hz]	Mean value of the first formant
F1 median [Hz]	Median value of the first formant
F1 std [Hz]	Standard deviation of the first formant
F1 min [Hz]	Minimum value of the first formant
F1 max [Hz]	Maximum value of the first formant
F2 mean [Hz]	Mean value of the second formant
F2 median [Hz]	Median value of the second formant
F2 std [Hz]	Standard deviation of the second formant
F2 min [Hz]	Minimum value of the second formant
F2 max [Hz]	Maximum value of the second formant
F3 mean [Hz]	Mean value of the third formant
F3 median [Hz]	Median value of the third formant
F3 std [Hz]	Standard deviation of the third formant
F3 min [Hz]	Minimum value of the third formant
F3 max [Hz]	Maximum value of the third formant
Signal duration [s]	Time duration of the audio file
% voiced	Percentage of voiced parts inside the whole signal
Voiced duration [s]	Total duration of voiced parts

5.2.3 Unsupervised learning: Clustering

A clustering problem assumes that an event space E, described by an observation matrix $n \times p$ comes from k different underlying divisible clusters. Several methods have been proposed, with the k-means algorithm being one of the most commonly used.

This algorithm initialises by selecting k values from the event space as candidate cluster centres. It then iteratively follows two steps:

- Assign: Each point in the event space is assigned to the nearest candidate cluster centre, resulting in k different clusters.
- Update: Each cluster centre is recalculated as the mean of the coordinates of all points in the cluster, yielding centroids.

The algorithm repeats these steps iteratively until convergence, where the cluster centres no longer change. Clustering efficiency was evaluated by considering the silhouette score, a value ranging between -1 and 1 that compares inter-cluster distances with intra-cluster ones. Moreover, accuracy was taken into account after performing manual diagnosis assignment.

Transformation

Two feature transformation techniques were applied to enhance the clusterability of the data: logarithm and cube-root transform. They were both used to reduce the skewness of a distribution, with the cube-root transform being less effective but suitable for both positive and negative values than the logarithm one.

Feature Selection

To reduce the problem's dimensionality and extract a subset of acoustic features containing the most valuable information, Principal Component Analysis (PCA) and correlation analysis were applied (removing parameters with $\rho > 0.8$).

Unaware Analysis

A first clustering problem investigated whether the k-means algorithm could separate data into two groups, corresponding to the BLVF and UVFP classes, for each gender. A second clustering problem explored the possibility of separating data into four groups, i.e., the BLVF subclasses (nodules, polyps, cysts) and UVFP. This experiment was performed for the F dataset only due to the low nodules numerosity of the M dataset.
Three approaches were followed:

- Vanilla: No preprocessing was involved, serving as a reference point for comparison.
- PCA: Outlier removal based on the Interquartile Range (IQR) was performed. Depending on data normality, features were also scaled and centred accordingly. PCA was then applied: the number of principal components m was chosen according to the k-means clustering performance.
- Unskew + PCA: Before applying the PCA pipeline, skewed data were cube-root or log-transformed.

Aware Analysis

To improve the identification of clustered structures in the analysed groups, all available information were leveraged to filter out irrelevant features. Hence, to identify an optimal subset, statistical tests were employed to determine which parameters presented significant differences between groups. An analogous pipeline to the one presented in subsection 5.2.3 was implemented.

5.2.4 Supervised learning: Machine Learning

Machine learning (ML) algorithms were implemented to conduct three classification problems for female and male datasets. Two experiments were conducted to test whether age could affect the results. The first concerned a classifier trained with the whole dataset (E1), while the second considered a subset of subjects aged 18-60 years only (E2). The third experiment (E3) refined the previous ones, as its purpose was to further separate the BLVF class into three subclasses: nodules, polyps, and cysts, as in [207]. This experiment was not performed for the M group as the sample size of the nodule patients was too small (2 subjects only, as shown in Table 5.1).
The following supervised classifiers were considered and validated using a k-fold cross-validation framework (k=10) [216]: SVM (with both linear and Gaussian kernel), Ensemble models (i.e., AdaBoost and RobustBoost using decision trees as weak learners), and k- Nearest Neighbors algorithm (k-NN). Depending on the type of experiment and the numerousness of each voice pathology condition across genders, Bayesian hyperparameters optimisation was used to maximise either the accuracy (ACC) or, in case of an imbalanced number of samples between groups, the Matthews Correlation

Coefficient (MCC). Indeed, the literature demonstrated that the MCC represents a more reliable metric for micro-averaging correction in multi-class frameworks. [217]. Specifically, for E1 and E2, the ACC was used for hyperparameter tuning, whereas for E3, the MCC was implemented. The number of iterations for Bayesian optimisation was set to 200.
Models were trained using three different sets of features:

1. Features extracted from the sustained vowel /a/.
2. Features extracted from the Italian word /aiuole/
3. Merge of the two sets of features 1) and 2). In this case, subjects without one of the vocal tasks were removed.

Feature Selection

Each predictor was standardised with a z-score procedure during each cross-validation step for the training and validation sets, according to the current training statistics, avoiding possible data leakage.
Furthermore, to reduce the number of predictors, the following feature selection methods were applied: Pearson correlation, ReliefF [218], LASSO [219], and mrMR [220]. More precisely, before ReliefF, LASSO, and mrMR, the highly correlated predictors were removed from the training and validation set, retaining only those with an absolute Pearson correlation coefficient (PCC) < 0.8 in the training set [221]. When ranked by the considered feature selection methods, the number of features and the nearest neighbours for ReliefF were also used as parameters during the Bayesian optimisation. The True Positive Rate (TPR) for each class and the area under the ROC curve (AUC) were also considered for binary classifiers. In contrast, for three-class problems, only the ACC, MCC, and AUC for each class were evaluated.

Explainability

Explainability was addressed both by implementing transparent ML algorithms (k-NN) and predictor importance as a post-hoc technique for ensemble methods. The importance measure for each feature is computed with a two-step procedure [222]. First, changes in the node risk due to split are summed, where a change is defined as the difference between the risk for the parent node and the total risk for the two children. Then, the summation is

divided by the total number of branch nodes. For example, if a tree splits a parent node (node 1) into two child nodes (nodes 2 and 3, respectively), then the algorithm increases the importance of the split predictor according to Equation 5.1:

$$imp = \frac{(R1 - R2 - R3)}{N_{branches}} \tag{5.1}$$

where *imp* is the feature importance, R_i is the node risk of node i, and N_{branch} is the total number of branch nodes. A node risk represents a node error or node impurity weighted by the node probability. This is expressed by Equation 5.2

$$R_i = P_i \times E_i \tag{5.2}$$

where P_i is the node probability of node i, and E_i is either the node error (for a tree grown by minimising the twoing criterion) or node impurity (for a tree grown by minimising an impurity criterion, such as the Gini index or the deviance) of node i.

Finally, a statistical analysis was performed for the F and the M cohorts, considering the parameters the AI models retained after feature selection. A Mann-Whitney test was used, with a level of significance $\alpha = 0.05$. These tests were performed to evaluate if statistical differences exist between BLVF and UFVP.

5.3 Results

5.3.1 Clustering

For all experiments, the best results were obtained with one single principal component. Figure 5.1 displays how the observations were distributed along the principal component.

Figure 5.1: Graphical representation of the data distribution along the principal component axis in the original dataset (right) and the predicted clusters (left).

Unaware Analysis

Table 5.3 shows the results of the three pipelines for the 2-groups unaware cluster analysis, divided by gender. S_n is the silhouette score computed for each pathological group, where $n = 1$ refers to the BLVF class and $n = 2$ to the ULVF one; on the other hand, S corresponds to the average silhouette score. The abbreviation A refers to accuracy. Finally, σ shows the percentage of variance explained by the single PCA component.

Table 5.4 displays the results of the three pipelines for the 4-groups unaware cluster analysis. S_n is the silhouette score computed for each pathological group, where $n = 1$ refers to the nodules, $n = 2$ to polyps, $n = 3$ to cysts, whereas $n = 4$ to UVFP. This experiment was not performed for the male dataset due to nodules low numerosity.

Table 5.3: Results of k = 2 clusters analysis on the males and females datasets. S = silhouette score, A = accuracy, σ = retained variance percentage.

F - Unaware	$\mathbf{S}_1$	$\mathbf{S}_2$	**S**	**A**	σ
Vanilla	0.47	0.37	0.42	0.54	1.00
PCA	0.68	0.62	0.65	0.55	0.19
PCA + Unskew	0.70	0.59	0.64	0.616	0.21
M - Unaware					
Vanilla	0.61	0.51	0.56	0.52	1.00
PCA	0.79	0.62	0.70	0.68	0.26
PCA + Unskew	0.77	0.64	0.70	0.69	0.28

Table 5.4: Results of k = 4 clusters analysis on the females dataset. S = silhouette score, A = accuracy, σ = retained variance percentage.

F - Unaware	$\mathbf{S}_1$	$\mathbf{S}_2$	$\mathbf{S}_3$	$\mathbf{S}_4$	**S**	**A**	σ
Vanilla	0.21	0.28	0.18	0.17	0.21	0.34	1.00
PCA	0.63	0.59	0.61	0.55	0.60	0.30	0.19
PCA + Unskew	0.57	0.67	0.66	0.60	0.62	0.34	0.21

Aware Analysis

Table 5.5 shows the results of the three pipelines for the 2-groups aware cluster analysis, divided by gender.

Table 5.5: Results of k = 2 clusters analysis on the males and females datasets, after retaining statistically significant features only.

F - Aware	S_1	S_2	**S**	**A**	σ
Vanilla	0.47	0.37	0.42	0.54	1.00
PCA	0.71	0.65	0.68	0.69	0.34
PCA + Unskew	0.68	0.72	0.70	0.665	0.35
M - Aware					
Vanilla	0.61	0.51	0.56	0.52	1.00
PCA	0.79	0.61	0.70	0.69	0.43
PCA + Unskew	0.77	0.64	0.70	0.70	0.42

Table 5.6 shows the results of the three pipelines for the 4-groups aware cluster analysis for the female dataset only.

Table 5.6: Results of k = 4 clusters on females dataset, after retaining statistically significant features only.

F - Aware	S_1	S_2	S_3	S_4	**S**	**A**	σ
Vanilla	0.21	0.28	0.18	0.17	0.21	0.34	1.00
PCA	0.52	0.56	0.54	0.60	0.58	0.33	0.43
PCA + Unskew	0.56	0.53	0.56	0.60	0.58	0.34	0.43

Figure 5.2 displays the percentage of misclustered observation for each voice disorder, divided per pipeline and gender.

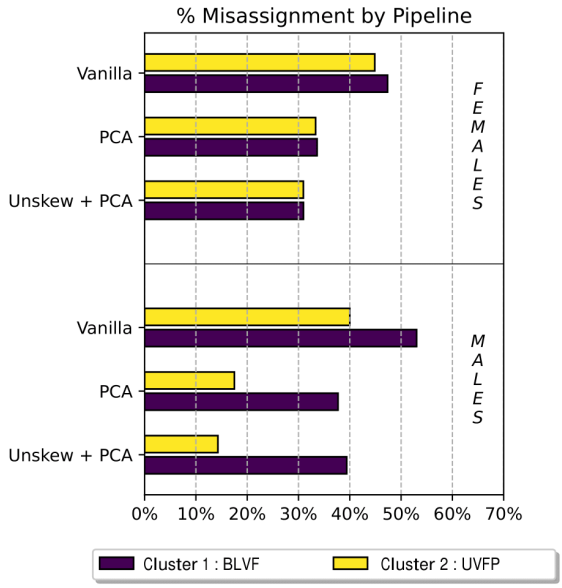


Figure 5.2: Barplot of the percentage of misassignments in each cluster found by the aware condition in the three pipelines.

5.3.2 Machine Learning

E1

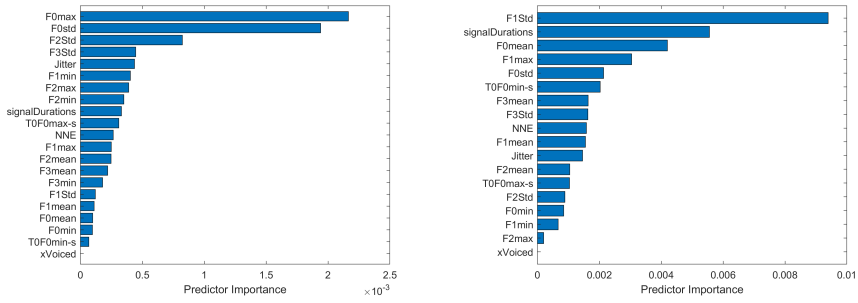
Table 5.7 reports the validation results for the classification experiment between BLVF and UVFP, divided per gender and utilised feature set. Only the models with the highest average accuracy in cross-validation are displayed.

Table 5.7: Results of cross-validation (k-fold=10) for the experiment E1 regarding the binary discrimination between BLVF and UVFP (indicating in the table their True Positive Rate with TPR^{BLVF} and TPR^{UVFP}, respectively). In this experiment, all the subjects were considered without preprocessing according to age. Mean and standard deviation ($\mu \pm \sigma$) statistics regarding the validation performance are reported. The metric used for hyperparameter tuning is highlighted in bold, whereas the feature set leading to the best model is highlighted in grey.

F dataset - E1					
Feature set	**ACC** (%)	TPR^{BLVF} (%)	TPR^{UVFP} (%)	MCC	AUC (%)
/a/	76±3	73±4	77±4	0.50±0.06	79±11
/aiuole/	74±2	74±4	72±6	0.46±0.06	75±12
/a/ + /aiuole/	76±3	76±5	79±4	0.55±0.05	83±9
M dataset - E1					
Feature set	**ACC** (%)	TPR^{BLVF} (%)	TPR^{UVFP} (%)	MCC	AUC (%)
/a/	81±2	84±5	80±5	0.64±0.05	86±14
/aiuole/	73±4	58±10	80±5	0.46±0.07	67±21
/a/ + /aiuole/	81±3	82±5	79±5	0.62±0.06	81±14

Figure 5.3 illustrates the predictor importance for these specific ensemble models.

Table 5.8 displays which acoustic parameters presented a significant difference between BLVF and UVFP in the age-unbalanced datasets for both female and male patients.



(a) F dataset (b) M dataset

Figure 5.3: Predictor Importance for the best ensemble models reported in Table 5.7 using Feature set /a/ both for F dataset (a) and M dataset (b).

Table 5.8: Statistical analysis results for acoustic parameters in E1.

F dataset - Significant features considered in the best model reported in Table 5.7			
Feature	p-value	BLVF - Median (IQR)	UVFP - Median (IQR)
F0 max /a/ [Hz]	e-06	249 (89)	302 (110)
F0 std /a/ [Hz]	e-04	8 (12)	14 (24)
F2 std /a/ [Hz]	0.001	59 (35)	75 (38)
Jitter /a/ [%]	0.007	0.98 (1.25)	1.42 (3.9)
F2 max /a/ [Hz]	0.01	1470 (267)	1522 (202)
Signal duration /a/ [s]	0.02	3.8 (2.1)	3.2 (1.9)
F1 max /a/ [Hz]	e-04	848 (177)	950 (252)
F1 std /a/ [Hz]	0.01	70 (36)	78 (52)
F0 mean /a/ [Hz]	e-04	189 (41)	209 (55)
M dataset - Significant features considered in the best model reported in Table 5.7			
Feature	p-value	BLVF - Median (IQR)	UVFP - Median (IQR)
F1 std /a/ [Hz]	e-07	38 (25)	66 (33)
Signal duration /a/ [s]	e-05	4.7 (2.3)	3.0 (2.1)
F1 max /a/ [Hz]	e-08	652 (137)	778 (105)
F0 std /a/ [Hz]	e-05	2.5 (5.2)	7.7 (27.9)
NNE /a/ [dB]	0.01	-21 (5)	-19 (13)
F1 mean /a/ [Hz]	0.01	547 (113)	598 (105)
Jitter /a/ [%]	e-04	0.84 (0.83)	3.02 (12.64)
F2 std /a/ [Hz]	e-05	40 (18)	57 (33)
F2 max /a/ [Hz]	0.001	1195 (115)	1271 (194)

E2

Table 5.9 reports the results related to the ML models considering an age-balanced dataset only for female patients diagnosed with BLVF and UVFP. A Mann-Whitney test showed that, for adult females, age does not affect the difference between the two pathologies with $p = 0.1534$. The subset comprises 61 subjects with BLVF and 43 with UVFP.

Table 5.9: Results of cross-validation (k-fold=10) for the experiment E2 regarding the binary discrimination between BLVF and UVFP on a balanced version of F dataset according to age. Mean (μ) and standard deviation (σ) statistics regarding the validation performance are reported. The metric used for hyperparameter tuning is highlighted in bold, whereas the feature set leading to the best model is highlighted in grey.

F dataset balanced according to the age - E2					
Feature set	**ACC** (%)	TPR^{BLVF} (%)	TPR^{UVFP} (%)	MCC	AUC (%)
/a/	77±4	80±4	71±8	0.52±0.08	82±11
/aiuole/	69±2	78±9	53±11	0.41±0.08	69±19
/a/ + /aiuole/	72±5	78±6	68±10	0.48±0.09	77±17

E3

Regarding the experiment on the classification of BLVF subclasses, i.e., nodules, polyps, and cysts, Table 5.10 reports the results obtained during the cross-validation.

Table 5.10: Results of cross-validation (k=10) for the experiment E3, on the F dataset regarding the possibility to discriminate among the sub-classes representing BLVF: nodules (acronym=n), polyps (p) and cysts (c). For each class, the AUC values are reported. Mean (μ) and standard deviation (σ) statistics regarding the validation performance are reported. The metric used for hyperparameter tuning is highlighted in bold, whereas the feature set leading to the best model is highlighted in grey.

F dataset - E3					
Feature set	ACC (%)	**MCC**	AUC^n (%)	AUC^p (%)	AUC^c (%)
/a/	51±9	0.26±0.07	57±9	64±5	41±16
/aiuole/	60±3	0.34±0.06	70±21	62±5	64±8
/a / + /aiuole/	54±10	0.32±0.10	73±20	59±17	64±8

5.4 Discussion

Unsupervised Learning

The unsupervised learning experiments served as an exploration study regarding the clusterability of patients diagnosed with BLVF and UVFP based on uncorrelated acoustic features (unaware condition) and significantly different ones between BLVF and UVFP (aware condition).
When considering the two groups classification problem, the best results for both genders were achieved using the aware condition. Indeed, for the female dataset, the average silhouette score and accuracy for the best pipeline (i.e., the Unskew + PCA one) are 0.70 and 0.67, respectively, compared to the unaware condition where $S = 0.64$ and $A = 0.62$. On the other hand, for the male dataset, the aware condition obtained similar results of those concerning the unaware one. In fact, both experiments present $S = 0.70$, but the first achieved a slightly higher accuracy ($A = 0.70$ vs $A = 0.69$). Interestingly, the preprocessing procedure helped improving the silhouette score for the UVFP class, especially for the female dataset. Indeed, clustering went from the $S_2 = 0.37$ of the Vanilla pipeline to the $S_2 = 0.72$ of the Unskew + PCA one. A model with such a property could be helpful in clinical practice because it recognised the most severe pathology, possibly reducing its misdiagnosis [223]. An analogous result was obtained for the male dataset, for which feature transformation allowed to obtain a $S_2 = 0.64$, starting from the Vanilla pipeline value of $S_2 = 0.51$.

The benefits of feature preprocessing were also supported by Figure 5.2, as indeed the upper panel concerning female patients shows a monotonical decrease of the number of misassignments from the vanilla to the Unskew + PCA pipelines and, for this latter one, a balance between misclustered BLVF and UVFP observations. For the male dataset, the lower panel of Figure 5.2 highlights a similar beneficial effect.
Finally, the right panel of Figure 5.1 shows that, even if the proposed approach achieved a good separation of the two considered pathologies, the original UVFP data distribution seemed to be more dispersed than BLVF ones. This could result from the severity degree of the paralysis, which should be considered in future studies as a confounding factor.
Overall, this outcome suggested the existence of unique vocal phenotypes associated with BLVF (taken as a whole class) and UVFP, validating the feasibility of a supervised learning framework.

The ML experiments were developed for the voice pathology identification task using acoustic features only extracted from the sustained vowel /a/ and the Italian word /aiuole/.
E1 vs E2 classification comparison aimed at highlighting the implicit bias of age. This study was performed on adult females only as some pathologies, especially nodules, present a strong gender-dependent prevalence [224]. Results in Tables 5.9 and 5.7 showed that model accuracy did not improve when considering a subset where age did not significantly differ between BLVF and UVFP. Notably, ageing has a direct influence on vocal properties, an outcome that is supported by both statistical analysis [225–227] and AI-based methods [228], but the similar age distribution in our dataset and the chosen tasks might have reduced this effect. For /aiuole/, the accuracy changed from 69% of balanced E2 to 74% of unbalanced E1. Thus, an articulation task seemed more sensitive to age than a sustained vowel. However, Spazzapan *et al.* [229] demonstrated that, for female subjects, both sustained phonation and speech tasks highlighted statistical differences between age groups. Therefore, it seems implausible that one of the two tasks performed better in identifying UVFP than BLVF or vice versa due to an age bias on vocal properties. According to these results, further analyses were performed on the unbalanced dataset only for both genders.

Supervised Learning

For the F dataset, the concurrent use of the acoustic features of /a/ and /aiuole/ did not bring relevant improvements. Indeed, as Table 5.7 illustrates, the accuracy was the same (76%). The /a/+/aiuole/ classifier presented a higher AUC value than the /a/ one. However, such a result was obtained with a larger number of features (40 vs 21). Therefore, further examination and importance analysis of the predictor was performed on the /a/ classifier. The analysis of the sole utterance /a/, as well as reducing the number of parameters, could be helpful for clinicians. The acoustic features extracted from the sustained vowel gave a higher TPR for the UVFP class, confirming a previous outcome in [230]. This result can be of clinical relevance as a higher sensitivity in detecting vocal fold paralysis may reduce the possibility that false-positive patients undergo long and expensive diagnostic procedures such as CT or MRI [231]. Unexpectedly, Figure 5.3a shows that the most relevant parameters in distinguishing BLVF and UVFP were F0 mean and its instability F0 std. Indeed, the vocal folds dynamic is quite different between these two pathologies. Moreover, the standard deviation of F2 was considered relevant by the AdaBoost ensemble model. The second formant is related to the degree of constriction of the tongue movements [213]. This result may suggest to otolaryngologist the use of imaging techniques such as ultrasonography or electromyography to track its activity to support and validate their diagnosis or treatment monitoring. The statistical analysis results in Table 5.8 show that the three most relevant parameters were also significant. UVFP presented higher values for the F0 max, F0 std, and F2 std than BLVF, which may underline higher vocal fatigue and instability [232].
For the M dataset, E1 highlighted results similar to those of the F group. No improvement was achieved by using the combined feature sets. The best classifier, in terms of accuracy, AUC, and number of parameters, was the /a/ RobustBoost. It was characterised by high TPR values for both BLVF and UVFP, 84% and 80%, respectively. Table 5.7 also shows that the sensitivity of voice pathology identification may depend on the vocal task, agreeing with [230]. Indeed, with the acoustic features of /a/, the TPR for benign lesions was higher than with the parameters of /aiuole/ (84% vs. 80%). In contrast, with the metrics of /aiuole/, the TPR for vocal fold paralysis was relevantly higher for the features of /a/ (80% vs 58%). A similar hypothesis was tested by Hosokawa *et al.* [233], showing that connected speech was

better correlated with perceptual assessments of dysphonia than sustained phonation. However, this analysis was carried out between disordered voice and normophonic subjects rather than across different pathologies, as in the present study. Nevertheless, this result suggested that some vocal tasks might be more appropriate for diagnosing and detecting specific voice disorders. Importantly, this outcome must be taken cautiously as the size of the M dataset is lower than the F one.
The most relevant acoustic feature for the E1 in the male group was F1 std (Figure 5.3b), which may be related to a difference in the instability of pharynx positioning and its degree of constriction [213]. Also, the overall signal duration, which can be considered an indicator of a patient's maximum phonation time, could represent a relevant clinical parameter to facilitate the diagnosis. Finally, the F0 mean appeared to be the third most relevant parameter for distinguishing the two pathologies. Overall, Figures 5.3a and 5.3b show an agreement for both M and F datasets between acoustic parameters capable of separating BLVF and UVFP. However, the gender-specific relevance order in the predictor importance outcome suggested a distinct relationship between acoustic parameters and voice pathology that could be due to direct morphological differences in vocal fold and tract structures. Finally, Table 5.8 shows that F1 std and signal duration significantly differed between BLVF and UVFP for the M dataset. The instability of the pharynx configuration was higher for the UVFP class, reflecting greater vocal fatigue.

Due to the small M dataset sample size concerning nodules and cysts, E3 was performed for female patients only. Table 5.10 demonstrated a noticeably lower performance than the E1 and E2 classification outcomes. The highest accuracy was achieved for the AdaBoost /aiuole/ classifier with a value of 60%, which aligned with the results of Marchese *et al.* [207]. The AUC for nodules (70%) was higher than those for polyps and cysts (62% and 64%), which could indicate that the acoustic features of /aiuole/ may be sensitive to the dimensions of vocal fold masses. Indeed, Rostampourgonbki *et al.* [234] proved that F0, jitter, and noise measures significantly differed between nodules, polyps, and cysts in female patients. However, this was not verified with predictor importance analysis in the current study due to the small sample size.

General Discussion

These findings demonstrate that explainable supervised and unsupervised AI techniques can provide clinicians with interpretable, physiologically meaningful information to guide diagnostic decisions. Such systems highlight the most discriminative vocal parameters, offering novel perspectives for differential diagnosis. As the prevalence of voice disorders, and especially of organic nature, could severely increase if mitigation strategies for air pollutants management are not rapidly implemented, this approach may support and improve early identification of laryngeal pathologies and reduce reliance on long, invasive or costly imaging procedures. This latter effect could be beneficial also for diminishing the healthcare carbon footprint by realising a more environmentally sustainable service delivery. Furthermore, these results also open the way toward patient-centred and remote monitoring strategies. The integration of explainable acoustic models into digital platforms as low-cost telehealth solutions could enhance individual health awareness, enabling patients to track their vocal status and recognise early signs of dysfunction. In future work, the implementation of self-supervised speech representation models, such as HuBERT and similar transformer-based architectures [235], should be explored. In the context of voice pathology assessment, these pretrained models can be leveraged as feature extractors, with a task-specific final layer added to carry out the actual classification problem. This approach could improve models' performance with respect to traditional handcrafted acoustic features. However, this gain in accuracy would potentially be accompanied by a reduced level of interpretability, as the learned representations do not map directly onto well-defined physiological or perceptual voice attributes [236]. To mitigate this limitation, the integration of dedicated explainability techniques becomes even more necessary. However, a wider database should be collected to reduce the risk of overfitting and hinder model generalisability. For these reasons, while deep learning-based embeddings represent a promising direction for future investigations, the present study prioritised explainable models grounded in interpretable acoustic features.

5.5 Conclusion

This chapter explored the potential of using acoustic features to differentiate between two voice disorders, i.e., benign lesions of the vocal folds and unilat-

eral vocal fold paralysis, through both unsupervised and supervised machine learning approaches to support the development of eco-friendly healthcare AI-based tools.

In the unsupervised learning experiments, the clusterability of patient data was assessed using acoustic features selected in two conditions: unaware (uncorrelated features) and aware (uncorrelated and statistically different features between BLVF and UVFP). Results showed that the aware condition consistently outperformed the unaware one, particularly for female patients. Notably, feature preprocessing (specifically, Unskew + PCA) improved clustering performance. Such a model could assist in clinical settings by better identifying more severe conditions like UVFP and reducing its misdiagnosis.

In the supervised learning experiments, binary classification tasks using features from both sustained phonation and word articulation confirmed that accurate identification of BLVF and UVFP is feasible using acoustic parameters alone. Evaluation of age-related effects in female patients showed minimal bias, validating the use of age unbalanced datasets for further analysis. For both genders, models based on the sustained vowel /a/ only achieved high accuracy and sensitivity, which is particularly important for UVFP detection, where a higher true positive rate could reduce the need for costly diagnostic imaging. Moreover, the analysis of feature importance revealed that F0 mean, F0 std, and F2 std were the most relevant predictors for female patients. For males, F1 std, signal duration, and F0 mean were the most informative. These results suggested a gender-specific relationship between acoustic parameters and underlying pathology, likely influenced by anatomical differences in vocal fold and tract structure. Finally, a multiclass classification task among nodules, polyps, and cysts achieved limited performance. Despite the modest results, higher AUC for nodules suggested that certain acoustic features may be sensitive to lesion size. However, the small sample size limited definitive conclusions and highlighted the need for further data collection.

Chapter 6

Climate Change and Health: Insight on Post-Treatment Voice Monitoring

The digitalisation of healthcare services aims at reducing the high carbon footprint and making them more accessible to fragile population who can not reach primary care facilities. This chapter proposes two studies that are focused on the automation of patients' follow-up after medical treatment of voice disorders. In the first study, machine learning classifiers achieved a good performance in recognising pre- and post-treatment voices, also highlighting which acoustic parameters were most relevant in such a distinction. In the second study, the concept of complexity matching was implemented to better detect patients that have not fully recovered their vocal output, in order to plan more quickly and efficiently the following interventions. [1]

[1]The first study has been partially published in "Towards an explainable Artificial intelligence system for voice pathology identification and post-treatment characterisation" in *Biomedical Signal Processing and Control, v. 104, p. 107530, 2025* [175]. The second one has been published as "On the Complexity Matching and Multiscale Nonlinear Perspective of Voice Restoration via Fat Injection Laryngoplasty in Unilateral Vocal Fold Paralysis" in *Scientific Reports, v. 15, n.1 , p. 31801, 2025.* [237]

6.1 Introduction

Voice analysis plays a critical role in the monitoring of the progression or recovery of vocal function after treatment. Indeed, objective acoustic measures can be used to track changes over time in response to surgical interventions, pharmacological treatments, or logopaedic approaches to evaluate the efficacy of interventions. Moreover, monitoring voice signals offers a non-invasive and cost-effective method that can be performed even in remote settings using portable devices or smartphones. As such, voice becomes a dynamic biomarker of vocal health and therapeutic response. Notably, treating voice disorders is not merely a matter of improving acoustic output. It is fundamentally tied to a person's psychosocial well-being and quality of life. Individuals with dysphonia frequently report limitations in communication, increased vocal effort, and feelings of frustration or social withdrawal. These effects are even more pronounced in vocally demanding professions such as teaching, acting, or singing, where vocal performance is essential to professional identity and livelihood. Hence, even if effective treatment mostly aims at restoring a proper functional communication, it can also indirectly enhance self-esteem, reduce anxiety, and improve overall life quality [238].
Recently, AI techniques has been implemented as alternative to statistical comparisons to develop automated, scalable tools that are accessible for widespread clinical and non-clinical use. In this, ML is particularly advantageous as it needs very few assumptions about the data-generating systems, which are instead relevant to obtain reliable statistical results (and, at the same time, they are not often met in clinical practice). ML models are capable of generalising underlying data patterns for prediction with completely new observations. Furthermore, several studies have highlighted that even when data do not show statistically significant differences, ML can still effectively identify parameters to differentiate observations [212]. This approach managed to recognise patients diagnosed with multiple vocal diseases that underwent voice therapy (achieving 87% accuracy in combination with clinical information) [239], with head and neck cancer [240] and COPD [241]. Moreover, Suppa *et al.* [242] investigated with an Artificial Neural Network the change in voice quality before and after the injection of botulinum toxin in patients diagnosed with spasmodic dysphonia. Bensoussan *et al.* [243] used a previously trained classifier on cis-male and cis-female voices to investigate the efficiency of gender reassignment therapies on a cohort of trans-female patients. These studies demonstrated that AI can be successfully deployed

for such purposes. Yet, they need further examination to determine which type of therapy can restore or achieve euphonia and to analyse how vocal properties evolve over time. In this context, attention to interpretability is still a clinically relevant aspect for assessing changes in voice quality, as already introduced in Chapter 5.

Concurrently, implementing novel signal processing techniques could enhance the sensitivity and specificity in the assessment of disordered voice, enabling a more nuanced tracking of therapeutic progress. Indeed, the computation of perturbation measures, which directly depend on the estimation of the F0, can become extremely challenging and unreliable, especially in patients that manifest highly irregular, breathy and partially aphonic utterances [244,245]. Moreover, recent literature works demonstrated the existence of several nonlinear phenomena in voice production that the algorithms embedded in common tools used for acoustic characterisation are insensitive to. Therefore, the mismatch between the mathematical framework and the biosignals attributes can lead to ambiguous results in acoustic analysis, hindering its advantages and applicability, [244, 246] as well as impeding a clear, robust evaluation of treatments efficiency. For instance, subglottic airflow turbulences, rheological and mechanical collision properties of the vocal folds, and asymmetry in left and right vocal fold movements may generate unique characteristics that require novel and alternative techniques to extract and investigate them [179, 247]. To address such a gap in the literature, nonlinear dynamical systems theory has been proposed as a promising candidate to quantify dysphonic voices with a broader perspective [244, 248].

The aim of this chapter was two fold:

- Study 1 investigated the possibility to characterise pre- and post-treatment voice quality with a XAI-based approach, encompassing results interpretability and misclassification errors management.

- Study 2 concerned the application of the complexity matching concept on voice recordings to delve deeper into the degree of recovery of UVFP patients treated with autologous fat injection laryngoplasty. A nonlinear, multivariate and multiscale approach was also introduced to support the latter.

6.2 Study 1

6.2.1 Patients

A dataset of 90 patients' voices (61 F, mean age = 45.9 ± 13.4 years, 29 M, mean age = 49.4 ± 16.4 years) that underwent diverse types of medical intervention was available for analysis. These subjects are a subgroup of the 287 patients database introduced in Section 5.2.1. Audio files were recorded following the same procedure. A perceptual evaluation with the GRB scale [249] was performed blindly by a speech therapist with long experience in voice disorders. The GRB scale is composed of three items:

- G (global grade of dysphonia): the judgement is based on the overall impression of voice quality deterioration.
- R (Roughness): the impression of irregular F0 and noise.
- B (Breathiness): turbulent noise related to air escape through the vocal folds.

The scores ranged from 0 to 3, where 0 is the rating for the best voice condition and 3 for the worst one.

6.2.2 Experimental Procedure

Feature extraction was performed with the BioVoice software [215] after manual segmentation, obtaining the same parameters displayed in Table 5.2 from /a/ and /aiuole/ vocal tasks.
Before treatment, all pathological patients were considered belonging to a single class, analogously to the post-treatment ones. Besides predictor importance, explainability was also actuated by analysing whether the models' errors in the post-treatment class were associated to a specific pre-treatment pathology (i.e., if the post-treatment error referred to a nodule, polyp, cyst, or UVFP). Additionally, the identification tag of each misclassified recording was stored to investigate a possible agreement between the AI model's outcome and the subjects' clinical voice assessment using the GRB scale [175]. Specifically, a Mann-Whitney test ($\alpha = 0.05$) was performed by considering G, R, and B indices between correctly classified and misclassified audio recordings.

The trained classifiers, the hyperparameters optimisation procedure, the feature selection methods, and feature importance assessment were the same as the ones presented in Section 5.2.4.

6.2.3 Results

The results are reported in Table 6.1. The MCC metric was used during the hyperparameters optimisation. Only models with the highest average MCC obtained during cross-validation are displayed. This analysis was performed for the F and M datasets separately, considering all the feature sets (/a/, /aiuole/, and /a/+/aiuole/).

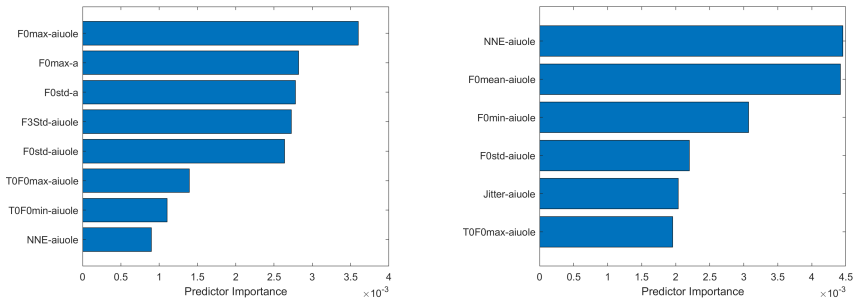
Table 6.1: Cross-validation results related to binary pre vs. post evaluation. In this case, the pre-class merges the observations for the two groups considered (BLVF and UVFP), and the post-class represents the recording related to the post-treatment analysis for both groups. Mean (μ) and standard deviation (σ) statistics regarding the validation performance are reported for F and M datasets. The metric used for hyperparameter tuning is highlighted in bold, whereas the feature set leading to the best model is highlighted in grey.

F dataset					
Feature set	ACC (%)	TPR^{pre} (%)	TPR^{post} (%)	**MCC**	AUC (%)
/a/	76±3	84±4	57±8	0.45±0.05	78±10
/aiuole/	75±2	85±4	53±5	0.43±0.05	75±12
/a/ + /aiuole/	77±4	83±7	64±6	0.52±0.07	81±11
M dataset					
Feature set	ACC (%)	TPR^{pre} (%)	TPR^{post} (%)	**MCC**	AUC (%)
/a/	71±5	81±3	55±8	0.46±0.05	72±14
/aiuole/	72±3	76±4	68±7	0.48±0.06	74±13
/a/ + /aiuole/	70±5	71±4	65±6	0.40±0.09	66±16

The predictor importance for both datasets is shown in Figure 6.1.

To investigate possible relationships between AI model output and perceptual evaluation, only the models with the highest accuracy shown in Table 6.1 were considered for both groups.

For the F population, a Mann-Whitney test highlighted that G, R, and B indices were significantly different between correctly classified and misclassified recordings of the /aiuole/ vocal task with $p-value = 0.02$, $p-value = 0.001$,



(a) F dataset (b) M dataset

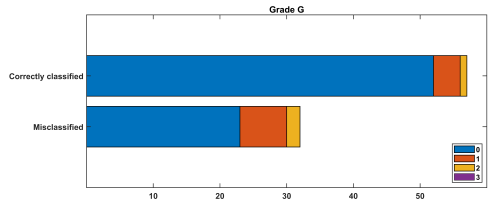
Figure 6.1: Predictor Importance for the best ensemble models reported in Table 6.1 regarding the pre- vs post-treatment analysis using feature set /a/+/aiuole/ for the F dataset (a), and feature set /aiuole/ for the M dataset (b).

and $p - value = 0.003$, respectively. Figure 6.2 displays the distribution of the perceptual assessment indices.

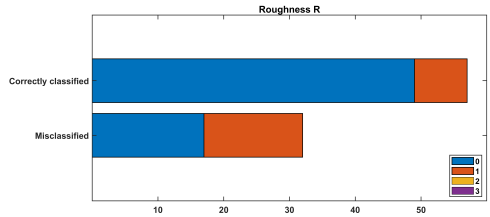
Moreover, a significant difference was also found for the B index relative to the /a/ vocal task with $p - value = 0.01$. Figure 6.3 shows the distribution of its values across correctly classified and misclassified post-treatment recordings.

For the M group, the misclassified recordings extracted from the /aiuole/ classifier did not present any significant difference in G, R, and B indices compared with correctly identified observations.

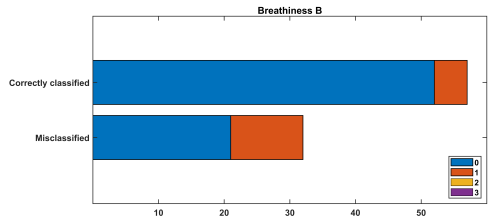
The highest classification error was related to the UVFP. The best model of the F dataset misclassified 27 observations. Their pre-treatment classes were UVFP (56%, 18 cases), nodules (3%, 1 case), polyps (25%, 8 cases) and cysts (15%, 5 cases). The best model of the M dataset misclassified 19 cases. Their pre-treatment classes were UVFP (95%, 18 cases) and polyps (5%, 1 case).
Finally, Table 6.2 displays which acoustic parameters presented a significant difference between pre- and post-treatment conditions for female and male patients.



(a) G values distribution.



(b) R values distribution.



(c) B values distribution.

Figure 6.2: Distribution of G, R, and B index values across correctly classified and misclassified observations from Table 6.1 best classifiers for the F dataset.

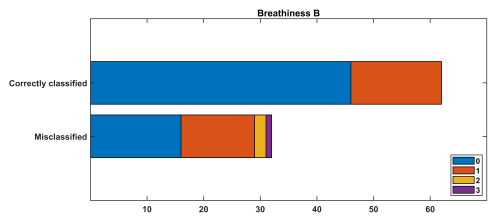


Figure 6.3: Distribution of B index values across correctly classified and misclassified observations from Table 6.1 best classifiers for the M dataset.

Table 6.2: Statistical analysis results for acoustic parameters.

F dataset - Significant features considered in the best model reported in Table 6.1			
Feature	p-value	PRE - Median (IQR)	POST - Median (IQR)
F0 max /aiuole/ [Hz]	e-09	286 (131)	223 (65)
F0 max /a/ [Hz]	e-10	270 (99)	212 (48)
F0 std /a/ [Hz]	e-10	11 (19)	4 (6)
F3 std /aiuole/ [Hz]	e-5	387 (137)	459 (161)
F0 std /aiuole/ [Hz]	e-10	28 (32)	17 (20)
M dataset - Significant features considered in the best model reported in Table 6.1			
Feature	p-value	PRE - Median (IQR)	POST - Median (IQR)
NNE /aiuole/ [dB]	0.03	-17 (7)	-18 (7)

6.2.4 Discussion

In the pre-post comparison, the best model for the F dataset was an /a/ + /aiuole/ RobustBoost classifier with a high accuracy of 77%. Table 6.1 shows that the TPR for the pre-class was slightly worse and less stable than the one from the /a/ classifier (83±7% vs 84±4%). However, the TPR for the post-class was higher (64% vs 57%). Therefore, the concurrent use of acoustic features extracted from different vocal tasks helped improving the sensitivity to detect post-treatment subjects. Even if TPR^{post} was relatively low, TPR^{pre} was promising. A high value means that, with the selected features, models tend to classify post-treatment recordings as *pre* treatment ones, and not vice versa. Thus, a patient diagnosed with a voice disorder is not considered a voice-recovered subject. Therefore, this model may reduce the possibility of under- or misdiagnoses. This could help clinicians monitor and evaluate the progress of therapies. Indeed, a post-treatment subject classified as a pre-treatment one could indicate a poor outcome of the intervention itself. Figure 6.1a highlights that relevant acoustic features were extracted from the /aiuole/ vocal task, with the most important one being F0 max. Hence, the subsequent articulation of four vocalic sounds could be a biomarker to detect a low voice quality after treatment. Moreover, vocal fold vibratory patterns expressed by F0-related metrics represented a meaningful characteristic that otolaryngologists should consider to assess the difference before and after therapy. Interestingly, all relevant metrics were related to the fundamental frequency except for the noise measure NNE extracted from /aiuole/. The examination of misclassified observations led to an investigation regarding which class presented the highest error rate. Then, a compari-

son was performed between classification errors and perceptual assessment of post-treatment voices to discover possible agreement between the results obtained with the proposed AI methods and the perceptual evaluation (Figure 6.2-6.3). The first analysis highlighted that the most misclassified class was the UVFP. In the current dataset, this pathology was surgically treated with a particular injection laryngoplasty procedure (see Section 6.3). Although beneficial, it does not intervene on muscles or nerves; therefore, the motility of vocal folds is not entirely restored. As a result, voice quality might not be relevantly improved compared with the overall pre-treatment class or with the techniques involved for treating BLVF. Figure 6.2a seems to confirm this hypothesis, as a larger number of high G post-treatment recordings were misclassified as the pre-class. Similar considerations can be made for roughness and breathiness indexes (Figures 6.2b and 6.2c). Moreover, the NNE extracted from /aiuole/ was relevant to distinguish pre- and post-treatment audio samples (Figure 6.1a). Similarly, the R-value of misclassified observations was significantly higher with respect to correctly identified ones. It denoted a possible agreement between perceptual assessment and the classification outcome. It could also confirm that the NNE parameter is related to the R index of the GRB scale [250], similarly to the HNR [251]. Finally, Figure 6.3 shows that breathiness B values for /a/ were higher for misclassified observations. Breathy voices typically characterise UVFP due to an incomplete glottic closure. Therefore, the results of our model could underline a classification rule based on the UVFP pathophysiology.

Table 6.1 shows that the best classifier for the M dataset was the RobustBoost model trained with the acoustic features of /aiuole/ only. Indeed, even if TPR^{pre} was lower than the /a/ model (76% vs. 81%), the TPR^{post} was notably higher (68% vs. 55%). This means that fewer pre-treatment voice samples were misclassified as post-treatment ones. From Figure 6.1b, the best acoustic metrics were NNE and F0 mean. Hence, noise level and vocal fold dynamics differed in the pre- and post-condition. No significant variation was identified in GRB values, possibly because the number of misclassified observations was too low to carry out a reliable statistical test. Indeed, when looking for the class that presented the highest error rate, it was found that the overall number of mistakes was equal to 19, out of which 95% were UVFP. This confirmed F dataset results: when considering BLVF and UVFP, models might face greater difficulty in separating pre- and post-treatment recordings due to the different surgical operations adopted [245].

Such an outcome could be reflected and explained by the relevance of F0 mean, as the surgical removal of vocal fold masses effectively restores their physiological motility. However, more data from a male cohort is needed to validate these hypotheses.
Differently from the predictor importance analysis displayed in Figure 5.3, Figure 6.1 shows that the pre/post-treatment models have considered alternative relevant features to separate the two classes, which, additionally, were also derived from the other vocal task /aiuole/. This might indicate time-intrinsic differences due to structural factors. Moreover, the different predictor importance outcome with respect to experiment E1 (5.3.2) may suggest that the relevant features for voice pathology recognition were not identical to those for pre/post-treatment voice quality separation. Therefore, such acoustic parameters should not be used interchangeably for these classification experiments and, possibly, in clinical practice.

6.3 Study 2

UVFP is a voice disorder characterised by impaired motility of a single vocal fold. In the general population, the incidence of UVFP among voice pathologies is estimated at 1.2% [252]. Its etiopathogenesis is wide and includes mechanical trauma to the head and neck, neoplasms, neurological diseases, idiopathic origin and, most commonly, iatrogenic causes (typically represented by injuries occurring in thyroid and cardiothoracic surgery). These phenomena lead to nervous system damages, causing abnormal laryngeal muscle contractions and limiting the adduction of one of the vocal folds. The most common and perceivable symptom of UVFP is dysphonia, with patients experiencing breathy voices due to glottal insufficiency. Moreover, UVFP may cause speech dyspnoea, swallowing problems and, in most severe cases, body stabilisation difficulties after physical exercise.
When logopaedic treatments are unsatisfactory, surgery becomes necessary [245]. There is no clinical consensus regarding the best intervention to close the glottic gap and recover the mucosal wave. A promising techniques is injection laryngoplasty (IL), which inserts a mouldable and biocompatible material (e.g., autologous fat and fascia) to medialise the paralysed vocal fold [245]. The main advantages of this approach are material wide availability, cost-effectiveness, that it does not require open surgery, in some cases it allows for adipose tissue revascularisation, and it does not hamper spon-

taneous reinnervation. Moreover, feedback on voice improvements can be collected in real-time: this is an important aspect as it can counteract the main and inevitable disadvantage of IL, fat reabsorption, because it may help the phonosurgeon to regulate the over-injection. With specific preparation procedures, the beneficial effect of IL on voice outcomes can last up to 5 years.

As a support to gold standard methods for treatment monitoring, features that do not rely on the computation of the F0 and can also account for well-known voice nonlinearities have been proposed. In acoustic analysis, measures describing the state space geometrical properties and trajectories (e.g., the correlation dimension) were firstly explored a few decades ago. However, they require speech dynamics to be purely deterministic and such an assumption can be proven only under specific circumstances. Moreover, this modelling approach does not account for randomness [237]. Therefore, information theory parameters (such as approximate and sample entropies) have been introduced to overcome this issue, since they do not consider the deterministic or non-deterministic nature of the signal. Additionally, self-similarity measures, as detrended fluctuation analysis (DFA) [253], have been implemented to study chaotic vibration.

Importantly, these features assume that the scale invariance does not depend on time and space. However, such variations often occur in biomedical signals, including voice and speech, and consequently indicate a multifractal, rather than monofractal, structure [254]. Therefore, multifractal DFA (MFDFA) could be a more adequate and promising technique to analyse and characterise the multiple components, interactions, and scales that the voice production system involves over time. Notably, MFDFA has been also implemented to study the so-called complexity matching (CM) between interacting systems. CM refers to a maximised flow of information between systems when they share similar complexity. Its occurrence has been demonstrated in participants that engaged in joint activities, as well as between two different effectors (i.e., lower limbs) of the same individual [255]. CM is still an unexplored phenomenon in speech analysis, as it has been mostly applied in psychoacoustic studies involving dyadic conversations [256]. Thus, more research is needed to understand its capabilities when considering disordered voices and pre/post treatment comparison.

6.3.1 Patients

A total of 69 participants were included in the study, of which 40 are females (mean age = 49.3 ± 12.0 years) and 29 are males (mean age = 49.7 ± 14.6 years). They were diagnosed with unilateral vocal fold paralysis through videolaryngostroboscopy with a flexible endoscope or a 70° rigid fiberoptic endoscope, supported by voice perceptual assessment performed with the GRB scale [249] (6.2.1). Patients affected by UVFP not achieving a full recovery of voice quality with logopaedic therapy were considered eligible for autologous fat IL (AFIL). Patients underwent surgical treatment with injection laryngoplasty [237]. To evaluate the efficiency of this procedure, voice quality was measured pre- and post-operatively with aerodynamic, auditory and acoustic analyses. In both conditions, recordings were acquired using a C1000S microphone (AKG Acoustics GmbH, Vienna, Austria) with a sampling frequency of 44.1kHz at a fixed distance of 5cm from the patient's mouth. Participants were asked to utter a sustained vowel /a/ for at least 3s at comfortable pitch and loudness.

6.3.2 Signal processing & Nonlinear analysis

Audio recordings were downsampled at 22.05kHz for computational efficiency and automatically segmented, selecting a central section of about 743ms (precisely 2^{14} samples, i.e., the power of two number of samples that is closest to 1000ms, to guarantee equal window size when scaling) to remove the unstable parts concerning vocal onset and offset.
Multi-Fractal Detrended Fluctuation Analysis, introduced in [257], can be divided into 5 steps that roughly follow the same procedure developed in [253] for Detrended Fluctuation Analysis.

It starts by centring and integrating the time series *x(i)* of length N according to Equation 6.1. Figure 6.4 displays this operation applied on a pre-treatment acoustic signal from our dataset.

$$X(k) = \sum_{i=1}^{k} [x(i) - \frac{1}{N} \sum_{i=1}^{N} x(i)] \tag{6.1}$$

In the second step, the time series X(k) is divided into N_n non-overlapping segments of length n, with s = 1, ..., N_n according to predefined scale sizes. In each window, a (typically linear) trend is computed with Equation 6.2.

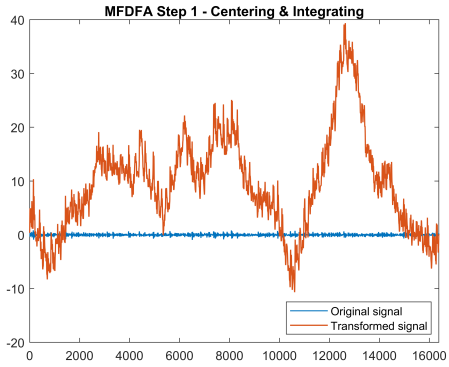


Figure 6.4: The original time series is centred and integrated, obtaining the orange signal.

$$z_s(k) = X_s(k) - a_s - b_s k \tag{6.2}$$

where the constants a_s and b_s are obtained after a linear fit. Figure 6.5 shows this operation using three scales (s = 128, s = 256, s = 512) used in this study.

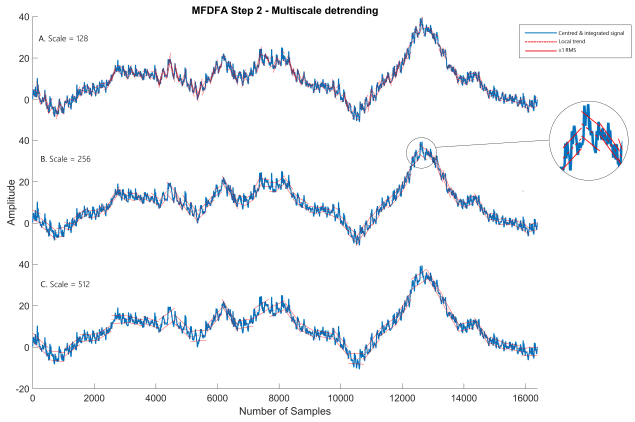


Figure 6.5: Trend computation in each segment across three different scales. A) s = 128, B) s = 256, C) s = 512.

After detrending, step 3 provides the calculation of the Root Mean Square (RMS) of the residuals (i.e., the deviation from the trend), obtaining the fluctuation function $F^2(n,s)$ for each scale s according to Equation 6.3. Figure 6.6 shows the $F^2(n,s)$ for the same above-mentioned scales.

$$F^2(n,s) = \frac{1}{n} \sum_{k=(s-1)n+1}^{sn} [X(k) - X_{n,s}(k)]^2 \tag{6.3}$$

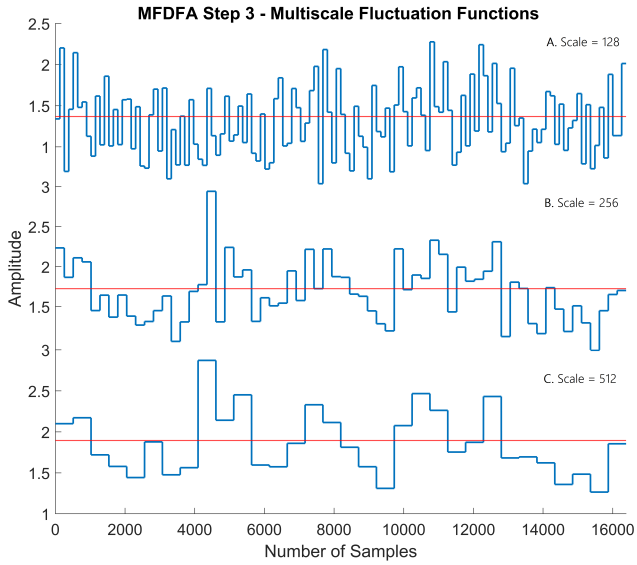


Figure 6.6: Fluctuation functions for three different scales. The red horizontal line represents the mean of $F(n,s)$.

At step 4, MFDFA deviates from the original DFA method to account for the multifractal structure of a signal. Indeed, the fluctuation functions are computed for different orders than $q = 2$, then averaged to derive the q-order fluctuation functions as in Equation 6.4. This allows highlighting the segments where extremely small or large RMS occurs, as Figure 6.7 illustrates.

$$F_q(n) = \{\frac{1}{N_n} \sum_{s=1}^{N_n} [F^2(n,s)]^{q/2}\}^{1/q} \tag{6.4}$$

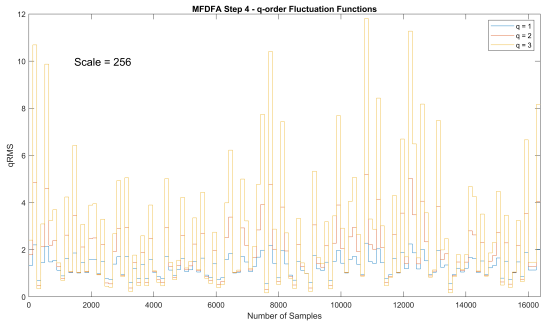


Figure 6.7: Fluctuation functions obtained considering a single scale (s = 256) and three different q-order moments (q = 1, q = 2 for reference to the B. panel of Figure 6.6 and q = 3). Since positive q are selected, only the largest variations are emphasised.

Finally, MFDFA step 5 can be subdivided into three smaller passages. Firstly, the $F_q(n)$ are plotted against the corresponding scale size n in a loglog graph (see Figure 6.8) and, if a long-term relationships exist, then $F_q(n)$ should increase with n following a power law:

$$F_q(n) \propto n^{h(q)} \tag{6.5}$$

where $h(q)$ is the scaling exponent computed as the slope of the linear regression of log $F_q(n)$ against log n.

Then, scaling exponents are converted into Renyi exponents $\tau(q)$ by Equation 6.6:

$$\tau(q) = qh(q) - 1 \tag{6.6}$$

from which the Lipschitz-Hölder exponents $\alpha(q)$ are obtained with:

$$\alpha(q) = \frac{d\tau(q)}{dq} \tag{6.7}$$

Therefore, MFDFA requires setting three parameters: the scale range, the q-order that weights local variations and the polynomial order for the detrending.

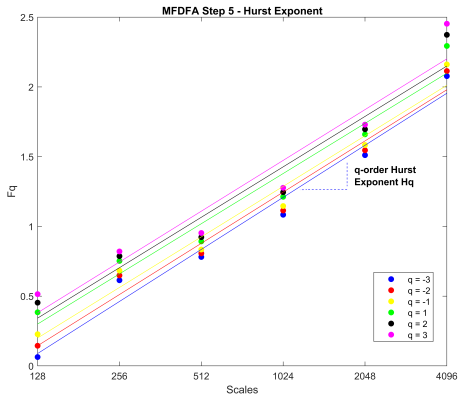


Figure 6.8: Hurst Exponent visualisation on a loglog graph for a set of different orders q and scales s.

Considering a mean F0 of 200Hz and the sampling frequency of 22.05kHz, the minimum segment size should be greater than 110.25 samples [254]. This value was later approximated to the closest power of two number, i.e., 128 samples. Furthermore, as suggested in [255], since physiological time series often present different scaling regimes over the short and long term, it is advisable to estimate the multifractal exponents firstly over the entire set of available ranges (in this context, from 2^7 to N/2, i.e., 2^{13}), then considering more limited sets, progressively excluding the shortest intervals. This strategy should better highlight a true CM effect between interacting systems and discard local corrections. Therefore, four scale ranges were implemented: the longest range (LLR, from 2^7 to 2^{13}), the long range (LR, 2^8 - 2^{13}), the short range (SR, 2^9 - 2^{13}) and the shortest range (SSR, 2^{10} - 2^{13}).
The q-order decides the weighting of local fluctuations. In a monofractal signal, the absence of both large and small magnitude fluctuations allows for the variation of the time series itself to be fully described by its variance, i.e., its second-order statistical moment. This is not the case for a signal exhibiting a multifractal structure. Therefore, multiple q-order statistical moments should be considered. To account for periods with both small and large variations, the set of q-order should encompass both negative and positive values. Additionally, to perform a more precise computation, large (in absolute value) orders should be avoided. This study investigated multifrac-

tality in the [-15, 15] range, according to [255].
Finally, as an exploratory analysis, the polynomial order implemented for detrending the signal was set to 1.
After calculating the individual q-order singularity exponents $\alpha(q)$, the CM was obtained by computing the correlation coefficient of $\alpha_1(q)$ and $\alpha_2(q)$, where $n = 1, 2$ refers to pre- and post-surgical treatment conditions, respectively. Since this study is interested in analysing how much information is retained after medical intervention in voice signals (which might reflect treatment inefficacy), the UVFP cohort was arbitrarily split into two populations based on the change of grade G of the GRB scale. Patients presenting a decrease of magnitude two or higher and a passage from G = 1 to G = 0 were considered as the fully recovered group (FR), whereas lower decrease values represented the mild recovery (MR) group. Moreover, to understand whether a common pattern of the disease recovery process exists or, rather, it is characterised by variations that are specifically related to each subject, the order of data for the MR group was randomised in the sets of α exponents from both pre- and post-operative conditions. It was hypothesised that in the case of patient-specific variations, the correlation coefficient value would drop.

A nonlinear multiscale analysis was also performed. It included the following parameters: correlation and fractal dimension (calculated as in [258]), the largest Lyapunov exponent, Sample Entropy (SE), and Hurst Exponent. Excluding the latter, each of them required the reconstruction of the state space. This operation was performed at each scale and individually for each recording, considering the first minimum of the mutual information function for computing the time delay τ and the false nearest neighbour method to calculate the embedding dimension, which was also retained as an additional feature to describe the physical properties of the state space. Moreover, regarding sample entropy, the tolerance was set at 20% of the signal standard deviation. A coarse-grained approach was employed to generate the scaled signals, according to Equation 6.8 [259].

$$y^{(s)}(j) = \frac{1}{s} \sum_{i=(j-1)s}^{js} x(i), 1 \leq j \leq \lfloor \frac{N}{s} \rfloor \tag{6.8}$$

Where *x(i)* is the i-th sample of the original time series of length N, s is the power of two scale factors (varied from 0, the original signal, to 7), and

the symbol $\lfloor . \rfloor$ denotes the integer part of its argument.

6.3.3 Statistical analysis

As explained in Section 6.3.2, after computing the scaling factors for each scale range and q-order moments, a Spearman correlation analysis was performed to discover possible relationships between pre- and post-treatment recordings. Correlation coefficients were computed separately for the FR and MR groups. Moreover, another correlation analysis was performed within the MR group only by randomly assigning patients to highlight individual or common variations. Finally, the same correlation coefficients were averaged within the four scale ranges (SSR, SR, LR, LLR) and separate t-tests were performed to underline different CM effects in each ($\alpha = 0.05$).
A preliminary multiscale analysis was conducted by considering the whole sample. As counter-intuitive results were obtained, it was decided to separate the population into the same FR and MR groups. After such a division, each nonlinear metric underwent either a paired sample t-test or a Wilcoxon signed test ($\alpha = 0.05$) for all scales individually.

6.3.4 Results

Figure 6.9 represents the MFDFA analysis: solid lines concern the FR group, whereas broken lines concern the MR group. Colours relate to the different scale ranges. All correlation coefficients of the FR group computed between pre- and post-operative scaling factors are not significant, whereas, for positive q, the coefficients of the MR group are significant.

Figure 6.10 shows the distribution of the correlation coefficients for the four scale ranges, highlighting significant differences between the FR and MR cohorts for each of them (p-values < 0.001). Effect size was calculated with Cohen's d and it is depicted in Figure 6.10.

On the contrary, Figure 6.11 represents the MFDFA results that investigated whether specific variations exist or a general pattern can be found in the MR group: in this case, solid lines refer to the pairwise correlation coefficients and broken lines to the coefficients obtained after randomising patients order, whereas colours still relate to the scale ranges. The coefficients for the randomised correlation analysis are not significant.

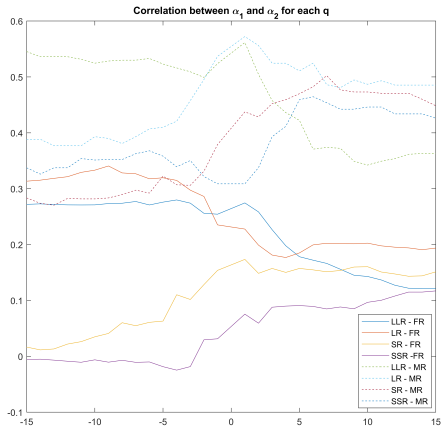


Figure 6.9: Complexity Matching between pre- and post-operative audio recordings in FR (solid) and MR (broken) patient groups. LLR = longest range $2^7 - 2^{13}$, LR = long range $2^8 - 2^{13}$, SR = short range $2^9 - 2^{13}$, SSR = shortest range $2^{10} - 2^{13}$.

The outcome of the multiscale nonlinear analysis produced some unexpected results, e.g., the LLE for the third scale was significantly higher in the post-operative condition than in the pre-operative one. Therefore, it was decided to carry out two additional statistical analyses accounting for the degree of recovery.
Figure 6.12 displays the trends of the means of nonlinear measures against the varying scales, along with standard error bars for pre- and post-operative conditions; furthermore, on the left, the FR group is represented, whereas the right side shows the results of the MR group. Effect size was calculated with Cohen's d and it is depicted in Figure 6.12.

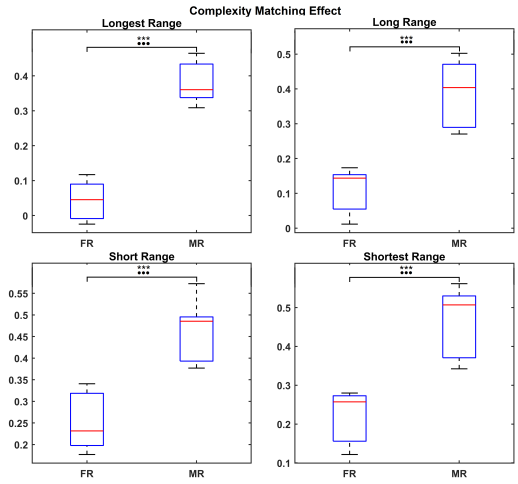


Figure 6.10: Averaged correlation coefficient for each scale range between FR and MR groups. A (***) refers to a significant difference < 0.001. A (●●●) represents a Cohen's d large effect size.

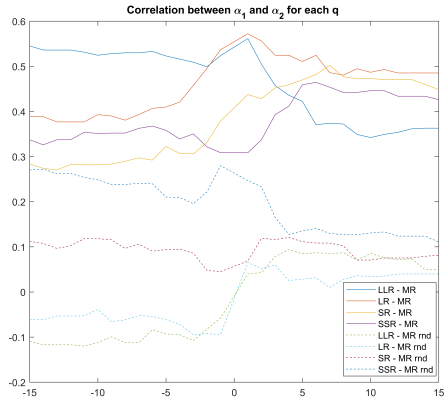


Figure 6.11: Complexity Matching between pre- and post-operative audio recordings in MR patients (solid) and randomised order MR patients (broken). LLR = longest range $2^7 - 2^{13}$, LR = long range $2^8 - 2^{13}$, SR = short range $2^9 - 2^{13}$, SSR = shortest range $2^{10} - 2^{13}$. The abbreviation "rnd" stands for "randomised".

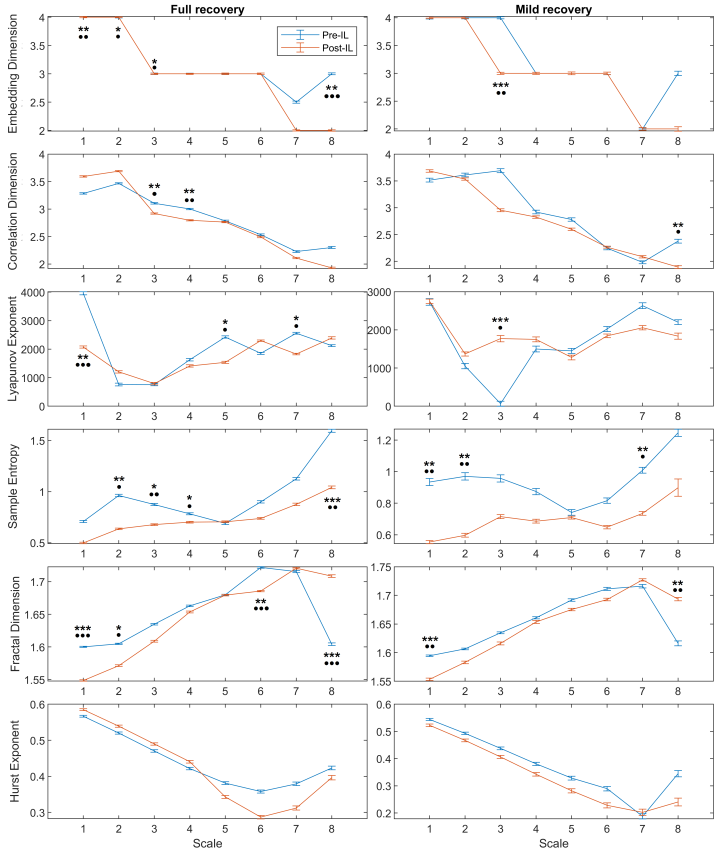


Figure 6.12: Trends with standard error bars of nonlinear parameters for FR patients (left) and MR patients (right). Blue and orange lines refer to pre- and post-operative conditions, respectively. A (•) refers to a significant difference between before and after treatment, computed separately for FR and MR groups, < 0.05, (••) to $p < 0.01$, and (• • •) to $p < 0.001$. A (*) represents a Cohen's d small effect size, (**) a medium one, (***) a large one.

6.3.5 Discussion

AFIL is a minimally invasive and regenerative surgical procedure used to reduce the glottic gap in several voice pathologies, including UVFP. Treatment efficacy is typically assessed with perceptual evaluation, which may suffer from high inter-rater variability, and linear parameters (e.g., F0, jitter), which might not be appropriate for studying the complex, nonlinear dynamics of disordered vocal folds' vibration.
This study proposed a multivariate multiscale nonlinear approach that for the first time also implemented the concept of complexity matching in voice analysis. Moreover, to account for the maladaptive behaviour that UVFP patients adopt to achieve better communication, a multiscale analysis was performed to evaluate the effect of IL on articulation dynamics. Indeed, hyperfunctional characteristics such as anterior-posterior or lateral compression of the glottis, general tightening of neck muscles, hyperadduction of the healthy contralateral vocal fold are observed [260–262]. By progressively filtering out specific frequency components from audio recordings, it is possible to highlight better the contribution of vocal tract resonance phenomena to voice timbre, thus obtaining relevant information that might reflect the coordination and arrangement of the articulators. Therefore, providing a wider insight into phonatory and articulatory capabilities with nonlinear features could help otolaryngologists evaluate and monitor treatment efficiency with a cost-effective and contactless automatic tool.

MFDFA demonstrated that UVFP patients who presented a mild perceptual voice quality recovery had the strongest correlation between scaling coefficients of the pre- and post-conditions, from the shortest to the longest ranges, and for each q-order statistical moment. However, as Figure 6.9 shows, the highest correlation values were below 0.5 and, excluding the ones concerning positive q-order LR to SSR scale ranges of the MR group, not significant, as opposed to the results of [255]. This could suggest that a complexity-matching effect never occurred. Nevertheless, these weaker relationships could be caused by the specific experimental protocol, which did not investigate simultaneously interacting physiological systems but the same system at two different time points. Therefore, what Figure 6.9 displays could be interpreted as the fact that nonlinear phenomena of voice production, which a multifractal approach is better suited to describe, presented the strongest similarity when, perceptually, vocal quality was not

fully restored. Interestingly, this could mean that a larger degree of information (e.g., frequency components, noise, irregular vibratory patterns) from the pre-operative condition is still present after surgery, indicating an agreement between the auditory and objective evaluations [237]. Moreover, the boxplots shown in Figure 6.10 highlighted that, for each scale range, the correlation coefficients between the scaling exponents were significantly higher for the MR group than the FR one. Hence, at all levels of voice production, patients with poorer surgical outcome presented similar vocal alterations to those assessed in the pre-treatment condition. Then, the complexity similarity obtained through the MFDFA efficiently distinguished between two populations. This outcome may also help clinicians to detect post-operative patients that need further medical care, e.g., more frequent follow-up visits or additional logopaedic therapy, to complete the rehabilitation process better and guarantee a proper recovery of both vocal and life quality. The results from Figure 6.11 suggested that complex interactions within the multifractal structure of voice were subject-specific. It was first hypothesised that a randomised order to compute the correlation would not affect the coefficient values if the MR group was characterised by an underlying common property of voice nonlinearities. However, the relevant drop in the strength of the relationships recommended planning precision medicine procedures to follow the rehabilitation of UVFP patients individually. Furthermore, such an evidence showed that correlation coefficients between scaling factors of pre- and post-operative conditions are not due to chance, but they reflect a true matching between acoustic signal complexities, hence, an agreement between the perceptual and objective evaluation of voice quality

For the multivariate and multiscale analysis, three significant differences were found in the FR group at the original signal level (scale = 1). Unexpectedly, the embedding dimension proved to be smaller in the pre-operative condition. This could result from greater sensitivity to the richer content of higher frequency harmonics that characterise normophonic voices, in contrast with dysphonic ones where that range is mostly dominated by noise. Although not significant, this outcome could also be supported by the higher correlation dimension. On the other hand, both the LLE and Fractal Dimension had a significantly larger value in the pre-operative condition. The LLE highlighted that IL surgery may decrease the divergence rate of close trajectories, making the attractor more compact. Similarly, the lower Fractal Dimension

suggested that fewer independent variables were required to describe it. Such an ambiguous result might indicate that, according to [263], the D2 and LLE parameters are not adequate for studying dysphonic voices, or at least the ones characterised by glottal insufficiency. Moreover, Figure 6.12 shows that pre- and post-operative trends intersected frequently, making these parameters seemingly less robust to evaluate voice quality improvements as their behaviour changed repeatedly across the different scales.
Sample entropy presented instead a coherent pattern, with the pre-operative condition values being constantly larger than the post-operative ones. At scale 2, where the habitual speaking range stands after removing higher frequency components (above 11.025 kHz), a more physiological periodicity of the acoustic signal seemed to be restored. Since a conversational sustained phonation of /a/ does not contain much more information beyond the threshold of the third formant (usually below 4kHz [213]), SE was found out to be significantly higher in the pre-operative condition also at scale three and, additionally, scale four, where the contribution of the third formant is filtered out. Unsurprisingly, this can indicate that UVFP, and its potential inappropriate articulatory adjustments, did not affect lip rounding capabilities. SE at scales 5 and 6 seemed to reveal that articulatory differences were mostly due to tongue movements and not the pharyngeal tract, as the contribution related to its frequency components did not lead to a significant difference between pre- and post-treatment. Interestingly, endoscopy evaluation demonstrated that after IL, patients presented a more relaxed pharynx musculature. This may suggest that the consequent change in pharyngeal degree of constriction was not sufficient to affect the first formant, but, since the pharyngeal wall is tightly linked with the tongue base, such relaxation may be reflected in its motility. This result may advise otolaryngologists to consider such an aspect when evaluating treatment efficiency to ensure that compensatory behaviours are progressively deserted. This can guarantee UVFP patients better vocal and life quality, as such behaviours are also known to determine brain adaptation mechanisms, e.g., alterations to the processing and integration of sensory and sensorimotor activity [264,265]. Non-invasive diagnostic tools, e.g., ultrasonography, may highlight additional aspects clinicians could use to better understand the post-treatment rehabilitation process and plan more personalised care. At scales 7 and 8, where the vibration of vocal folds is more evident, pre-operative SE remained larger than the post-operative ones, even if significance was achieved for the last

scale only. Such an outcome underlined beneficial effects on the vocal quality of the IL procedure at both phonatory and articulatory levels. Notably, SE appeared to be a reliable nonlinear parameter for studying also the MR group, as the trend remained lower for each scale in the post-operative condition. Moreover, significant differences were obtained at scales 1, 2, and 7, pointing to the improvements in vocal quality even in the mild recovery case, as both noise was reduced and vocal folds' vibration was regularised. Since breathy vocal quality is not solely associable with UVFP, this multivariate and multiscale strategy may be applied to other voice pathologies, encouraging its broad usability.

6.4 Conclusion

The pre/post classification problem achieved a clinically meaningful performance. The results from the female dataset highlighted the importance of combining acoustic parameters derived from diverse tasks to enhance sensitivity. A higher TPR for the pre-treatment class indicated the model's conservative tendency to err on the side of caution-classifying questionable samples as pre-treatment, possibly reducing misdiagnosis. Explainability analysis suggested that all relevant features, but NNE, were F0-related, emphasising the central role of vocal fold vibration in assessing voice recovery. Moreover, misclassifications analysis revealed that UVFP was the most frequently misclassified condition, in accordance with perceptual evaluations.
In the male dataset, the best performance was achieved by a model trained solely on /aiuole/. The most informative features were F0 and NNE, confirming the importance of vocal fold dynamics and noise levels in characterising treatment effects. UVFP still accounted for the majority of classification errors.
Predictor importance advised that feature selection should be task-specific, and clinical applications should avoid using the same acoustic parameters interchangeably across diagnostic and monitoring tasks.

This chapter also introduced a novel nonlinear multivariate and multiscale approach to assess the effects of AFIL in UVFP. As conventional assessment methods often fall short in capturing the complex, nonlinear behaviour of disordered phonation, the current study proposed the integration of complexity matching to characterise voice dynamics more comprehensively.

MFDFA revealed that MR patients exhibited the strongest correlations between pre- and post-operative conditions. This outcome suggested a retained complexity in vocal behaviour that aligned with perceptual moderate recovery. Moreover, the multifractal structure of voice seemed to be patient-, and not disorder-, specific, advocating for personalised rehabilitation protocols. The significantly higher correlation coefficients in the MR group compared to the FR one across all scale ranges further supported this interpretation. Consequently, complexity similarity may serve as a sensitive marker to identify patients requiring additional therapeutic attention.

The multivariate and multiscale analysis yielded further insights. Sample entropy emerged as a reliable and consistent nonlinear parameter. SE was higher in pre-operative recordings across articulatory and phonatory scales in both groups. This trend indicated a reduction in signal irregularity following AFIL, supporting the notion of treatment-induced stabilisation. Specifically, scales associated with lip and tongue movement may indicate that UVFP-related compensatory behaviours might be progressively abandoned. These observations suggested its potential use as a clinical indicator across different UVFP severity levels and even broader voice pathology contexts.

Notably, both studies illustrate two approaches that may better identify patients that need additional care. This outcome could lead to an improved handling of healthcare resources, potentially avoiding later intensive and environmentally unsustainable interventions and actions.

Chapter 7

Climate Change and Health: the Role of Smartphones

Smartphones are sensor-rich devices that enable mobile health solutions, from appointment reminders to data collection and analysis. Such a strategy can support the healthcare system by making its services more efficient, better organised and also more eco-friendly. This chapter explores the possibility to use smartphones' microphones to carry out a reliable acoustic analysis. Hence, the feasibility of this approach was initially tested on a cohort presenting marked alterations to the voice production apparatus, i.e., patients diagnosed with genetic syndromes. It was found out that voices from Cri du chat and Crisponi syndromes present several alterations in both phonatory and articulatory skills. Moreover, an AI-based tool was developed to distinguish four additional genetic disorders and healthy controls. Despite data limitedness, the models achieved promising accuracies fostering their future use in clinical practice as a robust screening system to better guide and select following examinations. [1]

[1]The individual analyses on Cri du chat and Crisponi syndromes have been published as "From phenotype to phonotype: a comprehensive description of voice features of Cri du chat syndrome" in *European Journal of Pediatrics, v. 184, n. 1, p. 60, 2025* [266] and as "Do rare genetic conditions exhibit a specific phonotype? A comprehensive description of the vocal traits associated with Crisponi/Cold-Induced Sweating Syndrome type 1" in *Genes, v. 16, n. 8, p. 881, 2025* [267]. The machine learning experiment has been partially published in "Artificial intelligence procedure for the screening of genetic

7.1 Introduction

A smartphone is defined as a electronic device that combines the functionality of a traditional mobile phone with the capabilities of a personal computer. It typically features a touchscreen interface to interact with applications, access the internet, and handle multimedia objects. Smartphones have grown in popularity over the past decade and, in 2025, their global penetration was estimated to be around 72.6%. This indicates a high level of their adoption worldwide, particularly in developed countries, and a continued growth trend in emerging markets [268]. Technological advancement allowed smartphones to be integrated with several sensors such as high-resolution cameras, global positioning system (GPS), accelerometers, gyroscopes, magnetometers, ambient light sensors, and microphones. Beyond offering communication, entertainment and activity monitoring, these sensor can be used to measure health parameters on demand and in real time, enabling so-called mobile health (mHealth) solutions [269, 270]. mHealth is part of a wider framework known as Ambient Assisted Living, a concept that was developed to improve the quality of life of people with specials needs, aiming to provide them independence and self-sufficiency. The ubiquity, easy-of-use and relative cost-effectiveness of smartphones made them an affordable strategy to improve health services' accessibility, especially for those with a significant shortage of healthcare resources [271]. Furthermore, such an approach also aligns with emerging sustainability goals in the healthcare system. Indeed, the latter is among the most resource-intensive and carbon-emitting service industries, accounting for a substantial share of global greenhouse gas emissions due to hospital infrastructure management, diagnostic imaging, laboratory testing, and patient and clinician mobility [14, 15]. Thus, DHTs held promise to reduce the need for in-person visits, limit unnecessary diagnostic procedures, and promote early screening and remote monitoring [30]. This clear potential has triggered a wave of interest and investment in health mobile apps to collect and analyse HRV, respiratory rate, medical images of skin and eyes, and voice. Regarding the latter, clinical use of smartphones in acoustic voice and speech assessment protocols has been growing among phoniatricians and otolaryngologists. Research has already focused on the comparability and clinical feasibility of smartphone-derived acoustic measures, considering both healthy and pathological speakers. The high correlations that were achieved

syndromes based on voice characteristics" in *Bioengineering, v. 10, n. 12, p. 1375, 2023* [76].

across recording devices (e.g., high- vs. low-cost smartphones and with respect to professional microphones) suggested their adequacy for a reliable recording and digitising of disordered voices [272].
Within this framework of accessible and environmentally sustainable digital healthcare, this thesis proposed the combination of smartphones and voice analysis as a screening tool in a cohort that presents also relevant anatomo-physiological alterations in the phonatory apparatus, i.e., rare genetic syndromes. This approach could assist clinicians better selecting the next diagnostic exams and therapeutic solutions, thus mitigating healthcare load and carbon footprint.
Up to now, the non-invasive semiotics of such diseases is generally carried out by evaluating somatic traits only. Therefore, their phenotyping could benefit from a more detailed description encompassing acoustic parameters that are unique to individual pathological conditions.

This study evaluated the practicability of a screening procedure, also based on AI-based techniques, that implemented smartphone recordings to identify the so-called *phonotype* of patients diagnosed with genetic syndromes. The objective is two-fold:

- To study whether the acoustic properties of patients diagnosed with genetic syndromes statistically differ from age- and gender-matched healthy population. Specifically, the Cri du chat syndrome was included due to its prominent acoustic characteristics. Then, patients with Crisponi syndrome were also involved to understand whether a phonotype exists even when peculiar vocal features are lacking.

- To study if and how these same parameters are capable of differentiating the genetic syndromes with each other. Particularly, the voice features of patients with Down, Noonan, Costello, and Smith-Magenis syndromes were used in a machine learning experiment to develop a potential automatic tool that may assist clinicians in early diagnosis.

7.2 Acoustic properties of genetic syndromes

Genetic disorders occur in the presence of one or more abnormalities in the genome. They can be caused by a mutation in a single gene (monogenic), multiple genes (polygenic) or by a chromosome abnormality (e.g., duplicated

or missing chromosomal material). Mutations can occur spontaneously before embryonic development, they can be inherited or even induced after a chronic exposure to certain substances (e.g., tobacco smoke, chemical compounds, radiations). The OMIM Gene Map Statistics accounts for more than 7500 different existing genetic conditions [273]. Notably, even if such diseases are rare themselves, the total population prevalence of rare syndromes is estimated being between 3.5-5.9%, which equates to 263-446 million persons affected globally [274]. Genetic syndromes typically imply some form of health-related issues, that also affect the structure and functioning of the voice production system [76]. However, information of vocal properties of these subjects are scarce and fragmented, thus, there's the need to develop a framework to obtain a clear, objective overview.

7.2.1 Cri du chat syndrome

Cri du chat syndrome (CdCS; OMIM #123450), also known as 5p-syndrome, is a neurodevelopmental disorder caused by deletions of variable size in the short arm of chromosome 5 with an incidence ranging from 1÷15000 to 1÷50000 live births [266]. The name of this condition refers to the newborns' high-pitched, feeble, cat-like cry, which is supposedly determined by larynx anomalies (i.e., small, narrow and diamond-shaped) and nervous central system defects [266, 275, 276]. Other potential voice-related findings include large nasal bridge, micrognathia, hypotonia and moderate to severe language delay. Usually, acoustic analysis focuses on cries, but few works conducted investigations on other type of utterances, e.g. babbling in children between 11-26 months of age and the longitudinal case report concerning the vowel production of a young girl (4.6-7 years old) [275, 277]. In both cases, a higher F0 with respect to an age- and gender-matched normophonic population was observed.

7.2.2 Crisponi syndrome

Crisponi/cold-induced sweating syndrome type 1 (CS/CISS1; OMIM #272430) is an autosomal recessive congenital disorder [278]. Although this genetic condition is ultra-rare (with a global incidence estimated at 1÷1000000 alive births), a higher prevalence is recorded in Sardinia (Italy) with an incidence of 1 case every 20700 newborns [279]. CS/CISS1 is caused by variants in the cytokine receptor-like factor 1 (CRLF1) gene, coding for a ligand of the

ciliary neurotrophic factor receptor. Common characteristics include facial muscle and larynx abnormal contractions and orofacial weakness, especially regarding mouth and masseter muscles, limiting mouth opening capabilities [279, 280]. Moreover, the presence of trismus observed in CS/CISS1 is associated with symptoms of motor dysfunction of the palatine veil, causing nasal regurgitation [281]. To date, no acoustic analysis has been carried out to evaluate the objective vocal properties of these patients.

7.2.3 Down syndrome

Down syndrome (DS; OMIM #190685) results from a microscopically demonstrable chromosomal aberration that consists in a triplicate state (trisomy) of all or a critical portion of chromosome 21. Its relatively high prevalence (estimated at 1÷700-800 live births) allowed to gain a clearer overview of DS phenotypical characteristics. For instance, MRI scans highlighted frontal cortical alterations in paediatric age [282], whereas laryngological findings include laryngomalacia, enlarged arytenoid cartilages, palatal anomalies, pharyngeal constriction and limited movement range of articulators due to hypotonia. Voice from patients with DS is typically low in pitch, rough and breathy. Objectively, studies underlined higher F0, larger frequency and amplitude perturbations of the F0 itself, a higher degree of vocal noise, overlapped and altered formant frequencies, and reduced articulatory space with respect to normophonic subjects [283].

7.2.4 Noonan syndrome

Noonan Syndrome (NS; OMIM #163950) is an autosomal dominant multisystem disorder characterised by a relatively high birth prevalence between 1÷1000-1÷2500. NS belongs to the so-called RASopathies, a term used for diseases caused by abnormalities in the RAS/MAPK3 signaling pathway. RAS are proteins that change their state between “on” and “off” during signal transduction. When activated, they trigger a series of phosphorylations that results in the activation of MAPK (Mitogen-Activated Protein Kinase), which is in turn associated with cell proliferation and growth [284]. Typical facial characteristics include deep filtrum, ogival palate, thicker tongue, and short or webbed neck [285]. Moreover, central nervous system alterations have been observed, such as cerebellar ectopia, temporal lobe malformation, and hydrocephalia [284], also leading to language delays and speaking dif-

ficulties. Indeed, dysarthria, i.e., the difficulty in sound articulation, represents a common acoustic sign of NS, up to 72% of total cases [286], associated with lip, tongue and jaw limited movements due to general hypotonia [285]. Despite the higher prevalence, only few studies have addressed acoustic properties of NS. An analysis of the sustained vowel /a/ from 11 children did not produce any statistical different with a normophonic group, except for the Soft Phonation Index [287]. In another study, co-articulation was evaluated with the number listing task, highlighting that NS have a tendency to higher speech rates than control subjects, which may be an indicator of lower intelligibility [192].

7.2.5 Costello syndrome

Costello syndrome (CS; OMIM #218040) is a rare multiple congenital syndrome (estimated prevalence 1÷300000), belonging to the RASopathies. Airway anomalies are known to occur in CS. These include choanal atresia, obstructive sleep apnoea due to adenotonsillar hypertrophy, tongue base obstruction and macroglossia, pharyngeal wall collapse, laryngomalacia, and subglottic stenosis [288]. Moreover, CS subjects are at risk to present macrocephalia, hydrocephalia, and cerebellar ectopia [289]. CS voice has been described as perceptually low in both tonality and intensity, with also hoarse or whispery quality [290]. Dysarthria has been also reported, which was associated with delayed oral-motor skills and language development [290]. To date, no acoustic analysis has been carried out to evaluate the objective vocal properties of these patients.

7.2.6 Smith-Magenis syndrome

Smith-Magenis syndrome (SMS; OMIM #182290) is a genetic disorder with low prevalence (1÷15000-25000 alive births). It is caused an interstitial microdeletion of the short arm of chromosome 17 (region 17p11.2). SMS characterises for several nervous system anomalies (e.g., posterior fossa deformation, ventriculomegalia, structural defects of frontal lobes [291]), an altered neuromotor profile (generalised hypotonia, oral-motor dysfunction [292]) and voice production system abnormalities (such as velopharyngeal insufficiency, tendency to nodule and polyps formation, as well as to incur in vocal fold paralysis [293]). These characteristics may determine a deep, hoarse voice quality in SMS patients that led research to delve deeper in objective acous-

tic analysis. High F0, excess vocal muscle stiffness, disfluencies, tachylalia, numerous syllabic simplifications, phoneme omissions, and high unintelligibility were discovered [292].

7.3 Materials and Methods

7.3.1 Patients

The current study recruited a total of 97 patients diagnosed with CdCS, CS/CISS1, DS, NS, CS and SMS. They were prospectively recruited after confirmed molecular diagnosis at the Department of Life Sciences and Public Health, Fondazione Policlinico Agostino Gemelli-IRCCS, Rome, Italy. Exclusion criterium was the concomitant presence of an acute inflammatory pathology of the upper respiratory tract. Taking into account gender- and age-related influence of vocal properties, the sample was divided into three categories: paediatric subjects (PS, age $\leq$ 12 years), adult females (AF) and adult males (AM). Table 7.1 displays their distribution, along with gender and age information.

Table 7.1: Mean and standard deviation of age of genetic syndrome patients divided per gender and age. In parentheses, the number of subject per group is reported.

Group	CdCS	CS/CISS1	DS	NS
PS	8.8 ± 2.4 (5)	8.3 ± 2.0 (6)	7.2 ± 3.6 (6)	10.7 ± 2.3 (5)
AF	24.4 ± 8.6 (12)	24.2 ± 11.6 (6)	21.2 ± 12.7 (4)	22.4 ± 7.7 (6)
AM	31.0 ± 9.3 (12)		18.3 ± 2.2 (3)	23.7 ± 8.4 (6)

Group	CS	SMS
PS	9.9 ± 2.0 (3)	8.0 ± 2.0 (8)
AF	16.4 ± 4.3 (5)	17.5 ± 1.3 (5)
AM	29.5 ± 2.1 (2)	16.3 ± 1.5 (3)

A Huawei 10 Mate smartphone was used for the recordings. Acquisitions were performed in empty rooms where only the experimenter and the patient were present (in the case of minors, a parent or a tutor was also present, and they were required to stay silent). According to [76], all items of the SIFEL protocol were recorded, but only those concerning the three sustained cardi-

nal vowels (/a/, /i/, and /u/) and the number listing task were considered. These were chosen because they were the most numerous, as they represent a relatively easy task, even for patients with severe cognitive or behavioural impairments. The acquisitions were used for both perceptual and objective voice assessments.
The recordings used in the AI experiments, associated with DS, NS, CS and SMS, were acquired longitudinally. Therefore, multiple audio files were collected for the same patients. Although such measures are not strictly independent, recordings with at least a 1-year interval between acquisitions were selected to find a compromise between reaching an adequate number of data to perform robust classification experiments and obtaining independent-like measures. This approach was necessary, as some of the considered diseases are characterised by extremely low prevalence (7.2).
The study included a healthy control (HS) group of 45 non-smoking subjects. In the 3-12 years subgroup, a total of n = 20 children were included (mean age 8.5 ± 2.3 years), n=10 adult females (mean age 28.2 ± 5.7 years), and n=15 adult males (mean age 28 ± 5.9 years).

7.3.2 Feature extraction

The feature extraction pipeline for the sustained vowels is analogous to the ones presented in Sections 5.2.2 and 6.2.2. Additionally, five derived parameters were included to evaluate articulation capabilities [76].
The vowel space area VSA (Equation 7.1) measures the vowel triangle area, thus quantifying the articulatory ability.

$$VSA = \frac{\mid (F1i * (F2a - F2u) + F1a * (F2u - F2i) + F1u * (F2i - F2a) \mid}{2} \tag{7.1}$$

The formant centralization ratio (FCR) represents a normalisation procedure to obtain an acoustical parameter that maximises dysarthria detection and minimises inter-variability. It is expressed as in Equation 7.2:

$$FCR = \frac{F2u + F2a + F1i + F1u}{F2i + F1a} \tag{7.2}$$

Formant ratios are other important parameters that evaluate tongue movements.

$$Ratio_{F1a/F1i} = \frac{F1a}{F1i} \tag{7.3}$$

$$Ratio_{F1a/F1u} = \frac{F1a}{F1u} \tag{7.4}$$

$$Ratio_{F2i/F2u} = \frac{F2i}{F2u} \tag{7.5}$$

Particularly, Equations 7.3 and 7.4 are sensitive to vertical tongue movements, and Equation 7.5 to horizontal ones.

The number listing task, which recording was processed as a whole without manual segmentation, was parametrised with [192, 266]:

- Average time between each number (aTBN), defined as the average time difference in seconds between the offset and onset of two consecutive numbers.
- Standard deviation time between each number (sTBN).
- Voice segment length (VSL), representing the temporal length in seconds of each number.
- Percentage of unvoiced segments (PUVS), which is the percentage of unvoiced segments in the whole recording.
- Task length, i.e., the length of the whole number listing sequence.

Notably, these features were extracted for the CdCS population only.

7.3.3 Statistical analysis & Machine learning

Data normality was assessed with a Shapiro-Wilk test ($\alpha = 0.05$). In the comparison between acoustic properties of patients diagnosed with genetic syndromes (specifically, CdCS and CS/CISS1) and normophonic subjects, its results guided the choice of either performing a t-test or a Mann-Whitney U test, as well as the computation of the effect sizes. Specifically, for normal data, the Cohen's d was used, whereas for not normally-distributed data the Cliff's Δ was applied. A preliminary correlation analysis was used to remove highly-related features, deleting the ones with $\rho \geq 0.75$.

For the AI-based experiment, three classifiers were developed for the PS, FA, and MA groups individually. Specifically, KNN, SVM, and Random Forest (RF) models were implemented. The training procedure is analogous to the one presented in 5.2.4. Moreover, a custom code was developed to guarantee

classifier generalisation capabilities. Specifically, it assigned the same indexes to all available recordings associated with a single participant so that, during the data split, they were strictly included either in the training or the validation set. This approach allowed classifiers to recognise participants' pathology rather than their identity, reducing possible data leakage.

7.4 Results

The phonation task in the CdCS population highlighted several significant differences in each group. Table 7.2 reports the related p-values and size effects computed either with Cohen's d or Cliff's Δ.

Table 7.2: Significant differences in acoustic features, divided per gender and age. A ($\bullet$) refers to significance values, whereas a (*) to effect size.

Group	**Acoustic Features**
PS	F0 mean /a/$^{\bullet\bullet\bullet}_{***}$, F3 std /a/$^{\bullet}_{**}$, F2 std /u/$^{\bullet}_{**}$, Ratio$_{F1aF1u}$ $^{\bullet}_{***}$
AF	F1 mean /a/$^{\bullet}_{***}$, Signal duration /a/$^{\bullet}_{**}$, F0 std /i/$^{\bullet}_{**}$, Jitter /i/$^{\bullet\bullet}_{**}$, F1 std /i/$^{\bullet\bullet\bullet}_{***}$, F2 mean /i/$^{\bullet\bullet\bullet}_{***}$, F3 mean /i/$^{\bullet\bullet}_{***}$, F3 mean /u/$^{\bullet}_{***}$, Signal duration /u/$^{\bullet}_{***}$, Ratio$_{F1aF1u}{}^{\bullet\bullet}_{***}$, VSA$^{\bullet}_{**}$
AM	F0 mean /a/$^{\bullet\bullet\bullet}_{***}$, F0 std /a/$^{\bullet\bullet\bullet}_{***}$, F1 mean /a/$^{\bullet\bullet}_{***}$, F1 std /a/$^{\bullet\bullet\bullet}_{***}$, F2 mean /a/$^{\bullet\bullet\bullet}_{***}$, F2 std /a/$^{\bullet\bullet\bullet}_{***}$, Signal Duration /a/$^{\bullet\bullet\bullet}_{***}$, F0 mean /i/$^{\bullet\bullet\bullet}_{***}$, F0 std /i/$^{\bullet\bullet}_{**}$, F1 mean /i/$^{\bullet\bullet\bullet}_{**}$, F1 std /i/$^{\bullet\bullet\bullet}_{***}$, F2 std /i/$^{\bullet\bullet}_{***}$, F0 std /u/$^{\bullet\bullet\bullet}_{**}$, F1 mean /u/$^{\bullet\bullet\bullet}_{***}$, F1 std /u/$^{\bullet\bullet\bullet}_{***}$, F2 mean /u/$^{\bullet\bullet\bullet}_{**}$, F2 std /u/$^{\bullet\bullet\bullet}_{***}$

The co-articulation task underlined other significant alterations too, only for the PS and AM groups. Specifically, the first presented a higher PUVS with respect to HS ($p = 0.03_{***}$), whereas the latter a higher sTBCM than HS ($p = 0.002_{***}$).

Regarding the CS/CISS1 sample, the results of the statistical analysis are reported in Table 7.3.

On the other hand, Table 7.4 shows the performance of the ML classifiers for the PS, FA, and MA groups. It displays the mean value of the evaluation

Table 7.3: Significant differences in acoustic features, divided per gender and age. A (•) refers to significance values, whereas a (*) to effect size.

Group	**Acoustic Features**
PS	F1 mean /i/$^{\bullet}_{***}$, F0 std /u/$^{\bullet}_{**}$
AF	F3 mean /i/$^{\bullet}_{**}$

metric across the 10 cross-validation folders and the standard deviation (std).

Table 7.4: Performance of the best classifiers for PS, FA and MA groups. Mean value ± std

PS					
Parameter	**SMS**	**NS**	**CS**	**DS**	**HS**
Precision	88±19%	88±23%	100±0%	93±14%	100±0%
Recall	100±0%	80±42%	50±53%	100±0%	100±0%
Specificity	92±14%	96±10%	100±0%	95±11%	100±0%
F1-score	93±12%	92±15%	100±24%	96±8%	100±0%)
AUC	99±1%	97±3%	85±4%	81±10%	99±0%
Validation Accuracy	87±9%				
FA					
Parameter	**SMS**	**NS**	**CS**	**DS**	**HS**
Precision	60±33%	88±34%	100±0%	100±0%	100±0%
Recall	90±32%	70±48%	60±52%	90±32%	100±0%
Specificity	70±23%	100±11%	100±0%	100±0%	100±0%
F1-score	77±19%	100±0%	100±0%	100±0%	100±0%
AUC	95±5%	96±2%	79±9%	93±8%	100±0%
Validation Accuracy	77±19%				
MA					
Parameter	**SMS**	**NS**	**CS**	**DS**	**HS**
Precision	78±36%	89±33%	94±17%	100±0%	100±0%
Recall	89±33%	89±33%	90±32%	80±42%	80±30%
Specificity	88±19%	97±11%	95±16%	100±0%	100±0%
F1-score	92±15%	100±0%	96±11%	100±0%	87±17%
AUC	93±5%	97±2%	100±0%	100±0%	94±5%
Validation Accuracy	84±17%				

Figure 7.1 illustrates ROC plots for the PS, FA, and MA best classifiers, respectively.

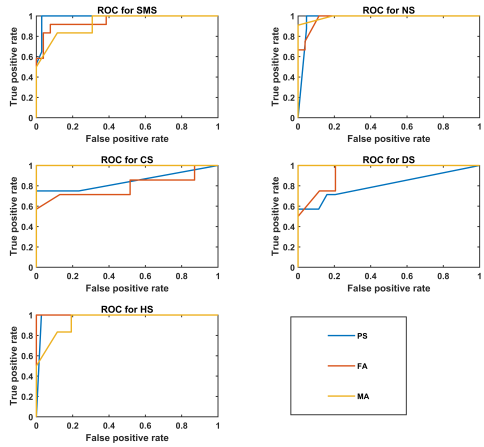


Figure 7.1: ROC plots for the best classifier of PS, FA, and MA groups.

7.5 Discussion

Genetics is responsible for the physiology and for shaping the structures involved in speech production. Therefore, voice analysis might serve as a phenotypic marker. Indeed, several genetic syndromes exhibit distinct voice anomalies that may assist in their early detection and categorisation. Alterations can result from factors like vocal cord function, respiratory rate, neuromuscular control, or hormonal influences.

In the study of CdCS vocal properties, children presented a significantly higher F0 mean /a/, agreeing the perceptual evaluation of high pitched voices even after infancy. This outcome was also confirmed for the AM group, but not for the AF one. Perturbation and noise measures (e.g., F0 std, jitter) were also significant between pathological and healthy population, indicating a possible alignment with reported higher scores in roughness and breathiness indices of the GRB scale [266]. This instability could be determined by neurological causes. Concerning articulatory capabilities, the first formant was significantly altered only in the adult groups, possibly showing that pharynx control and positioning can be difficult for these subjects. Furthermore, significant differences were observed for F2 mean, especially for more

complex sounds such as /i/ and /u/. Finally, F3 mean values for the AF group were significantly lower in the pathological condition. These results suggested that neurological and possibly morphological anomalies concerning tongue and lips motility were present in female CdCS patients, contributing to produce an abnormal voice. These results were supported by a reduced VSA area, which implied less articulatory capacity (and, as a consequence, less intelligibility). This finding matched the perceptual judgements of this samples, as in some cases it was noted that phonation of /i/ and /u/ were practically identical [266]. Finally, $Ratio_{F1aF1u}$ showed a significant difference between CdCS and healthy controls, especially in the PS and AF populations. Indeed, it seems that they experienced greater difficulties in controlling and positioning the tongue in the oral cavity to articulate /u/ rather than /a/. This was not unexpected, as it can be associated with the detected alteration in F3 mean values due to their relationships with the lip rounding skills that were necessary to utter it [213].
The co-articulation task underlined some relevant results to further outline the phonotype of the CdCS. For instance, the AM group presented a significantly higher sTBN with respect to control subject. Numbers as "quattro" (IPA transcription: /'kwattro/), "sette" (IPA transcription: /'sɛtte/), and "otto" (IPA transcription /'ɔtto/) contain complex utterance patterns, such as the double "tt" sound, which may require longer times to re-arrange the articulators to utter the subsequent number. Additionally, in the PS group, larger PUVS were found. This could be the result of biomechanical difficulties, but considering the age of the subjects, longer silence could be also associated with cognitive impairments [192]. Indeed, in severe CdCS cases, learning numbers and uttering them in the correct sequence may be challenging. As a matter of fact, a qualitative analysis highlighted that two subjects skipped a number, and another one uttered the numbers in random order [266]. On the other hand, when mild cases are considered, longer times could be caused by a delayed memory recalling: e.g., three subjects were helped by assistants or parents in cases of longer pauses.

The acoustical analysis showed that the phonatory activity of Crisponi patients did not significantly deviate from that of healthy controls, agreeing with the perceptual evaluation. Indeed, the F0 mean for each corner vowel in the PS and AF groups were comparable with those of age- and gender-matched healthy populations. Concerning articulation, F1 mean /i/ was

significantly altered, suggesting that the frequent paroxysmal tightening of larynx muscles may have a direct detrimental effect on the pharyngeal cavity resonance, especially for more complex cardinal vowels. A difficulty in lowering the jaw might have not allowed the tongue to execute the appropriate vertical movements to reach the correct position to articulate vowels /i/ and /u/. For the AF group, another significant difference was discovered: the standard deviation of F3 mean /i/, suggesting that articulation could also be altered at the uppermost part of the vocal tract, i.e., the lips, due to abnormal facial muscle contractions [213, 267, 279]).

Altogether, these findings showed that smartphones can be reliably used to collect and analyse voice samples. Due to their ubiquity and ease-of-use, these devices could be implemented in domestic recording campaigns that may benefit both patients (e.g., limiting their transportation and interaction with unknown people) and the healthcare system.

In the machine learning experiment, Table 7.4 shows an interesting result: HS is always correctly identified in PS and FA groups. As supported by statistical analysis [76], the voice quality of normophonic and pathological subjects differed, allowing an almost complete separation between these two macro-classes.
Particularly, the KNN classifier of the PS group achieved the highest mean accuracy of 87%. Such a result was expected due to the larger size of the paediatric dataset. The CS class showed a high precision (100%). However, the low recall value (50%), along with its high standard deviation, suggested that vocal properties might not be specific solely to this syndrome. Indeed, a hoarse voice quality has been recognised as a typical property also of other conditions (e.g., SMS or DS 7.2). SMS and NS present a more stable outcome, especially with high specificity scores. However, NS was characterized by a variable recall value (80±42%), which may mean that the NS vocal phenotype might not be easy to define. Since acoustic analyses have shown no peculiar difference between NS and normophonic children in phonation production, this task might be not adequate. On the other hand, since dysarthria is well documented for NS, this suggests that a refined phonotype could emerge considering other vocal emissions, e.g. the number listing one. The SVM model of the FA group performed well on the CS and DS classes. Therefore, these two syndromes seemed to present specific voice characteris-

tics that avoided the other pathologies to be classified as NS and CS due to their 100% specificity value. Oppositely, the SMS class showed poor performance, as some NS, CS, and DS observations were classified as SMS, possibly because it was the most numerous class. In any case, the high SMS recall value suggested that most true observations were correctly identified.
Considerations similar to those of the FA group can be made in the male cohort. High performance also characterised CS and DS recognition. Notably, the HS were not all correctly identified for the MA group, and some observations were misclassified as DS. This could mean that DS presented some vocal similarities with the normophonic group and, thus, alterations might be more difficult to perceive with just a sustained phonation task. The accuracy is similar to the PS group, 84%, but this result must be taken cautiously as the MA cohort was the smallest in this study.
Overall, this output highlighted that the AI-based tools may support the differential diagnosis of genetic syndromes with acoustic features only. Such an automatic approach could be used to better identify following exams and therapies, limiting the arrangement of unnecessary interventions that could delay proper patients care and misuse resources.

7.6 Conclusion

Voice analysis holds significant potential as a non-invasive tool for identifying, characterising, and monitoring syndromic disorders with a genetic origin. Given that speech production is inherently influenced by physiological and anatomical structures, in turn shaped by genetics, voice features can serve as robust phenotypic markers. The preliminary studies performed with smartphone based recordings on CdCS and CS/CISS1 syndromes revealed that individuals diagnosed with the first exhibit distinct vocal anomalies, such as elevated F0 and increased instability in phonation, particularly in male participants. Altered formant patterns, reduced vowel space area, and limited articulatory range were also observed, all of which suggested underlying neuromuscular and morphological impairments that compromised speech intelligibility. The co-articulation task further enriched the phonotypic characterisation of CdCS, capturing dynamic behaviours not evident in sustained phonation tasks. Higher voice-to-unvoiced ratios and prolonged silent segments among paediatric and adult males suggested diminished motor planning and coordination, possibly exacerbated by cognitive impair-

ments. These findings emphasised the importance of task selection in clinical voice assessment for capturing the subtle motor deficits in this population. In contrast, individuals with Crisponi syndrome showed overall comparable phonatory activity to healthy controls, yet formant ratio alterations and reduced vowel space area suggested difficulties in tongue and jaw control, particularly in paediatric subjects. These observations were consistent with CS/CISS1 known clinical symptoms.

The application of machine learning demonstrated that acoustic features can effectively discriminate between pathological and normophonic voice, despite data limitedness. Classifiers such as KNN and SVM achieved high accuracy in distinguishing healthy controls from syndromic populations, particularly in the paediatric cohort. However, variable performance across syndromes, especially in identifying NS patients, suggested that the phonation task only may not fully capture syndrome-specific vocal traits. This strengthened the idea that a more comprehensive vocal profile may be necessary to enhance screening accuracy.
The proposal of a novel framework for vocal phenotyping holds a potential clinical utility for personalised care, improved quality of life, better management of healthcare resources to reduce both its high burden and carbon footprint.

Chapter 8

Conclusion

This doctoral research firstly explored the potential of immersive technologies and psychophysiological paradigms to investigate and influence climate change perception, bridging the explicit and implicit dimensions of environmental attitudes. Two separate studies employed a combination of climate change-related stimulation and physiological monitoring with wearable devices to gain a nuanced insight into how individuals cognitively and emotionally respond to this unfolding crisis, and how these responses can inform and shape effective communication strategies.

The VR experience demonstrated that immersive, emotionally charged simulations can foster significant attitudinal shifts, particularly when tailored to users' cultural and geographical contexts. Participants exposed to a wildfire scenario set in a familiar landscape determined increased concern, awareness, and perceived agency in addressing climate change. These effects were more pronounced among younger participants, males, and those with prior connection to the setting, suggesting that demographic details and context are critical factors in the design of impactful VR experiences. Physiological responses, such as elevated electrodermal activity and heart rate during key narrative moments, corroborated self-reported data regarding presence and immersion, underscoring the emotional salience of the simulation. Moreover, behavioural indicators, such as in-VR navigation patterns, provided further objective evidence of engagement and perceived realism, adding an important layer to traditional questionnaire-based methodologies.
Complementing these findings, the investigation into implicit attitudes using

oculometric data during a Single Category Implicit Association Test revealed a stronger cognitive load when participants processed conflicting environmental information. Eye-tracking measures, particularly gaze recurrence, pupil dynamics, and blink patterns proved sensitive to moments of internal evaluative dissonance. Moreover, these indicators reflected attentional instability and efforts, particularly during incongruent trials, suggesting that physiological markers can underline latent mental mechanisms not captured by reaction time metrics alone.

This research also addressed the intersection of climate change consequences and mitigation strategies by investigating how digital health technologies, particularly based on voice analysis, can improve healthcare delivery with a resource-efficient and environmentally sustainable approach. Considering the relevant contribution of healthcare systems to global carbon emissions, this work aimed to develop and validate non-invasive, data-driven solutions that may reduce diagnostic reliance on energy-intensive medical imaging and in-person procedures, while maintaining or enhancing clinical efficacy.
Through the application of machine learning models solely trained with acoustic features, this thesis demonstrated that voice recordings can reliably differentiate between common laryngeal pathologies such as benign lesions of the vocal folds and unilateral vocal fold paralysis. Importance predictor analyses confirmed the clinical potential of acoustic parameters, especially fundamental frequency and its perturbations, as effective diagnostic markers. Gender-specific patterns in feature relevance highlighted how anatomical differences influence vocal signal characteristics also in disordered voice, reinforcing the need for tailored strategies.
The thesis further explored how voice features can monitor treatment efficacy. Using pre- and post-treatment recordings, machine learning models were able to track vocal recovery following clinical intervention with a high degree of accuracy. Notably, features sensitive to therapeutic change differed from those most effective for diagnosis, emphasising the task-specific nature of vocal markers. This distinction is critical for designing intelligent clinical systems that adapt their analytic frameworks based on the intended purpose. Additionally, explainability analyses and perceptual assessments provided converging evidence that vocal fold dynamics were central to detecting recovery, while the UVFP cohort remained challenging to monitor due to its complex and nonlinear phonatory behaviour.

To address this latter limitation, the current project introduced a novel nonlinear multivariate and multiscale approach for analysing treatment outcomes in UVFP. This methodology incorporated MFDFA and SE to describe the underlying, nonlinear properties in voice dynamics that conventional tools often overlook. On the one hand, findings revealed that complexity matching could reliably distinguish between patients with full and marginal recovery. On the other hands, SE could represent a powerful vocal marker to separate pre- and post-treatment recordings, also identifying specific articulatory deficits in order to guarantee patients a full recovery. These results might advocate for the integration of nonlinear voice metrics into personalised rehabilitation protocols, enhancing clinical decision-making while aligning with sustainable healthcare principles.
Finally, the thesis extended voice-based analysis to the phenotyping of syndromic disorders with a genetic basis. Studies on Cri du chat and Crisponi syndromes revealed distinct and measurable acoustic deviations, particularly in formant structure, vowel space area, and phonation stability. Machine learning classifiers demonstrated that such features can reliably distinguish between healthy and pathological voices, especially in paediatric populations. While some syndromes exhibited clear vocal markers, others proved more elusive, suggesting the need for richer, context-sensitive vocal tasks. This outcome could lay the foundation for a novel framework of vocal phenotyping, which may enable earlier, non-invasive identification of rare genetic conditions and guide more personalised and sustainable care trajectories.

8.1 Summary of contribution

The major contribution of this doctoral project can be summarised as:

- Theoretical Contribution. It advanced the understanding of the value-action gap by demonstrating how immersive media and embodied experiences can realign attitudes, especially when affective and spatial dimensions are leveraged effectively.

- Practical Implications. It provided actionable design guidelines for VR climate interventions, emphasising the need for cultural proximity, sensory realism, and demographic inclusivity to maximise impact.

- Methodological Innovation. It introduced oculometric parameters, such as gaze entropy and pupil diameter, as viable tools to detect implicit

climate attitudes, offering a promising alternative to traditional methods.

- Interdisciplinary Integration. By combining environmental psychology, media studies, and physiological monitoring as well as analysis, this thesis modelled a multi-layered approach to tackle complex behavioural topics such as climate change.

- Advancing Voice Diagnostics. It demonstrated that features from voice recordings can be stand-alone indicators for differentiating between specific laryngeal pathologies, thereby reducing reliance on resource-intensive diagnostic tools.

- Supporting Treatment Monitoring. It established the feasibility of acoustic markers to evaluate therapeutic outcomes, also revealing that diagnosis and monitoring required distinct analytic strategies.

- Pioneering Nonlinear Voice Characterisation. It introduced the concept of complexity matching in voice analysis to better capture post-treatment dynamics in hard-to-classify conditions like UVFP.

- Promoting Vocal Phenotyping. It provided preliminary evidence for the use of voice as a phenotypic marker in syndromic disorders, particularly in paediatric care, with implications for screening and tailored interventions.

- Aligning Clinical Innovation with Sustainability, It showed that intelligent, voice-based systems can offer scalable, low-carbon alternatives to traditional diagnostic pathways, contributing to the healthcare's decarbonisation agenda.

8.2 Directions for future work

Building on the outcomes of this research, several avenues for future investigation and practical implementation emerged. For the VR-based climate change simulation, expanding the sample size and including a more diverse participant pool, particularly international users, would enhance the generalisability of findings and uncover potential cultural moderators of engagement. Moreover, it could better prove the relevance of affective attachment

and spatiotemporal closeness of the virtual setting. The integration of neurophysiological measures such as EEG could provide deeper insights into the cognitive processes underlying attitudinal shifts and emotional involvement, complementing the existing physiological and behavioural data. The VR experience could also benefit from enriching the sensory landscape with olfactory and haptic stimuli to increase realism and immersion. Additionally, a compelling future feature could be represented by a gamified restoration phase, where users could interactively reverse the environmental damage caused by the simulated wildfire. This may potentially foster a stronger sense of agency and reinforcing pro-environmental behaviours. Finally, an active biofeedback system could be deployed: through wearable sensors and AI techniques it would be possible to change and adapt the scenario considering the participants' physiological responses and, with a real-time signal processing framework, to realise a dynamic, engaging environment.
Similarly, future work on the SC-IAT paradigm should aim to recruit a larger and more stratified sample to better examine subgroup differences and refine the sensitivity of implicit attitude detection. Incorporating EEG headset in the experimental paradigm would provide a more fine-grained understanding of the temporal dynamics involved in evaluative conflict, particularly during incongruent trials. To enhance emotional engagement and improve sensitivity, visual stimuli, such as images or pictograms, could replace or complement textual cues, potentially triggering stronger and more intuitive responses. In addition, the current findings underscore the potential for developing a physiologically derived metric, drawing on oculometric and autonomic indicators (e.g., pupil dilation, blink rate, gaze patterns), to support or even augment the traditional D-score. Such a composite index could offer a more nuanced and objective measure of implicit attitudes, broadening the applicability of SC-IATs in environmental psychology and beyond.
In the domain of voice analysis, several extensions are envisioned to maximise both scientific and societal impact. Expanding the range of pathologies under investigation and the inclusion of nonlinear acoustic metrics into classification models could enhance diagnostic precision and reveal subtle pathological signatures overlooked by traditional features. The successful application of complexity matching and multiscale entropy to UVFP should also be replicated across other treatment contexts to validate these tools as universal markers of vocal recovery in breathy dysphonia. Most importantly, the development of an accessible digital platform or mobile applications, de-

signed for both clinicians and the general population, could translate these innovations into useful tools, enabling self-monitoring, early screening, and personalised rehabilitation while promoting sustainable, remote healthcare models.

Appendix A

This appendix reports the questions of the custom-made quiz that aimed at evaluating the level of attention and assimilation of the information given during the simulated radio interview phase of the VR experience.

- Climate change alters the frequency and intensity of heatwaves only.
- The increase in ocean temperatures influences the frequency and severity of wildfires.
- Wildfires can, in turn, worsen the global consequences of climate change.
- Psychological distress caused by wildfires can have long-term effects.
- Home hardening only involves structural modifications, such as using fire-resistant materials for walls and roofs.
- Forest thinning can help slow the spread of wildfires, but it is only effective when performed deep inside dense forests.
- Forest thinning also involves the removal of undergrowth, including small bushes and lower branches.
- Campfires are a minor cause of wildfires and do not require close supervision.

Appendix B

B.1 Guided Tour Scripts

The script of the first stop welcomed participants into the virtual landscape and explained how they need to move to activate the recordings of the tour. In English, it read:
Welcome to São Leonardo de Galafura! I'm your guide and I will help you exploring this panoramic viewpoint. Turn to your right and head toward the green ball to begin the tour!
In Portuguese:
Bem-vindo a São Leonardo de Galafura! Eu sou o seu guia e irei acompanhar a sua visita. Vire à direita e posicione-se junto à bola verde para começar o tour!

The script of the second stop gave geographical information about the Douro River, how it was used to transport Port wine toward the city of Porto. The english version read:
The Douro River originates in Spain, in the Urbión Mountains, and stretches for about 897 km before flowing into the Atlantic Ocean in the city of Porto. Throughout the centuries, its waters served as an important transport route, especially for Port wine. To accomplish this, the iconic Rabelo boats were used, designed to withstand the river's strong currents. Measuring between 19 and 23 meters in length, each boat could carry up to 100 barrels of wine, making its way to Vila Nova de Gaia, where the wine would age and develop its characteristic flavor. Today, Rabelos no longer transport wine, but they still grace the Douro, serving as a tourist attraction. Despite its beauty, the Douro has always been a challenging river to navigate, with strong currents

making the journey risky. One of the most intriguing stories dates back to 1861, when Baron Forrester, a well-known figure in the wine world, fell into the water during a shipwreck. It is said that while the ladies accompanying him were carried by the current and survived, the baron, weighed down by his hefty belt full of gold and money, quickly sank. A tragic story? Perhaps. But also a good reminder of the burden of material wealth!
Now, let's take a turn and head for the opposite site of the viewpoint.
In Portuguese:
O rio Douro nasce em Espanha, na serra de Urbión, e percorre cerca de 897 km até desaguar no oceano Atlântico, na cidade do Porto. Ao longo dos séculos, as suas águas foram uma importante via de transporte, especialmente para o vinho do Porto. Utilizavam-se os icónicos barcos Rabelos, especialmente projetados para enfrentar as fortes correntes do rio. Com um comprimento entre 19 e 23 metros, cada embarcação podia carregar até 100 barris de vinho do Porto, que seguiam viagem até Vila Nova de Gaia, onde envelheciam e ganhavam o seu sabor característico. Hoje em dia, os barcos Rabelo já não têm esta atividade, mas continuam a embelezar o rio Douro, servindo de atração turística. Apesar da sua beleza, o Douro sempre foi um rio de navegação desafiadora, com correntes fortes que tornavam a viagem arriscada. Um dos episódios mais curiosos remonta a 1861, quando o barão de Forrester, uma figura icónica neste contexto, caiu água durante um naufrágio. Diz-se que, enquanto as senhoras que o acompanhavam foram arrastadas pela corrente e acabaram por sobreviver, o barão, com o seu volumoso cinto cheio de ouro e dinheiro, afundou-se rapidamente. Uma história trágica? Talvez. Mas também uma boa lição sobre o peso dos bens materiais! Agora, viremos e sigamos para o lado oposto do miradouro.

The third stop explained the traditional production methods of Port wine and its history. The English text read:
Right now, we are surrounded by the beautiful vineyards that continue to uphold this incredible tradition. And in the distance, you can spot Galafura, a small village that still preserves its historic wine cellars. Port wine is much more than just a drink - it is a true symbol of the region's identity! Produced in the Alto Douro Wine Region, the world's oldest demarcated wine region, established in 1756 by the Marquis of Pombal, this wine stands out for its unmistakable flavor and the traditional methods that are still preserved today. During the harvest season, grapes are carefully hand-picked and then

trodden in stone tanks, a practice that not only keeps the tradition alive but also ensures the unique quality of the wine. But Port wine hasn't always had an easy journey! Its international fame led to an increase in counterfeits, with imitations being sold at much lower prices. This situation eventually caused the Lavoura crisis in the 19th century, severely affecting the region's winemakers. At the time, the Douro was seen as a land of hardship, a reputation that lingered for years. Fortunately, time brought change, and today, Port wine shines brighter than ever, celebrated worldwide as one of Portugal's greatest treasures! Well... as we proceed to the next stop of this guided tour toward the entrance of the chapel, let's admire the enchanting flora surrounding this incredible viewpoint.

In Portuguese:

Neste momento, estamos rodeados pelas vinhas que continuam a manter viva a bonita tradição da vindima. Ao longe, podem avistar Galafura, uma pequena aldeia que ainda preserva as suas históricas caves de vinho do Porto. Este vinho é, sem dúvida, muito mais do que uma simples bebida - é um verdadeiro símbolo da identidade da região! O vinho do Porto, produzido no Alto Douro Vinhateiro, a mais antiga região do mundo, demarcada em 1756 pelo Marquâs de Pombal, distingue-se pelo seu sabor inconfundível e pelo método tradicional que ainda hoje se mantém: a vindima. Neste processo, as uvas são cuidadosamente colhidas e depois pisadas de forma artesanal em lagares, um processo que não só preserva a tradição, mas também garante a qualidade única do vinho. Mas nem tudo foi sempre um mar de rosas para o vinho do Porto! A sua fama internacional levou a um aumento de fraudes, com imitações vendidas a preços muito mais baixos. Esta situação acabou por causar a crise da Lavoura, no século XIX, afetando gravemente os produtores da região. Nessa altura, o Douro ficou com a imagem de um lugar empobrecido, um rótulo difícil de apagar. Felizmente, o tempo trouxe mudanças, e hoje o vinho do Porto brilha mais do que nunca, sendo apreciado em todo o mundo como um dos grandes tesouros de Portugal! Bem... enquanto nos dirigimos para o próximo ponto da visita, em direção à entrada da capela, vamos admirar a encantadora flora que envolve este fantástico miradouro.

In the fourth stop, participants became aware about the flora that surrounds the viewpoint, focusing on oaks and pines, as well as the dangers and damages caused by wildfire. The the English text read:

What we are looking here are pine trees, one of the most dominant trees in the Alto Douro Region, thanks to its remarkable ability to adapt to different types of soil and high temperatures. Its wide canopy not only provides shelter from the intense summer sun but also helps keep the ground cooler. Its widespread diffusion is most probably due to its interesting fire-resistant behavior. Indeed, pine trees have a peculiar strategy for dealing with fire: instead of resisting the flames, as oaks do with their thick, resistant bark, pine trees burn quickly but release their seeds into the fertile ground left by the fire, allowing them to regenerate rapidly. For this very reason, the Pinheiro-manso has been essential for the reforestation of several areas after devastating fires. However, their replanting must follow very strict rules as its conical shape and the structure of its leaves and branches facilitate the rapid spread of fire, making it highly flammable. Keep in mind that about half of the burned areas in the 2017 fires in Portugal were composed of pine trees, highlighting the need for better forest management practices.
Ok, let's go back to where we started to conclude this visit of the viewpoint.
In Portuguese:
As árvores que aqui vemos são pinheiros, uma das espécies mais dominantes na região do Alto Douro, graças à sua impressionante capacidade de adaptação a diferentes tipos de solo e temperaturas elevadas. A sua ampla copa não só oferece abrigo contra o intenso sol de verão, como também ajuda a manter o solo mais fresco. A sua vasta difusão deve-se, provavelmente, ao seu interessante comportamento em relação ao fogo. Na verdade, os pinheiros têm uma estratégia peculiar para lidar com as chamas: em vez de resistirem às chamas, como fazem os carvalhos graças à sua casca espessa e resistente, os pinheiros ardem rapidamente, mas libertam as suas sementes no solo fértil deixado pelo incândio, permitindo-lhes regenerar com grande rapidez. Por essa mesma razão, o Pinheiro-manso tem sido essencial para a reflorestação de diversas áreas após incândios devastadores. No entanto, esta espécie apresenta também um lado preocupante: a sua forma cónica e a estrutura das suas folhas e ramos facilitam a propagação rápida do fogo, tornando-o altamente inflamável. Cerca de metade da área ardida nos incândios de 2017 em Portugal era composta por pinheiros, evidenciando a necessidade de melhores práticas de gestão florestal.
Ok, vamos agora regressar ao ponto de partida para concluir esta visita ao miradouro..

The fifth and final stop thanked participants for their attention and conveyed a final message about ecosystem fragility and responsibility toward natures. In English, it followed this script:
Here we are, thank you for your attention. Before leaving, I want to share with you one last thing: this ecosystem is very fragile, and wildfires pose significant threat everyday. It is our responsibility to ensure the safety of the environment, otherwise the places we know, the beautiful memories we develop in there, would not be accessible for the future generations.
In Portuguese:
Agradeço a atenção dispensada! Para terminar, quero referir uma última coisa: este ecossistema é extremamente frágil, e os incândios representam uma ameaça constante. Cabe-nos a todos garantir a segurança deste sítio, pois, sem essa responsabilidade, o que hoje nos deslumbra e marca a nossa memória, poderá não estar acessível para as futuras gerações!

B.2 Radio Interview

B.2.1 The scripts

The simulated radio news occurred between one host and two experts (EXP1 and EXP2, respectively). During the interview, information about how wildfires originates and their deep connection with climate change were unveiled. In addition to the well-known detrimental effects that such a disaster can bring to health and economy, the interview aimed also at sharing relevant information about prevention and mitigation strategies, so that participants could become aware and possibly actuate them in real life.
The English script read:

HOST: Breaking news! A massive wildfire is currently raging in the Alto Douro region, impacting hundreds of residents. In response to this unfolding crisis, we have brought together a panel of experts to provide you with crucial, life-saving information-helping you understand how these fires start and, most importantly, how to stay safe. Stay tuned for the insights coming up next.

HOST: First of all what is a wildfire?

EXP1: Well, wildfires are one of the most destructive natural disasters, with severe impacts on both human health and ecosystem conservation. They ignite and spread due to three key factors: heat, oxygen, and fuel. Climate change has intensified this threat, driving more frequent and extreme heatwaves, altering ocean temperatures, and disrupting wind and rainfall patterns. As a result, higher temperatures, stronger winds, and prolonged droughts are accelerating the occurrence, spread, and severity of wildfires. Despite ongoing environmental research emphasizing the urgent need for action, projections indicate that wildfire incidents could increase by up to 25% by 2030.

EXP2: Yes, and we all know that Portugal has faced particularly devastating wildfires in recent years. The 2017 fires were especially catastrophic, burning 3.8% of the country's total land area and destroying around 170,000 hectares of forest. In 2024, although fewer wildfires occurred, the worsening climate crisis led to more intense and destructive fires. On average, the area of burned land reached approximately 1,500 hectares per fire-double the average recorded in 2017.

HOST: And the consequences are typically catastrophic, right?

EXP1: Of course. In addition to destroying vital natural ecosystems-which in turn worsens climate change-wildfires pose a serious threat to human health, particularly for the elderly. Wildfire smoke contains hazardous gases and fine particles known to increase hospital admissions for cardiovascular and respiratory conditions. Moreover, scientific studies have shown that wildfires can cause severe psychological distress, including post-traumatic stress disorder, especially when homes and loved ones are lost.

HOST: But we're not powerless, are we? Let's go over some essential guidelines-key rules that everyone should know to help minimize fire damage and stay safe.

EXP2: Certainly. For example, when constructing new homes, it's crucial to consider home hardening-a set of measures designed to make walls and roofs more fire-resistant through the use of appropriate materials. This

approach also emphasizes the importance of regularly clearing debris, such as dead leaves from roofs and gutters, and installing dual-pane windows for added protection.

EXP2: Forest thinning is another important measure, particularly in areas where rural and urban environments meet.

HOST: It basically makes trees less dense while getting closer to houses, right?

EXP2: Exactly, it reduces excess vegetation, as smaller bushes and lower branches, to help slow the spread of wildfires and protect nearby communities.

EXP1: Then of course, it is also important to talk about safe behaviors when starting a fire at campsites. This is indeed one of the most common cause of a wildfire. To avoid this, it is important to attend it at every time and be extremely careful when extinguishing it.

HOST: Thank you, everyone. The flames now appear to be under control, thanks to the fire department. Remember these key points, stay vigilant, and always use good judgment. Take care, and goodbye.

The Portuguese script of the radio interview read:

INT: E agora... uma notícia de última hora! Um grande incêndio está a decorrer neste momento na região do Alto Douro, afetando centenas de pessoas que lá vivem. Face a esta situação, convidamos um painel de especialistas para nos ajudar a compreender como estes fenómenos têm origem e, mais importante, como nos devemos proteger.

INT: Antes de mais, o que é um incêndio florestal?

ESP1: Bem, um incêndio florestal é uma catástrofe natural devastadora que tem um forte impacto tanto na saúde humana como na conservação dos ecossistemas. Cada incêndio começa devido a três fatores principais:

calor, oxigénio e combustível. As alterações climáticas têm provocado ondas de calor mais frequentes e extremas, além de alterarem a temperatura dos oceanos, modificando os padrões dos ventos e da precipitação. Como resultado, a combinação de temperaturas mais elevadas, ventos mais fortes e períodos de seca prolongados tem contribuído para o aumento da frequência, da propagação e da intensidade dos incêndios florestais. A investigação ambiental conduzida recentemente tem alertado para a urgência de combater estes desafios. No entanto, infelizmente, prevê-se que este fenómeno continue a crescer, podendo aumentar até 25% em 2030.

ESP2: Sim, é um facto que Portugal tem enfrentado incêndios florestais particularmente dramáticos nos últimos anos, especialmente em 2017, quando, assustadoramente, 3,8% da área total do país ardeu, resultando na perda de cerca de 170.000 hectares de floresta. Em 2024, embora tenham ocorrido menos incêndios, o agravamento da crise climática levou a fogos mais perigosos e violentos, com a área média ardida a rondar os 1.500 hectares - o dobro do valor registado em 2017!

INT: E as consequências são, normalmente, catastróficas, certo?

ESP1: Sem dúvida. Para além da destruição de ecossistemas essenciais, que agrava ainda mais as alterações climáticas, os incêndios florestais representam um grande perigo para a saúde humana, especialmente para pessoas mais frágeis e para os idosos. O fumo dos incêndios contém gases e partículas altamente nocivos, que temos percebido estarem associados ao aumento das admissões hospitalares por problemas cardiovasculares e respiratórios. Além disso, vários estudos têm demonstrado que estas situações podem causar um grande sofrimento psicológico, como ansiedade, depressão e perturbação de stress pós-traumático, sobretudo quando se perdem bens materiais e entes queridos.

INT: Mas acham que a população tem cumprido bem o seu dever? Talvez seja importante abordar alguns pontos essenciais que podem ajudar a minimizar os danos causados pelos incêndios.

ESP2: Claro: por exemplo, ao construir novas habitações, é fundamental considerar o conjunto de intervenções que tornam as paredes e os telhados

mais resistentes ao fogo através da escolha de materiais adequados - um conceito que se chama "reforço da habitação". Além disso, recomenda-se a remoção frequente de detritos, como folhas secas dos telhados e caleiras, bem como a instalação de janelas de vidro duplo.

ESP2: Depois, o desbaste florestal é outra medida importante, especialmente nas zonas de transição entre áreas rurais e urbanas.

INT: Basicamente, essa medida reduz a densidade de mato nas zonas mais próximas das habitações, é assim?

ESP2: Exato. Além disso, remove os ramos mais baixos e arbustos mais pequenos, reduzindo significativamente a propagação das chamas que tende a começar por aí.

ESP1: Além disso, é fundamental falar sobre o correto comportamento ao acender fogueiras, por exemplo, em parques de campismo. Esta é, comprovadamente, uma das causas mais comuns de incêndios florestais. Para evitá-los, é essencial vigiar a fogueira em todos os momentos e ter extremo cuidado na hora de a apagar.

INT: Agradeço muito a vossa intervenção! Recebemos, entretanto, a informação de que as chamas parecem já estar controladas, graças aos esforços dos bombeiros. Ainda assim, alertamos para que não se esqueçam das informações aqui partilhadas pelos especialistas.
Por agora, despedimo-nos: muito obrigada pela atenção e até à próxima.

B.2.2 Images and Graphs

The first intervention of EXP/ESP1 was accompanied by Figure B.1. The left panel displays the three most relevant factors that facilitate the triggering and spreading of a wildfire [294]. The right panel represents a projection of how fire will impact Europe later in this century, highlighting the greater risks of southern Europe countries [295].

The first intervention of EXP/ESP2 is supported by Figure B.2. It displays the amount of burnt land hectares in Portugal in 2024 [296].

The second intervention of EXP/ESP1 was supported by Figure B.3. On the left side, two stacked bar plots display the overall amount of particulate

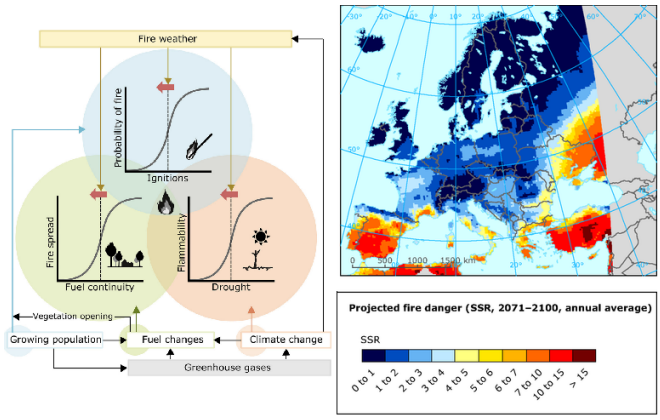


Figure B.1: Left panel: Conceptual model of relationships between fire parameters and their drivers. Right panel: Projected fire danger in late XXI century.

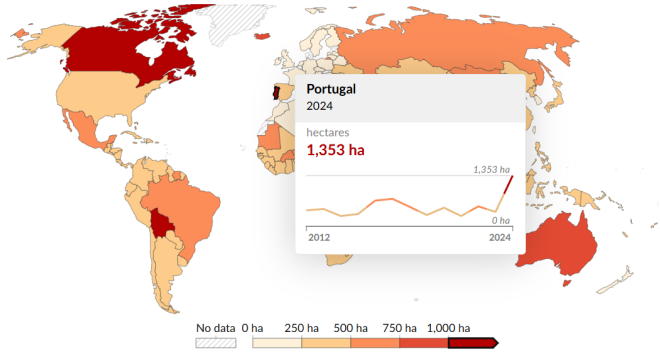


Figure B.2: Burn hectares in Portugal in 2024.

matter (especially the PM2.5) and CO_2 that is released by wildfires in each year [296]. On the right one, a satellite image shows the smoke tracks of the September 2024 northern and central Portugal fires [297].

The second intervention of EXP/ESP2 was accompanied by Figure B.4. It graphically explains which precautions should be taken when building a new house to protect it from the dangers of wildfires (e.g., dual-pane tempered screen, fire resistive materials) [298].

The third intervention of EXP/ESP2 followed the information depicted

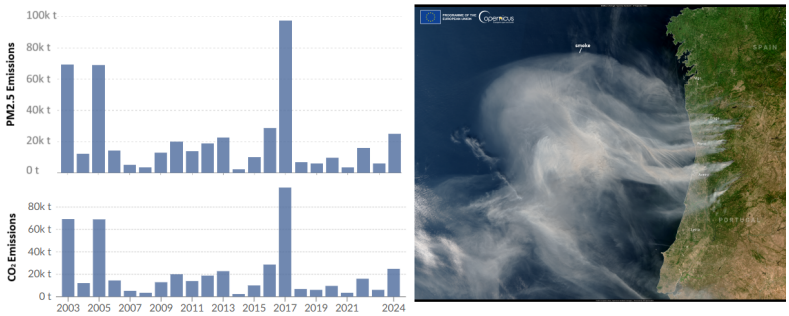


Figure B.3: Left side: Particulate matter and CO_2 emission. Right site: satellite image of 2024 Portugal wildfires.

Figure B.4: Home hardening strategies.

in Figure B.5. It shows how forest thinning should be performed at the urban-rural interface to reduce the possibility of wildfires reaching urban areas [299].

The third and final intervention of EXP/ESP1 was supported by Figure B.6. It displays a few basic principles and rules about the handling of fires at campfire sites [300].

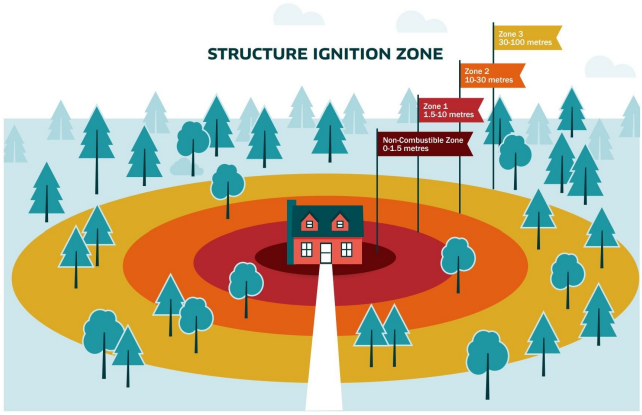


Figure B.5: Forest thinning techniques.

Figure B.6: Campfire management common rules.

Appendix C

This Appendix reports the results from the repeated measure ANOVA, only for the Block Type x Word Category interaction. Table C.1 shows the outcome for the RQA metrics extracted from gaze.

Table C.1: RQA-related eye-tracking metrics statistical analysis results. The acronyms m.d. and GG stand for mean difference and Green-Gaussler correction, respectively.

Parameter	**F**	**p-value**	**T2 vs T4 m.d.**	**T2 vs T4 p-value**	**T2 vs T4 effect size**
RR	913.73	.001	-0.0091	.955	-0.23
DET	2668	.001 (GG)	-0.0039	.999	-0.06
ADL	305.68	.001 (GG)	-0.03	1	-0.01
EDL	355.19	.001 (GG)	-0.30	1	-0.05
LAM	7012	.001 (GG)	-0.055	.997	-0.12
TT	342.03	.001(GG)	-0.26	.984	-0.18
LLVL	513.64	.001 (GG)	-0.67	.984	-0.18
RT1T	591.24	.001	0.039	.999	0.06
RT2T	1349	.001 (GG)	-0.22	.999	-0.09
RPDE	7389	.001	-0.005	.995	-0.14
CC	8781	.001	-0.005	.989	-0.16
TRAN	7535	.001	-0.006	.986	-0.17

Table C.2 displays the results for pupil dilation.

Table C.2: Pupil-related metrics statistical analysis results. The acronyms m.d. and GG stand for mean difference and Green-Gaussler correction, respectively.

Parameter	**F**	**p-value**	**T2 vs T4 m.d.**	**T2 vs T4 p-value**	**T2 vs T4 effect size**
D_{mean}	1126	.001 (GG)	0.069	.0963	0.39
D_{std}	112.94	.001 (GG)	-0.001	1	-0.12
D_{min}	1099	.001 (GG)	0.073	.151	0.42
D_{max}	1148	.001 (GG)	0.069	.193	0.39
$Lat_{D_{min}}$	5739	.001	0.013	1	0.16
$Lat_{D_{max}}$	6405	.001	-0.03	.840	-0.05
V_{dil}	45.88	.001	-1.23	.184	-0.44
$V_{dil_{minmax}}$	48.23	.001 (GG)	-0.66	.870	-0.18
SampEn	241.97	.001	-0.02	.0218	-0.67
IPA	3486	.001	0.013	.747	0.20
LLE	57.54	.001	-0.018	1	-0.05

Bibliography

[1] Linda Steg. Psychology of climate change. *Annual Review of Psychology*, 74(1):391–421, 2023.

[2] Kimberly S Wolske and Paul C Stern. Contributions of psychology to limiting climate change: Opportunities through consumer behavior. In *Psychology and climate change*, pages 127–160. Elsevier, 2018.

[3] Alexa Spence, Wouter Poortinga, and Nick Pidgeon. The psychological distance of climate change. *Risk Analysis: An International Journal*, 32(6):957–972, 2012.

[4] Yuanchao Gong, Shuai Wang, Yang Li, and Yan Sun. Discrepant implicit and explicit attitudes toward climate change: implications for climate change communications. *Sustainability Science*, 18(3):1367–1377, 2023.

[5] Tien Ming Lee, Ezra M Markowitz, Peter D Howe, Chia-Ying Ko, and Anthony A Leiserowitz. Predictors of public climate change awareness and risk perception around the world. *Nature climate change*, 5(11):1014–1020, 2015.

[6] Anthony Leiserowitz. Climate change risk perception and policy preferences: The role of affect, imagery, and values. *Climatic change*, 77(1):45–72, 2006.

[7] James Hansen, Makiko Sato, and Reto Ruedy. Perception of climate change. *Proceedings of the National Academy of Sciences*, 109(37):E2415–E2423, 2012.

[8] Laurie A Rudman, Meghan C McLean, and Martin Bunzl. When truth is personally inconvenient, attitudes change: the impact of extreme weather on implicit support for green politicians and explicit climate-change beliefs. *Psychological science*, 24(11):2290–2296, 2013.

[9] Kelly S Fielding and Matthew J Hornsey. A social identity analysis of climate change and environmental attitudes and behaviors: Insights and opportunities. *Frontiers in psychology*, 7:121, 2016.

[10] Bertram Gawronski and Galen V Bodenhausen. Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological bulletin*, 132(5):692, 2006.

[11] A. G. Greenwald, D. E. McGhee, and J. L. Schwartz. Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6):1464, 1998.

[12] Geoffrey Beattie and Laura McGuire. The modifiability of implicit attitudes to carbon footprint and its implications for carbon choice. *Environment and Behavior*, 52(5):467–494, 2020.

[13] Grace Kelly, Osatohamwen I Idubor, Sophie Binney, Paul J Schramm, Maria C Mirabelli, and Joy Hsu. The impact of climate change on asthma and allergic-immunologic disease. *Current allergy and asthma reports*, 23(8):453–461, 2023.

[14] Kathleen Gray. Climate change, human health, and health informatics: a new view of connected and sustainable digital health. *Frontiers in Digital Health*, 4:869721, 2022.

[15] Carolynn L Smith, Yvonne Zurynski, and Jeffrey Braithwaite. We can't mitigate what we don't monitor: using informatics to measure and improve healthcare systems' climate impact and environmental footprint. *Journal of the American Medical Informatics Association*, 29(12):2168–2173, 2022.

[16] Caroline Hickman, Elizabeth Marks, Panu Pihkala, Susan Clayton, R Eric Lewandowski, Elouise E Mayall, Britt Wray, Catriona Mellor, and Lise Van Susteren. Climate anxiety in children and young people and their beliefs about government responses to climate change: a global survey. *The Lancet Planetary Health*, 5(12):e863–e873, 2021.

[17] Daniel Helldén, Camilla Andersson, Maria Nilsson, Kristie L Ebi, Peter Friberg, and Tobias Alfvén. Climate change and child health: a scoping review and an expanded conceptual framework. *The Lancet Planetary Health*, 5(3):e164–e175, 2021.

[18] Alan P Jacobsen, Yii Chun Khiew, Eamon Duffy, James O'Connell, Evans Brown, Paul G Auwaerter, Roger S Blumenthal, Brian S Schwartz, and John William McEvoy. Climate change and the prevention of cardiovascular disease. *American Journal of Preventive Cardiology*, 12:100391, 2022.

[19] GBD 2015 Chronic Respiratory Disease Collaborators et al. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the global burden of disease study 2015. *The Lancet. Respiratory Medicine*, 5(9):691, 2017.

[20] Fatima Al-Shihabi, Anna Moore, and Tahseen A Chowdhury. Diabetes and climate change. *Diabetic Medicine*, 40(3):e14971, 2023.

[21] Sanjay Rajagopalan, Robert D Brook, Pedro RVO Salerno, Brendan Bourges-Sevenier, Philip Landrigan, Mark J Nieuwenhuijsen, Thomas Munzel, Salil V Deo, and Sadeer Al-Kindi. Air pollution exposure and cardiometabolic risk. *The Lancet Diabetes & Endocrinology*, 12(3):196–208, 2024.

[22] Leen Rasking, Kenneth Vanbrabant, Hannelore Bové, Michelle Plusquin, Katrien De Vusser, Harry A Roels, and Tim S Nawrot. Adverse effects of fine particulate matter on human kidney functioning: a systematic review. *Environmental Health*, 21(1):24, 2022.

[23] Aditi Bunker, Jan Wildenhain, Alina Vandenbergh, Nicholas Henschke, Joacim Rocklöv, Shakoor Hajat, and Rainer Sauerborn. Effects of air temperature on climate-sensitive mortality and morbidity outcomes in the elderly; a systematic review and meta-analysis of epidemiological evidence. *EBioMedicine*, 6:258–268, 2016.

[24] Ollie Jay, Anthony Capon, Peter Berry, Carolyn Broderick, Richard de Dear, George Havenith, Yasushi Honda, R Sari Kovats, Wei Ma, Arunima Malik, et al. Reducing the health effects of hot weather and heat extremes: from personal cooling strategies to green cities. *The Lancet*, 398(10301):709–724, 2021.

[25] Yoram Epstein and Ran Yanovich. Heatstroke. *New England Journal of Medicine*, 380(25):2449–2459, 2019.

[26] Ana Catalina Alvarez-Elias, Barry M Brenner, and Valerie A Luyckx. Climate change and its influence in nephron mass. *Current opinion in nephrology and hypertension*, 33(1):102–109, 2024.

[27] CL Reddington, EW Butt, DA Ridley, P Artaxo, WT Morgan, H Coe, and DV Spracklen. Air quality and human health improvements from reductions in deforestation-related fire in brazil. *Nature Geoscience*, 8(10):768–771, 2015.

[28] Mark Greener. The weather's no joke for people living with diabetes. *Practical Diabetes*, 40(3):39–41, 2023.

[29] Zhenxu Ning, Shuzhen He, Xinghao Liao, Chunguang Ma, and Jing Wu. Cold waves and fine particulate matter in high-altitude chinese cities: assessing their interactive impact on outpatient visits for respiratory disease. *BMC Public Health*, 24(1):1377, 2024.

[30] Zerina Lokmic-Tomkins, Dinesh Bhandari, Chris Bain, Ann Borda, Timothy Charles Kariotis, and David Reser. Lessons learned from natural disasters around digital health technologies and delivering quality healthcare. *International journal of environmental research and public health*, 20(5):4542, 2023.

[31] Christiana Westlin, Jordan E Theriault, Yuta Katsumi, Alfonso Nieto-Castanon, Aaron Kucyi, Sebastian F Ruf, Sarah M Brown, Misha Pavel, Deniz Erdogmus, Dana H Brooks, et al. Improving the study of brain-behavior relationships by revisiting basic assumptions. *Trends in cognitive sciences*, 27(3):246–257, 2023.

[32] Danielle S Bassett and Michael S Gazzaniga. Understanding complexity in the human brain. *Trends in cognitive sciences*, 15(5):200–209, 2011.

[33] Feliberto de la Cruz, Andy Schumann, Stefanie Köhler, Jürgen R Reichenbach, Gerd Wagner, and Karl-Jürgen Bär. The relationship between heart rate and functional connectivity of brain regions involved in autonomic control. *Neuroimage*, 196:318–328, 2019.

[34] Saroj Jayasinghe. Complexity science to conceptualize health and disease: is it relevant to clinical medicine? In *Mayo Clinic Proceedings*, volume 87, page 314, 2012.

[35] Luca Mainardi, Enzo Pasquale Scilingo, and Laura Astolfi, editors. *Unlocking the mind and emotions: cutting-edge bioengineering techniques from computational pshysiology to clinical applications*, chapter Wearable technologies for emotion monitoring: measuring physiological and behavioral signals. Pàtron Editore, 2025.

[36] Jayoung Kim, Alan S Campbell, Berta Esteban-Fernández de Ávila, and Joseph Wang. Wearable biosensors for healthcare monitoring. *Nature biotechnology*, 37(4):389–406, 2019.

[37] Subhas Chandra Mukhopadhyay. Wearable sensors for human activity monitoring: A review. *IEEE sensors journal*, 15(3):1321–1330, 2014.

[38] Tuba Yilmaz, Robert Foster, and Yang Hao. Detecting vital signs with wearable wireless sensors. *Sensors*, 10(12):10837–10862, 2010.

[39] Frank Van Overwalle, Mario Manto, Zaira Cattaneo, Silvia Clausi, Chiara Ferrari, John DE Gabrieli, Xavier Guell, Elien Heleven, Michela Lupo, Qianying Ma, et al. Consensus paper: cerebellum and social cognition. *The Cerebellum*, 19:833–868, 2020.

[40] Kenhub. `https://www.kenhub.com/`. Accessed: 26-12-2025.

[41] Andrea Stocco, Christian Lebiere, and John R Anderson. Conditional routing of information to the cortex: a model of the basal ganglia's role in cognitive coordination. *Psychological review*, 117(2):541, 2010.

[42] John Leonard Wilkinson. *Neuroanatomy for medical students*. Butterworth-Heinemann, 2014.

[43] Stephen Maren. Long-term potentiation in the amygdala: a mechanism for emotional learning and memory. *Trends in neurosciences*, 22(12):561–567, 1999.

[44] Jahangir Moini, Nicholas Avgeropoulos, and Mohtashem Samsam. *Epidemiology of brain and spinal tumors*. Academic Press, 2021.

[45] Vilayanur S Ramachandran. *Encyclopedia of the human brain*. Elsevier, 2002.

[46] Maria Samara, Cristina Farmaki, Nikolaos Zacharioudakis, Matthew Pediaditis, Myrto Krana, and Vangelis Sakkalis. Comparison between dry and wet EEG electrodes in an SSVEP-based BCI for robot navigation. In *2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 333–338. IEEE, 2022.

[47] Sven Leach, Ku-young Chung, Laura Tüshaus, Reto Huber, and Walter Karlen. A protocol for comparing dry and wet EEG electrodes during sleep. *Frontiers in neuroscience*, 14:586, 2020.

[48] Yu Mike Chi, Tzyy-Ping Jung, and Gert Cauwenberghs. Dry-contact and noncontact biopotential electrodes: Methodological review. *IEEE reviews in biomedical engineering*, 3:106–119, 2010.

[49] Cardiac conduction system. `https://my.clevelandclinic.org/health/body/21648-heart-conduction-system`. Accessed: 26-12-2025.

[50] Michael Davis. The role of the amygdala in fear and anxiety. *Annual review of neuroscience*, 15(1):353–375, 1992.

[51] Julian F Thayer and Richard D Lane. A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of affective disorders*, 61(3):201–216, 2000.

[52] Fred Shaffer and Jay P Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, 5:258, 2017.

[53] Julian F Thayer, Fredrik Åhs, Mats Fredrikson, John J Sollers III, and Tor D Wager. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews*, 36(2):747–756, 2012.

[54] Structure and function of eccrine, apocrine and sebaceous glands. `https://plasticsurgerykey.com/structure-and-function-of-eccrine-apocrine-and-sebaceous-glands/`. Accessed: 26-12-2025.

[55] António Banganho, Marcelino Santos, and Hugo Plácido da Silva. Electrodermal activity: Fundamental principles, measurement, and application. *IEEE Potentials*, 41(5):35–43, 2022.

[56] Michael E Dawson, Anne M Schell, Diane L Filion, et al. The electrodermal system. *Handbook of psychophysiology*, 2:200–223, 2007.

[57] AB Sereno and MS Bolding. Executive functions: eye movements and human neurological disorders. 2017.

[58] AG Shaikh. Eye movement disorders. 2017.

[59] Robert G Alexander and Susana Martinez-Conde. Fixational eye movements. *Eye movement research: An introduction to its scientific foundations and applications*, pages 73–115, 2019.

[60] Riccardo Brambilla, Francesco Onorati, Vincenzo Russo, Maurizio Mauri, Lorenzo Magrassi, Luca T Mainardi, and Riccardo Barbieri. A stimulus-response processing framework for pupil dynamics assessment during isoluminant stimuli. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 400–403. IEEE, 2018.

[61] Sebastiaan Mathôt. Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, 1(1), 2018.

[62] Chin-An Wang and Douglas P Munoz. A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Current opinion in neurobiology*, 33:134–140, 2015.

[63] Andy Schumann, Stephanie Kietzer, Juliane Ebel, and Karl Jürgen Bär. Sympathetic and parasympathetic modulation of pupillary unrest. *Frontiers in neuroscience*, 14:178, 2020.

[64] Janet C Rucker. Normal and abnormal lid function. *Handbook of clinical neurology*, 102:403–424, 2011.

[65] Matteo Bologna, Giulia Paparella, Josep Valls-Solé, Mark Hallett, and Alfredo Berardelli. Neural control of blinking. *Clinical Neurophysiology*, 2024.

[66] Antonio Maffei and Alessandro Angrilli. Spontaneous blink rate as an index of attention and emotion during film clips viewing. *Physiology & Behavior*, 204:256–263, 2019.

[67] Alfonso Magliacano, Salvatore Fiorenza, Anna Estraneo, and Luigi Trojano. Eye blink rate increases as a function of cognitive load during an auditory oddball paradigm. *Neuroscience Letters*, 736:135293, 2020.

[68] Otto Lappi. Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference. *Neuroscience & Biobehavioral Reviews*, 69:49–68, 2016.

[69] Gregory Hickok and Steven L Small. *Neurobiology of language*. Academic press, 2015.

[70] Camille Finck and Lionel Lejeune. Structure and oscillatory function of the vocal folds. In *Handbook of Behavioral Neuroscience*, volume 19, pages 427–438. Elsevier, 2010.

[71] Julian Chengjun Chen. *Elements of human voice*. World Scientific, 2016.

[72] Christiane Brems. Understanding the anatomy of breath and breathing. In *Therapeutic Breathwork: Clinical Science and Practice in Healthcare and Yoga*, pages 41–97. Springer, 2024.

[73] Niv Mor, Kristina Simonyan, and Andrew Blitzer. Central voice production and pathophysiology of spasmodic dysphonia. *The Laryngoscope*, 128(1):177–183, 2018.

[74] Jaehong Park, Seonmi Choi, Jun Takatoh, Shengli Zhao, Andrew Harrahill, Bao-Xia Han, and Fan Wang. Brainstem control of vocalization and its coordination with respiration. *Science*, 383(6687):eadi8081, 2024.

[75] Jan G Svec and Svante Granqvist. Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*, 19(4):356–368, 2010.

[76] Federico Calà, Lorenzo Frassineti, Elisabetta Sforza, Roberta Onesimo, Lucia D'Alatri, Claudia Manfredi, Antonio Lanata, and Giuseppe Zampino. Artificial intelligence procedure for the screening of genetic syndromes based on voice characteristics. *Bioengineering*, 10(12):1375, 2023.

[77] Federico Calà, Mariana Magalhães, António Coelho, and Antonio Lanatà. Burning reality: Experiencing climate change through virtual reality. 2025 IEEE 14th Global Conference on Consumer Electronics (GCCE). IEEE, 2025.

[78] Miguel Melo, Hugo Coelho, Guilherme Gonçalves, Nieves Losada, Filipa Jorge, Mário Sérgio Teixeira, and Maximino Bessa. Immersive multisensory virtual reality technologies for virtual tourism: A study of the user's sense of presence, satisfaction, emotions, and attitudes. *Multimedia Systems*, 28(3):1027–1037, 2022.

[79] Yétindranathsingh Dhunnoo, Adrian Carter, Daniel O'Hare, James Birt, and Martin Skitmore. Improving climate change awareness through immersive virtual reality communication: A case study. *Sustainability*, 15(17):12969, 2023.

[80] David M Markowitz and Jeremy N Bailenson. Virtual reality and the psychology of climate change. *Current Opinion in Psychology*, 42:60–65, 2021.

[81] Anna CM Queiroz, Géraldine Fauville, Adina T Abeles, Aaron Levett, and Jeremy N Bailenson. The efficacy of virtual reality in climate change education increases with amount of body movement and message specificity. *Sustainability*, 15(7):5814, 2023.

[82] David M Markowitz, Rob Laha, Brian P Perone, Roy D Pea, and Jeremy N Bailenson. Immersive virtual reality field trips facilitate learning about climate change. *Frontiers in psychology*, 9:2364, 2018.

[83] Marijn HC Meijers, Ragnheiur "Heather" Torfadóttir, Anke Wonneberger, and Ewa Maslowska. Experiencing climate change virtually: The effects of virtual reality on climate change related cognitions, emotions, and behavior. *Environmental Communication*, 17(6):581–601, 2023.

[84] Stefan P Thoma, Matthias Hartmann, Jonas Christen, Boris Mayer, Fred W Mast, and David Weibel. Increasing awareness of climate change with immersive virtual reality. *Frontiers in Virtual Reality*, 4:897034, 2023.

[85] Laura D Cosio, Oğuz'Oz' Buruk, Daniel Fernández Galeote, Isak De Villiers Bosman, and Juho Hamari. Virtual and augmented reality for environmental sustainability: A systematic review. In *Proceedings of the 2023 chi conference on human factors in computing systems*, pages 1–23, 2023.

[86] Tommaso Luzzati, Stefano Baraldi, Sara Ermini, Claudia Faita, Valeria Faralla, Pietro Guarnieri, Luca Lusuardi, Vincenzo Santalucia, Sara Scipioni, Matteo Sirizzotti, et al. Can improving climate change perception lead to more environmentally friendly choices? evidence from an immersive virtual environment experiment. *Ecological Economics*, 229:108475, 2025.

[87] Carolin Wienrich, Stephanie Vogt, Nina Döllinger, and David Obremski. Promoting eco-friendly behaviour through virtual reality-implementation and evaluation of immersive feedback conditions of a virtual co2 calculator. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–9, 2024.

[88] Chloe Hurrell, Andreas Chai, Heather Green, and Graham Bradley. Virtual reality facilitates pro-environmental behavioural intentions. *Environmental Education Research*, 30(10):1856–1883, 2024.

[89] Yuke Pi, Xueni Pan, Mel Slater, and Justyna Świdrak. Embodied time travel in vr: from witnessing climate change to action for prevention. *Frontiers in Virtual Reality*, 5:1499835, 2025.

[90] Zixiang Xu, Yuan Liang, Abraham G Campbell, and Soumyabrata Dev. Climate crisis in virtual environments: Exploration and evaluation of virtual reality and mixed reality for climate change education in sea level rise simulation. In *Proceedings of the Future Technologies Conference*, pages 564–577. Springer, 2023.

[91] Lara Ditrich and Martin Lachmair. A 'front-row seat' to catastrophe: testing the effect of immersive technologies on sympathy and pro-environmental behavior in the context of rising sea levels. *Environmental Education Research*, pages 1–19, 2025.

[92] Imed Nouri, Hajer Zorgati, and Rym Bouzaabia. The impact of immersive 360° video on environmental awareness and attitude toward climate change: the moderating role of cybersickness. *Journal of Social Marketing*, 2025.

[93] Mengqi Liao, Pejman Sajjadi, and S Shyam Sundar. How does VR affect emotional appeal and persuasiveness of gain versus loss-framed messages? *Science Communication*, 46(3):276–304, 2024.

[94] Lisa Rebenitsch and Charles Owen. Review on cybersickness in applications and visual displays. *Virtual reality*, 20:101–125, 2016.

[95] Sun Joo Ahn, Jesse Fox, Katherine R Dale, and J Adam Avant. Framing virtual experiences: Effects on environmental efficacy and behavior over time. *Communication Research*, 42(6):839–863, 2015.

[96] Guanxiong Pei, Qian Shang, Shizhen Hua, Taihao Li, and Jia Jin. EEG-based affective computing in virtual reality with a balancing of the computational efficiency and recognition accuracy. *Computers in Human Behavior*, 152:108085, 2024.

[97] Zhiyuan Yu and Shengwen Guo. A low-cost, wireless, 4-channel EEG measurement system used in virtual reality environments. *HardwareX*, 17:e00507, 2024.

[98] Jin Woo Choi, Haram Kwon, Jaehoon Choi, Netiwit Kaongoen, Chaeeun Hwang, Minuk Kim, Byung Hyung Kim, and Sungho Jo. Neural applications using immersive virtual reality: a review on EEG studies. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1645–1658, 2023.

[99] David Mendez-Encinas, Aaron Sujar, Sofia Bayona, and David Delgado-Gomez. Attention and impulsivity assessment using virtual reality games. *Scientific Reports*, 13(1):13689, 2023.

[100] Mariana Magalhães, António Coelho, Miguel Melo, and Maximino Bessa. Measuring users' emotional responses in multisensory virtual reality: A systematic literature review. *Multimedia Tools and Applications*, 83(14):43377–43417, 2024.

[101] Pia Spangenberger, Jule M Krüger, Sonja M Geiger, Georg Felix Reuth, Lena Baumann, and Steve Nebel. Compassion is key: How virtually embodying nature increases connectedness to nature. *Journal of Environmental Psychology*, 102:102521, 2025.

[102] Elif Erisen, Funda Yildirim, Emirhan Duran, Batuhan Şar, and Irmak Kalkan. Exploring the effectiveness of virtual reality in combating misinformation on climate change. *Political Psychology*, 2024.

[103] Vishal Kiran Kuvar, Jeremy N Bailenson, and Caitlin Mills. A novel quantitative assessment of engagement in virtual reality: Task-unrelated thought is reduced compared to 2D videos. *Computers & Education*, 209:104959, 2024.

[104] R. Christensen and G. Knezek. The climate change attitude survey: Measuring middle school student beliefs and intentions to enact positive environmen-

tal change. *International Journal of Environmental and Science Education*, 10(5):773–788, 2015.

[105] Teresa Pereira, Teresa Freire, and Dionísia Tavares. Portuguese validation of the climate change attitude survey: Psychometric properties and relations with positive youth development. *European Journal of Developmental Psychology*, 20(5):875–888, 2023.

[106] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. The experience of presence: Factor analytic insights. *Presence*, 10(3):266–281, 2001.

[107] Jacinto Vasconcelos-Raposo, Maximino Bessa, Miguel Melo, Luis Barbosa, Rui Rodrigues, Carla Maria Teixeira, Luciana Cabral, and António Augusto Sousa. Adaptation and validation of the Igroup Presence Questionnaire (IPQ) in a portuguese sample. *Presence*, 25(3):191–203, 2016.

[108] Mariana de Oliveira Magalhães. *Unlocking the virtual world: a study on the influence of multisensory stimuli on users' emotional responses and vividness of mental imagery in the context of virtual tourism*. PhD thesis, Universidade do Porto, 2024.

[109] Heeseok Oh and Wookho Son. Cybersickness and its severity arising from virtual reality content: A comprehensive study. *Sensors*, 22(4):1314, 2022.

[110] Elevenlabs. `https://elevenlabs.io/`. Accessed: 25-02-2025.

[111] Akli Benali, Ana CL Sá, João Pinho, Paulo M Fernandes, and José MC Pereira. Understanding the impact of different landscape-level fuel management strategies on wildfire hazard in central portugal. *Forests*, 12(5):522, 2021.

[112] Alberto Greco, Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo, and Luca Citi. cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE transactions on biomedical engineering*, 63(4):797–804, 2015.

[113] Jiapu Pan and Willis J Tompkins. A real-time QRS detection algorithm. *IEEE transactions on biomedical engineering*, (3):230–236, 2007.

[114] Hooman Sedghamiz. Complete Pan Tompkins implementation ECG QRS detector. `https://www.mathworks.com/matlabcentral/fileexchange/45840-complete-pan-tompkins-implementation-ecg-qrs-detector`. Accessed: 25-02-2025.

[115] Andrea Baldini, Sergio Frumento, Danilo Menicucci, Angelo Gemignani, Enzo Pasquale Scilingo, and Alberto Greco. Subjective fear in virtual reality: a linear mixed-effects analysis of skin conductance. *IEEE Transactions on Affective Computing*, 13(4):2047–2057, 2022.

[116] Rita Balocchi, Federico Cantini, Maurizio Varanini, Gianfranco Raimondi, Jacopo M Legramante, and Alberto Macerata. Revisiting the potential of time-domain indexes in short-term hrv analysis. 2006.

[117] Seema Arora-Jonsson. Virtue and vulnerability: Discourses on women, gender and climate change. *Global environmental change*, 21(2):744–751, 2011.

[118] European Investment Bank. Portuguese people among most informed about climate change consequences in european union, eib survey finds.

[119] Mustafa Can Gursesli, Federico Calà, Pietro Tarchi, Lorenzo Frassineti, Andrea Guazzini, and Antonio Lanatà. Eyetracking correlated in the matching pairs game. In *2023 IEEE Gaming, Entertainment, and Media Conference (GEM)*, pages 1–6. IEEE, 2023.

[120] Federico Calà, Mustafa Can Gursesli, Pietro Tarchi, Lorenzo Frassineti, Andrea Guazzini, and Antonio Lanata. Pupil dilation dynamics in serious game emotional activation. In *2024 IEEE Gaming, Entertainment, and Media Conference (GEM)*, pages 1–4. IEEE, 2024.

[121] Federico Calà, Pietro Tarchi, Lorenzo Frassineti, Mustafa Can Gursesli, Andrea Guazzini, and Antonio Lanata. Eye-tracking correlates in the Implicit Association Test. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 400–403. IEEE, 2023.

[122] Mustafa Can Gursesli, Federico Cala, Febri Abdullah, Ruck Thawonmas, Mirko Duradoni, Andrea Guazzini, and Antonio Lanata. Web-agile facial emotion recognition and eye-tracking system (wafer-et). In *2023 IEEE 13th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, pages 147–148. IEEE, 2023.

[123] S. Capstick, L. Whitmarsh, W. Poortinga, N. Pidgeon, and P. Upham. International trends in public perceptions of climate change over the past quarter century. *Wiley Interdisciplinary Reviews: Climate Change*, 6(1):35–61, 2015.

[124] Y. Luo and J. Zhao. Motivated attention in climate change perception and action. *Frontiers in psychology*, 10:1541, 2019.

[125] K. Lee, N. Gjersoe, S. O'Neill, and J. Barnett. Youth perceptions of climate change: A narrative synthesis. *Wiley Interdisciplinary Reviews: Climate Change*, 11(3):e641, 2020.

[126] A. Leiserowitz, C. Roser-Renouf, J. Marlon, and E. Maibach. Global warming's six americas: a review and recommendations for climate change communication. *Current Opinion in Behavioral Sciences*, 42:97–103, 2021.

[127] D. Evensen, L. Whitmarsh, P. Devine-Wright, J. Dickie, P. Bartie, C. Foad, M. Bradshaw, A. Ryder, S. ad Mayer, and A. Varley. Growing importance

of climate change beliefs for attitudes towards gas. *Nature Climate Change*, 13:240–243, 2023.

[128] R. E. Petty, D. T. Wegener, and L. R. Fabrigar. Attitudes and attitude change. *Annual Review of psychology*, 48(1):609–647, 1997.

[129] William R Catton Jr and Riley E Dunlap. A new ecological paradigm for post-exuberant sociology. *American behavioral scientist*, 24(1):15–47, 1980.

[130] Yuko Heath and Robert Gifford. Free-market ideology and environmental degradation: The case of belief in global climate change. *Environment and behavior*, 38(1):48–71, 2006.

[131] W. Hofmann, B. Gawronski, T. Gschwendner, H. Le, and M. Schmitt. A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and social psychology bulletin*, 31(10):1369–1385, 2005.

[132] B. Schiller, L. R. Gianotti, T. Baumgartner, K. Nash, T. Koenig, and D. Knoch. Clocking the social mind by identifying mental processes in the iat with electrical neuroimaging. *Proceedings of the National Academy of Sciences*, 113(10):2786–2791, 2016.

[133] Manon Moreno, Luis Gutiérrez-Rojas, and Alejandro Porras-Segovia. Implicit cognition tests for the assessment of suicide risk: A systematic review. *Current psychiatry reports*, 24(2):141–159, 2022.

[134] Danielle M Young, Laurie A Rudman, Helen M Buettner, and Meghan C McLean. The influence of female role models on women's implicit science cognitions. *Psychology of women quarterly*, 37(3):283–292, 2013.

[135] Aaron Smith-McLallen, Blair T Johnson, John F Dovidio, and Adam R Pearson. Black and white: The role of color bias in implicit race bias. *Social cognition*, 24(1):46–73, 2006.

[136] Claudia Menzel, Julia Brom, and Lea Marie Heidbreder. Explicitly and implicitly measured valence and risk attitudes towards plastic packaging, plastic waste, and microplastic in a German sample. *Sustainable Production and Consumption*, 28:1422–1432, 2021.

[137] Giulia Songa, Hendrik Slabbinck, Iris Vermeir, and Vincenzo Russo. How do implicit/explicit attitudes and emotional reactions to sustainable logo relate? A neurophysiological study. *Food Quality and Preference*, 71:485–496, 2019.

[138] Gerald Steiner, Bernhard Geissler, Günther Schreder, and Lukas Zenk. Living sustainability, or merely pretending? From explicit self-report measures to implicit cognition. *Sustainability Science*, 13:1001–1015, 2018.

[139] Laura McGuire and Geoffrey Beattie. Talking green and acting green are two different things: An experimental investigation of the relationship between

implicit and explicit attitudes and low carbon consumer choice. *Semiotica*, 2019(227):99–125, 2019.

[140] Emily O'Donnell, Shaun Maskrey, Glyn Everett, and Jessica Lamond. Developing the implicit association test to uncover hidden preferences for sustainable drainage systems. *Philosophical Transactions of the Royal Society A*, 378(2168):20190207, 2020.

[141] Melanie C Steffens. Is the implicit association test immune to faking? *Experimental psychology*, 51(3):165–179, 2004.

[142] Jessica Röhner, Ronald R Holden, and Astrid Schütz. IAT faking indices revisited: Aspects of replicability and differential validity. *Behavior Research Methods*, 55(2):670–693, 2023.

[143] Max J McDaniel, Margaret E Beier, Andrew W Perkins, Stephen Goggin, and Brian Frankel. An assessment of the fakeability of self-report and implicit personality measures. *Journal of Research in Personality*, 43(4):682–685, 2009.

[144] Klaus Fiedler, Claude Messner, and Matthias Bluemke. Unresolved problems with the "I", the "A", and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). *European review of social psychology*, 17(1):74–147, 2006.

[145] Christoph M Michel and Micah M Murray. Towards the utilization of EEG as a brain imaging tool. *Neuroimage*, 61(2):371–385, 2012.

[146] Michael X Cohen. Where does EEG come from and what does it mean? *Trends in neurosciences*, 40(4):208–218, 2017.

[147] Yoon Jung Choi and Jang-Han Lee. The effect of virtual covert sensitization on reducing alcohol craving in heavy social drinkers. *Virtual Reality*, 19:111–117, 2015.

[148] JP Caneiro, Peter O'Sullivan, Anne Smith, G Lorimer Moseley, and Ottmar V Lipp. Implicit evaluations and physiological threat responses in people with persistent low back pain and fear of bending. *Scandinavian journal of pain*, 17(1):355–366, 2017.

[149] Sylvia Gabel, Aldin Alijagic, Özün Keskin, and Andreas Gegenfurtner. Teacher gaze and attitudes toward student gender: evidence from eye tracking and implicit association tests. *Social Psychology of Education*, 28(1):72, 2025.

[150] Geoffrey Beattie and Laura McGuire. See no evil? Only implicit attitudes predict unconscious eye movements towards images of climate change. *Semiotica*, 2012(192):315–339, 2012.

[151] Maria Laura Mele, Stefano Federici, and John Lawrence Dennis. Believing is seeing: fixation duration predicts implicit negative attitudes. *PloS One*, 9(8):e105106, 2014.

[152] Andrea Zangrossi, Liisa Camilla Gatto, Virginia Lanfranchi, Cristina Scarpazza, Miriam Celli, and Giuseppe Sartori. Autobiographical implicit association test and eye movements: fixations topography enables detection of autobiographical memories. *Frontiers in Psychology*, 15, 2024.

[153] Tokihiro Ogawa, Natsu Todoriki, and Michiko Tsuneoka. The use of pupillometry in autobiographical implicit association test. *Frontiers in Psychology*, 12:729897, 2021.

[154] A. Karpinski and R. B. Steinman. The single category implicit association test as a measure of implicit social cognition. *J Pers Soc Psychol*, 91(1):16–32, 2006.

[155] Antonio Lanatà, Alberto Greco, Mirco Ciardelli, Allison Uvelli, Elisabetta Fratini, Diego Manzoni, Enzo P Scilingo, Enrica L Santarcangelo, and Laura Sebastiani. Linear and non linear measures of pupil size as a function of hypnotizability. *Scientific Reports*, 11(1):5196, 2021.

[156] Antonio Lanata, Laura Sebastiani, Francesco Di Gruttola, Stefano Di Modica, Enzo Pasquale Scilingo, and Alberto Greco. Nonlinear analysis of eye-tracking information for motor imagery assessments. *Frontiers in Neuroscience*, 13:1431, 2020.

[157] M. E. Kret and E. E. Sjak-Shie. Preprocessing pupil size data: Guidelines and code. *Behavior research methods*, 51:1336–1342, 2019.

[158] H Kenneth Walker, W Dallas Hall, and J Willis Hurst. Clinical methods: the history, physical, and laboratory examinations. 1990.

[159] Yanfang Wang, Sonia S Toor, Ramesh Gautam, and David B Henson. Blink frequency and duration during perimetry and their relationship to test–retest threshold variability. *Investigative ophthalmology & visual science*, 52(7):4546–4550, 2011.

[160] Andrew D Peckham and Sheri L Johnson. Spontaneous eye-blink rate as an index of reward responsivity: validation and links to bipolar disorder. *Clinical Psychological Science*, 4(3):451–463, 2016.

[161] Ali A Abusharha. Changes in blink rate and ocular symptoms during different reading tasks. *Clinical optometry*, pages 133–138, 2017.

[162] Joseph P Zbilut and Charles L Webber Jr. Embeddings and delays as derived from quantification of recurrence plots. *Physics letters A*, 171(3-4):199–203, 1992.

[163] Norbert Marwan, M Carmen Romano, Marco Thiel, and Jürgen Kurths. Recurrence plots for the analysis of complex systems. *Physics reports*, 438(5-6):237–329, 2007.

[164] Thomas Koelewijn, Hilde de Kluiver, Barbara G Shinn-Cunningham, Adriana A Zekveld, and Sophia E Kramer. The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing research*, 323:81–90, 2015.

[165] Hugo Mitre-Hernandez, Roberto Covarrubias Carrillo, Carlos Lara-Alvarez, et al. Pupillary responses for cognitive load measurement to classify difficulty levels in an educational video game: Empirical study. *JMIR Serious Games*, 9(1):e21620, 2021.

[166] Andrew T Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–13, 2018.

[167] Sandra P Marshall. The index of cognitive activity: Measuring cognitive workload. In *Proceedings of the IEEE 7th conference on Human Factors and Power Plants*, pages 7–7. IEEE, 2002.

[168] Francesco Onorati, Luca Tommaso Mainardi, Fabiola Sirca, Vincenzo Russo, and Riccardo Barbieri. Nonlinear analysis of pupillary dynamics. *BMT*, 61(1):95–106, 2016.

[169] Mate Gambiraža, Ivan Kesedžić, Marko Šarlija, Siniša Popović, and Krešimir Ćosić. Classification of cognitive load based on oculometric features. In *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*, pages 377–382. IEEE, 2021.

[170] M. Fiorenza, M. Duradoni, G. Barbagallo, and A. Guazzini. Implicit association test (IAT) toward climate change: A prisma systematic review. *Current Research in Ecological and Social Psychology*, 4:100103, 2023.

[171] Vasileios Skaramagkas, Giorgos Giannakakis, Emmanouil Ktistakis, Dimitris Manousos, Ioannis Karatzanis, Nikolaos S Tachos, Evanthia Tripoliti, Kostas Marias, Dimitrios I Fotiadis, and Manolis Tsiknakis. Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Reviews in Biomedical Engineering*, 16:260–277, 2021.

[172] Lei Fan, Junjie Wang, Qi Li, Zhenhao Song, Jinhui Dong, Fangjun Bao, and Xiaofei Wang. Eye movement characteristics and visual fatigue assessment of virtual reality games with different interaction modes. *Frontiers in Neuroscience*, 17:1173127, 2023.

[173] GV Portnova, KM Liaukovich, LN Vasilieva, and EI Alshanskaia. Autonomic and behavioral indicators on increased cognitive loading in healthy volunteers. *Neuroscience and Behavioral Physiology*, 53(1):92–102, 2023.

[174] Federico Calà, Francesco Correnti, Lorenzo Frassineti, Giovanna Cantarella, Giulia Buccichini, Ludovica Battilocchi, and Antonio Lanatà. Unveiling vocal phenotypes of dysphonia with unsupervised learning. In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOSIGNALS*, pages 860–867. INSTICC, SciTePress, 2025.

[175] Federico Calà, Lorenzo Frassineti, Giovanna Cantarella, Giulia Buccichini, Ludovica Battilocchi, Claudia Manfredi, and Antonio Lanatà. Towards an explainable artificial intelligence system for voice pathology identification and post-treatment characterisation. *Biomedical Signal Processing and Control*, 104:107530, 2025.

[176] Chaudhery Mustansar Hussain and Rustem Kecili. *Modern environmental analysis techniques for pollutants*. Elsevier, 2019.

[177] Robert Thayer Sataloff. *Clinical assessment of voice*. Plural publishing, 2017.

[178] Sue Sherratt. What are the implications of climate change for speech and language therapists? *International Journal of Language & Communication Disorders*, 56(1):215–227, 2021.

[179] Jorge Andrés Gómez-García, Laureano Moro-Velázquez, and Juan Ignacio Godino-Llorente. On the design of automatic voice condition analysis systems. Part I: Review of concepts and an insight to the state of the art. *Biomedical Signal Processing and Control*, 51:181–199, 2019. doi: 10.1016/j.bspc.2018.12.024.

[180] Lorraine Olson Ramig and Katherine Verdolini. Treatment efficacy: voice disorders. *Journal of Speech, Language, and Hearing Research*, 41(1):S101–S116, 1998. doi: 10.1044/jslhr.4101.s101.

[181] Regina Helena Garcia Martins, Henrique Abrantes do Amaral, Elaine Lara Mendes Tavares, Maira Garcia Martins, Tatiana Maria Gonçalves, and Norimar Hernandes Dias. Voice disorders: etiology and diagnosis. *Journal of voice*, 30(6):761.e1–761.e9, 2016. doi: 10.1016/j.jvoice.2015.09.017.

[182] Jonathan Reid, Preet Parmar, Tyler Lund, Daniel K Aalto, and Caroline C Jeffery. Development of a machine-learning based voice disorder screening tool. *American Journal of Otolaryngology*, 43(2):103327, 2022. doi: 10.1044/jslhr.4101.s101.

[183] Elizabeth Erickson Levendoski and M Preeti Sivasankar. Vocal fold ion transport and mucin expression following acrolein exposure. *The Journal of membrane biology*, 247(5):441–450, 2014.

[184] Elizabeth Erickson Levendoski, Ciara Leydon, and Susan L Thibeault. Vocal fold epithelial barrier in health and injury: a research review. *Journal of Speech, Language, and Hearing Research*, 57(5):1679–1691, 2014.

[185] Remigiusz Ziarno, Anna Suska, Wojciech Kulinowski, Aleksandra Grudzień-Ziarno, Magdalena Kostrzon, Magdalena Paciorek, and Jacek Składzień. Does smog affect the incidence of exacerbations of chronic laryngitis? Analysis based on the inhabitants of lesser poland voivodeship. *Polish Journal of Otolaryngology*, 71(3):10–19, 2017.

[186] Nur Ain Nabila Za'im, Fahad Taha Al-Dhief, Mawaddah Azman, Majid Razaq Mohamed Alsemawi, Nurul Mu' azzah Abdul Latiff, and Marina Mat Baki. The accuracy of an online sequential extreme learning machine in detecting voice pathology using the malaysian voice pathology database. *Journal of Otolaryngology-Head & Neck Surgery*, 52(1):62–72, 2023.

[187] Abdul-Latif Hamdan, Christopher Jabbour, Elie Khalifee, Anthony Ghanem, and Aya El Hage. Tolerance of patients using different approaches in laryngeal office-based procedures. *Journal of Voice*, 37(2):263–267, 2021. doi: 10.1016/j.jvoice.2020.12.009.

[188] Zulfiqar Ali, Ghulam Muhammad, and Mohammed F Alhamid. An automatic health monitoring system for patients suffering from voice complications in smart cities. *IEEE Access*, 5:3900–3908, 2017. doi: 10.1109/ACCESS.2017.2680467.

[189] John D Birkmeyer, Amber Barnato, Nancy Birkmeyer, Robert Bessler, and Jonathan Skinner. The impact of the COVID-19 pandemic on hospital admissions in the United States: study examines trends in US hospital admissions during the COVID-19 pandemic. *Health Affairs*, 39(11):2010–2017, 2020. doi: 10.1377/hlthaff.2020.00980.

[190] Hamzeh Ghasemzadeh and Meisam K Arjmandi. Toward optimum quantification of pathology-induced noises: An investigation of information missed by human auditory system. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:519–528, 2019. doi: 10.1109/TASLP.2019.2959222.

[191] Juan Ignacio Godino-Llorente, Rubén Fraile, Nicolás Sáenz-Lechón, Víctor Osma-Ruiz, and P Gómez-Vilda. Automatic detection of voice impairments from text-dependent running speech. *Biomedical Signal Processing and Control*, 4(3):176–182, 2009. 10.1016/j.bspc.2009.01.007.

[192] Lorenzo Frassineti, Federico Calà, Elisabetta Sforza, Roberta Onesimo, Chiara Leoni, Antonio Lanatà, Giuseppe Zampino, and Claudia Manfredi. Quantitative acoustical analysis of genetic syndromes in the number listing task. *Biomedical Signal Processing and Control*, 85:104887, 2023. doi: 10.1016/j.bspc.2023.104887.

[193] Zulfiqar Ali, Mansour Alsulaiman, Irraivan Elamvazuthi, Ghulam Muhammad, Tamer A Mesallam, Mohamed Farahat, and Khalid H Malki. Voice pathology detection based on the modified voice contour and SVM. *Biologically Inspired Cognitive Architectures*, 15:10–18, 2016. 10.1016/j.bica.2015.10.004.

[194] Rui Xu and Donald Wunsch. Survey of clustering algorithms. *IEEE Trans Neural Netw Learn Syst*, 16(3):645–678, 2005.

[195] Amit Saxena, Mukesh Prasad, Akshansh Gupta, Neha Bharill, Om Prakash Patel, Aruna Tiwari, Meng Joo Er, Weiping Ding, and Chin-Teng Lin. A review of clustering techniques and developments. *Neurocomputing*, 267:664–681, 2017.

[196] Iratxe Niño-Adan, Diana Manjarres, Itziar Landa-Torres, and Eva Portillo. Feature weighting methods: A review. *Expert Syst Appl*, 184:115424, 2021.

[197] Sarah K Diehl, Antje S Mefferd, Ya-Chen Lin, Jessie Sellers, Katherine E McDonell, Michael de Riesthal, and Daniel O Claassen. Motor speech patterns in huntington disease. *Neurology*, 93(22):e2042–e2052, 2019.

[198] Jan Rusz, Tereza Tykalova, Michal Novotny, David Zogala, Karel Sonka, Evzen Ruzicka, and Petr Dusek. Defining speech subtypes in de novo parkinson disease: response to long-term levodopa therapy. *Neurology*, 97(21):e2124–e2135, 2021.

[199] Laura Verde, Giuseppe De Pietro, and Giovanna Sannino. Voice disorder identification by using machine learning techniques. *IEEE access*, 6:16246–16255, 2018. doi: 10.1109/ACCESS.2018.2816338.

[200] HyunBum Kim, Juhyeong Jeon, Yeon Jae Han, YoungHoon Joo, Jonghwan Lee, Seungchul Lee, and Sun Im. Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy. *Journal of Clinical Medicine*, 9(11):3415, 2020.

[201] Shih-Hau Fang, Yu Tsao, Min-Jing Hsiao, Ji-Ying Chen, Ying-Hui Lai, Feng-Chuan Lin, and Chi-Te Wang. Detection of pathological voice using cepstrum vectors: A deep learning approach. *Journal of Voice*, 33(5):634–641, 2019.

[202] I Hammami, L Salhi, and S Labidi. Voice pathologies classification and detection using EMD-DWT analysis based on higher order statistic features. *Irbm*, 41(3):161–171, 2020. 10.1016/j.irbm.2019.11.004.

[203] Kevin Degila, Rahhal Errattahi, and Asmaa El Hannani. The UCD system for the 2018 FEMH voice data challenge. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 5242–5246. IEEE, 2018. doi: 10.1109/BigData.2018.8622604.

[204] Ioanna Miliaresi and Aggelos Pikrakis. A modular deep learning architecture for voice pathology classification. *IEEE Access*, pages 80465–80478, 2023. doi: 10.1109/ACCESS.2023.3300795.

[205] Pavol Harar, Zoltan Galaz, Jesus B Alonso-Hernandez, Jiri Mekyska, Radim Burget, and Zdenek Smekal. Towards robust voice pathology detection: Investigation of supervised deep learning, gradient boosting, and anomaly detection approaches across four databases. *Neural Computing and Applications*, 32(20):15747–15757, 2020. doi: 10.1007/s00521-018-3464-7.

[206] Bogdan Woldert-Jokisz. Saarbruecken Voice Database. 2007. [Online].

[207] Maria Raffaella Marchese, Federico Sensoli, Silvia Campagnini, Matteo Cianchetti, Andrea Nacci, Francesco Ursino, Lucia D'Alatri, Jacopo Galli, Maria Chiara Carrozza, Gaetano Paludetti, et al. Artificial intelligence for the recognition of benign lesions of vocal folds from audio recordings. *Acta Otorhinolaryngologica Italica*, 43(5):317–323, 2023. doi: 10.14639/0392-100X-N2309.

[208] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.

[209] Amina Adadi and Mohammed Berrada. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, 6:52138–52160, 2018.

[210] Nabeel Seedat, Vered Aharonson, and Yaniv Hamzany. Automated and interpretable m-health discrimination of vocal cord pathology enabled by machine learning. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pages 1–6. IEEE, 2020.

[211] Wasifur Rahman, Sangwu Lee, Md Saiful Islam, Victor Nikhil Antony, Harshil Ratnu, Mohammad Rafayet Ali, Abdullah Al Mamun, Ellen Wagner, Stella Jensen-Roberts, Emma Waddell, et al. Detecting parkinson disease using a web-based speech task: Observational study. *Journal of medical Internet research*, 23(10):e26305, 2021.

[212] Federico Calà, Lorenzo Frassineti, Claudia Manfredi, Philippe Dejonckere, Federica Messina, Sergio Barbieri, Lorenzo Pignataro, and Giovanna Cantarella. Machine learning assessment of spasmodic dysphonia based on acoustical and perceptual parameters. *Bioengineering*, 10(4):426, 2023.

[213] John R Deller Jr, John H S Hansen, and John G Proakis. *Discrete-time processing of speech signals*. Macmillan Publishing Co, New York, 1993.

[214] Barbara L Herrington-Hall, Linda Lee, Joseph C Stemple, Kim R Niemi, and Molly Miller McHone. Description of laryngeal pathologies by age, sex, and occupation in a treatment-seeking sample. *Journal of Speech and Hearing Disorders*, 53(1):57–64, 1988. doi: 10.1044/jshd.5301.57.

[215] Maria Sole Morelli, Silvia Orlandi, and Claudia Manfredi. Biovoice: A multipurpose tool for voice analysis. *Biomedical Signal Processing and Control*, 64:102302, 2021. doi: 10.1016/j.bspc.2020.102302.

[216] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer New York, New York, 2009.

[217] Davide Chicco and Giuseppe Jurman. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1):1–13, 2020. doi: 10.1186/s12864-019-6413-7.

[218] Kenji Kira and Larry A Rendell. A practical approach to feature selection. In *Machine learning proceedings 1992*, pages 249–256. Elsevier, 1992. doi: 10.1016/B978-1-55860-247-2.50037-1.

[219] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996. doi: 10.1111/j.2517-6161.1996.tb02080.x.

[220] Chris Ding and Hanchuan Peng. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02):185–205, 2005. doi: 10.1142/S0219720005001004.

[221] Lorenzo Frassineti, Claudia Manfredi, Benedetta Olmi, and Antonio Lanatà. A generalized linear model for an ECG-based neonatal seizure detector. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 471–474. IEEE, 2021. doi: 10.1109/EMBC46164.2021.9630841.

[222] Estimates of predictor importance for classification ensemble of decision trees - MATLAB. `https://www.mathworks.com/help/stats/classreg.learning.classif.compactclassificationensemble.predictorimportance.html`.

[223] Daniel M Low, Vishwanatha Rao, Gregory Randolph, Phillip C Song, and Satrajit S Ghosh. Identifying bias in models that detect vocal fold paralysis from audio recordings using explainable machine learning and clinician ratings. *PLOS Digit Health*, 3(5):e0000516, 2024.

[224] Alisa Zhukhovitskaya, Danielle Battaglia, Sid M Khosla, Thomas Murry, and Lucian Sulica. Gender and age in benign vocal fold lesions. *The Laryngoscope*, 125(1):191–196, 2015. doi: 10.1002/lary.24911.

[225] Aline Oliveira Santos, Juliana Godoy, Kelly Silverio, and Alcione Brasolotto. Vocal changes of men and women from different age decades: An analysis from 30 years of age. *Journal of Voice*, 37(6):840–850, 2021. doi: 10.1002/lary.24911.

[226] Sammi Taylor, Christopher Dromey, Shawn L Nissen, Kristine Tanner, Dennis Eggett, and Kim Corbin-Lewis. Age-related changes in speech and voice: spectral and cepstral measures. *Journal of Speech, Language, and Hearing Research*, 63(3):647–660, 2020. doi: 10.1044/2019_JSLHR-19-00028.

[227] Susanne Gahl and R Harald Baayen. Twenty-eight years of vowels: Tracking phonetic variation through young to middle age adulthood. *Journal of Phonetics*, 74:42–54, 2019.

[228] Leonardo A Forero Mendoza, Edson Cataldo, Marley MBR Vellasco, Marco A Silva, and José A Apolinário Jr. Classification of vocal aging using parameters extracted from the glottal signal. *Journal of Voice*, 28(5):532–537, 2014.

[229] Evelyn Alves Spazzapan, Vanessa Moraes Cardoso, Eliana Maria Gradim Fabron, Larissa Cristina Berti, Alcione Ghedini Brasolotto, and Viviane Cristina de Castro Marino. Acoustic characteristics of healthy voices of adults: from young to middle age. In *CoDAS*, volume 30. SciELO Brasil, 2018. doi: 10.1590/2317-1782/20182017225.

[230] Federico Calà, Lorenzo Frassineti, Giovanna Cantarella, Ludovica Battilocchi, Giulia Buccichini, Antonio Lanatà, and Claudia Manfredi. AI techniques applied to acoustical features of paralytic dysphonia versus dysphonia due to benign vocal fold masses. In *Models and Analysis of Vocal Emissions for Biomedical Applications: 13th International Workshop, September, 12-13, 2023*, pages 83–86. Firenze University Press, 2023.

[231] Stephen Politano, Fernando Morell, Kevin Calamari, Brad DeSilva, and Laura Matrka. Yield of imaging to evaluate unilateral vocal fold paralysis of unknown etiology. *The Laryngoscope*, 131(8):1840–1844, 2021. doi: 10.1002/lary.29152.

[232] Yehia A Ras, Mohamed Imam, Manal M El-Banna, and Nesrine H Hamouda. Voice outcome following electrical stimulation-supported voice therapy in cases of unilateral vocal fold paralysis. *The Egyptian Journal of Otolaryngology*, 32(4):322–334, 2016. doi: 10.4103/1012-5574.192543.

[233] Kiyohito Hosokawa, Toshihiko Iwahashi, Mio Iwahashi, Shinobu Iwaki, Chieri Kato, Misao Yoshida, Daichi Yoshida, Itsuki Kitayama, Masanori Umatani, Naoki Matsushiro, et al. The significant influence of hoarseness levels in connected speech on the voice-related disability evaluated using

voice handicap index-10. *Journal of Voice*, 37(2):290.e7–290.e16, 2020. doi: 10.1016/j.jvoice.2020.11.024.

[234] Mohaddeseh Rostampourgonbaki, Niloofar Deghanpour, Keivan Kiakojouri, Mehdi Dehghan, and Hemmat Gholinia. Acoustic voice measures in benign mass lesions. *Iranian Rehabilitation Journal*, 20(3):363–368, 2022. doi: 10.32598/irj.20.3.1596.1.

[235] Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM transactions on audio, speech, and language processing*, 29:3451–3460, 2021.

[236] Federico Calà. Vocal biomarkers of dysphonia and their interpretability: Challenges for an understandable ai. In *Models and Analysis of Vocal Emissions for Biomedical Applications: 15th International Workshop*, pages 35–38, 2025.

[237] Federico Calà, Lorenzo Frassineti, Giovanna Cantarella, Giulia Buccichini, Ludovica Battilocchi, and Antonio Lanata. On the complexity matching and multiscale nonlinear perspective of voice restoration via fat injection laryngoplasty in unilateral vocal fold paralysis. *Scientific Reports*, 15(1):31801, 2025.

[238] Federico Calà, Pietro Tarchi, Lorenzo Frassineti, Giovanna Cantarella, Giulia Buccichini, Ludovica Battilocchi, and Antonio Lanata. Acoustic and sentiment analyses for voice and life quality characterisation after medical intervention in voice pathologies. In *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 400–403. IEEE, 2025.

[239] Ji-Yeoun Lee, Ji-Hye Park, Ji-Na Lee, and Ah-Ra Jung. Investigation of the clinical effectiveness and prognostic factors of voice therapy in voice disorders: A pilot study. *Applied Sciences*, 13(20):11523, 2023.

[240] Yashes Srinivasan, Amy Liu, and Anaïs Rameau. Machine learning in the evaluation of voice and swallowing in the head and neck cancer patient. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 32(2):105–112, 2024.

[241] Andreas Triantafyllopoulos, Markus Fendler, Anton Batliner, Maurice Gerczuk, Shahin Amiriparian, Thomas M Berghaus, and Björn W Schuller. Distinguishing between pre-and post-treatment in the speech of patients with chronic obstructive pulmonary disease. *arXiv preprint arXiv:2207.12784*, 2022.

[242] Antonio Suppa, Francesco Asci, Giovanni Saggio, Luca Marsili, Daniele Casali, Zakarya Zarezadeh, Giovanni Ruoppolo, Alfredo Berardelli, and Giovanni Costantini. Voice analysis in adductor spasmodic dysphonia: Objective diagnosis and response to botulinum toxin. *Parkinsonism & related disorders*, 73:23–30, 2020. doi: 10.1016/j.parkreldis.2020.03.012.

[243] Yael Bensoussan, Christopher Park, Michael Johns III, Sarah Brown, Jeremy Pinto, Joseph Chang, and Mark Courey. A comparison of an artificial intelligence tool to fundamental frequency as an outcome measure in people seeking a more feminine voice. *The Laryngoscope*, 131(11):2567–2571, 2021. doi: 10.1002/lary.29605.

[244] Max Little, Patrick Mcsharry, Stephen Roberts, Declan Costello, and Irene Moroz. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Nature Precedings*, pages 1–1, 2007.

[245] Giovanna Cantarella, Riccardo F Mazzola, Michele Gaffuri, Elisabetta Iofrida, Pietro Biondetti, Laura V Forzenigo, Lorenzo Pignataro, and Sara Torretta. Structural fat grafting to improve outcomes of vocal folds' fat augmentation: long-term results. *Otolaryngology–Head and Neck Surgery*, 158(1):135–143, 2018. doi: 10.1177/0194599817739256.

[246] Wen-Dien Chang, Sheng-Hwa Chen, Ming-Hsui Tsai, and Yung-An Tsou. Autologous fat injection laryngoplasty for unilateral vocal fold paralysis. *Journal of Clinical Medicine*, 10(21):5034, 2021.

[247] Leonardo Wanderley Lopes, Vinícius Jefferson Dias Vieira, Silvana Luciene do Nascimento Cunha Costa, Suzete Élida Nóbrega Correia, and Mara Behlau. Effectiveness of recurrence quantification measures in discriminating subjects with and without voice disorders. *Journal of Voice*, 34(2):208–220, 2020.

[248] Vasileios Skaramagkas, Anastasia Pentari, Dimitrios I Fotiadis, and Manolis Tsiknakis. Using the recurrence plots as indicators for the recognition of Parkinson's disease through phonemes assessment. 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pages 1–4. IEEE, 2023.

[249] Minoru Hirano. Clinical examination of voice. *Disorders of human communication*, 5:1–99, 1981. ISSN 0173-170X.

[250] Ben Barsties v. Latoszek, Youri Maryn, Ellen Gerrits, and Marc De Bodt. A meta-analysis: Acoustic measurement of roughness and breathiness. *Journal of Speech, Language, and Hearing Research*, 61(2):298–323, 2018. doi: 10.1044/2017_JSLHR-S-16-0188.

[251] Tarika Bhuta, Linda Patrick, and James D Garnett. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *Journal of voice*, 18(3):299–304, 2004.

[252] Chloe Walton, Erin Conway, Helen Blackshaw, and Paul Carding. Unilateral vocal fold paralysis: a systematic review of speech-language pathology management. *Journal of Voice*, 31(4):509–e7, 2017.

[253] C-K Peng, Shlomo Havlin, H Eugene Stanley, and Ary L Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: an interdisciplinary journal of nonlinear science*, 5(1):82–87, 1995.

[254] Espen AF Ihlen. Introduction to multifractal detrended fluctuation analysis in Matlab. *Frontiers in Physiology*, 3:141, 2012.

[255] Zainy MH Almurad, Clément Roume, and Didier Delignières. Complexity matching in side-by-side walking. *Human Movement Science*, 54:125–136, 2017.

[256] Drew H Abney, Alexandra Paxton, Rick Dale, and Christopher T Kello. Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General*, 143(6):2304, 2014.

[257] Jan W Kantelhardt, Stephan A Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, and H Eugene Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):87–114, 2002.

[258] Karim Ansari-Asl, Guillaume Chanel, and Thierry Pun. A channel selection method for EEG classification in emotion assessment based on synchronization likelihood. pages 1241–1245, 2007.

[259] Madalena Costa, Ary L Goldberger, and C-K Peng. Multiscale entropy analysis of biological signals. *Physical Review E-Statistical, Nonlinear, and Soft Matter Physics*, 71(2):021906, 2005.

[260] Xinlin Xu, Peiyun Zhuang, Azure Wilson, and Jack J Jiang. Compensatory movement of contralateral vocal folds in patients with unilateral vocal fold paralysis. *Journal of Voice*, 35(2):328–e23, 2021.

[261] Lucia D'Alatri, Stefania Galla, Mario Rigante, Olga Antonelli, S Buldrini, and Maria Raffaella Marchese. Role of early voice therapy in patients affected by unilateral vocal fold paralysis. *The Journal of Laryngology & Otology*, 122(9):936–941, 2008.

[262] Luis MT Jesus, Joana Martinez, Andreia Hall, and Aníbal Ferreira. Acoustic correlates of compensatory adjustments to the glottic and supraglottic structures in patients with unilateral vocal fold paralysis. *BioMed Research International*, 2015(1):704121, 2015.

[263] Julian D Arias-Londono, Juan I Godino-Llorente, Nicolás Sáenz-Lechón, Víctor Osma-Ruiz, and Germán Castellanos-Domínguez. Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. *IEEE Transactions on Biomedical Engineering*, 58(2):370–379, 2010.

[264] Asanori Kiyuna, Norimoto Kise, Munehisa Hiratsuka, Hiroyuki Maeda, Hitoshi Hirakawa, Akira Ganaha, and Mikio Suzuki. Brain activity in patients with unilateral vocal fold paralysis detected by functional magnetic resonance imaging. *Journal of Voice*, 36(5):738–e1, 2022.

[265] Marie Dedry, Laurence Dricot, Vinciane Van Parys, Donatienne Boucquey, Nicolas Delinte, Julie van Lith-Bijl, Arnaud Szmalec, Youri Maryn, and Gauthier Desuter. Brain adaptation following various unilateral vocal fold paralysis treatments: A magnetic resonance imaging based longitudinal case series. *Frontiers in Neuroscience*, 16:947390, 2022.

[266] Elisabetta Sforza, Federico Calà, Claudia Manfredi, Antonio Lanatà, Andrea Guala, Cesare Danesino, Angelina Cistaro, Matelda Mazzocca, Lucia D'Alatri, Roberta Onesimo, et al. From phenotype to phonotype: a comprehensive description of voice features of cri du chat syndrome. *European Journal of Pediatrics*, 184(1):1–11, 2025.

[267] Federico Calà, Elisabetta Sforza, Lucia D'Alatri, Lorenzo Frassineti, Claudia Manfredi, Roberta Onesimo, Donato Rigante, Marika Pane, Serenella Servidei, Guido Primiano, et al. Do rare genetic conditions exhibit a specific phonotype? A comprehensive description of the vocal traits associated with crisponi/cold-induced sweating syndrome type 1. *Genes*, 16(8):881, 2025.

[268] Josh Howarth. How many people own smartphones? (2025-2029). `https://explodingtopics.com/blog/smartphone-stats`. Accessed: 20-06-2025.

[269] Lorena Parra, Sandra Sendra, José Miguel Jiménez, and Jaime Lloret. Multimedia sensors embedded in smartphones for ambient assisted living and e-health. *Multimedia Tools and Applications*, 75:13271–13297, 2016.

[270] Sumit Majumder and M Jamal Deen. Smartphone sensors for health monitoring and diagnosis. *Sensors*, 19(9):2164, 2019.

[271] Rakibul Hoque and Golam Sorwar. Understanding factors influencing the adoption of mhealth by the elderly: An extension of the UTAUT model. *International journal of medical informatics*, 101:75–84, 2017.

[272] Claudia Manfredi, Jean Lebacq, Giovanna Cantarella, Jean Schoentgen, Silvia Orlandi, Andrea Bandini, and Philippe H DeJonckere. Smartphones offer new opportunities in clinical voice research. *Journal of Voice*, 31(1):111–e1, 2017.

[273] OMIM. Omim gene map statistics. `https://www.omim.org/statistics/geneMap`. Accessed: 18-06-2025.

[274] Stéphanie Nguengang Wakap, Deborah M Lambert, Annie Olry, Charlotte Rodwell, Charlotte Gueydan, Valérie Lanneau, Daniel Murphy, Yann Le Cam, and Ana Rath. Estimating cumulative point prevalence of rare diseases: analysis of the orphanet database. *European journal of human genetics*, 28(2):165–173, 2020.

[275] Linda Sohner and Pamela Mitchell. Phonatory and phonetic characteristics of prelinguistic vocal development in cri du chat syndrome. *Journal of communication disorders*, 24(1):13–20, 1991.

[276] Doris Braunstein, Holly Jones, and Colleen Heffernan. Laryngeal findings in a 20-month-old with cri du chat syndrome. *The Laryngoscope*, 135(5):1774–1776, 2025.

[277] Kristian E Kristoffersen. Vowel production in cri du chat-syndrome–a case study. In *The 15th International Congress of Phonetic Sciences, Barcelona.(Forthcoming)*, 2003.

[278] Giangiorgio Crisponi. Autosomal recessive disorder with muscle contractions resembling neonatal tetanus, characteristic face, camptodactyly, hyperthermia, and sudden death: a new syndrome? *American journal of medical genetics*, 62(4):365–371, 1996.

[279] Insa Buers, Ivana Persico, Lara Schöning, Yvonne Nitschke, Maja Di Rocco, Angela Loi, Puneet Kaur Sahi, Gulen Eda Utine, Bilge Bayraktar-Tanyeri, Giuseppe Zampino, et al. Crisponi/cold-induced sweating syndrome: Differential diagnosis, pathogenesis and treatment concepts. *Clinical Genetics*, 97(1):209–221, 2020.

[280] Ingrid Anne Mandy Schierz, Gregorio Serra, Vincenzo Antona, Ivana Persico, Giovanni Corsello, and Ettore Piro. Infant developmental profile of Crisponi syndrome due to compound heterozygosity for CRLF1 deletion. *Clinical dysmorphology*, 29(3):141–143, 2020.

[281] Roberta Onesimo, Elisabetta Sforza, Federica Palermo, Valentina Giorgio, Chiara Leoni, Donato Rigante, Valentina Trevisan, Cristiana Agazzi, Domenico Limongelli, Francesco Proli, et al. Feeding and nutritional key features of crisponi/cold-induced sweating syndrome. *Genes*, 15(9):1109, 2024.

[282] Taralee Hamner, Manisha D Udhnani, Karol Z Osipowicz, and Nancy Raitano Lee. Pediatric brain development in down syndrome: a field in its infancy. *Journal of the International Neuropsychological Society*, 24(9):966–976, 2018.

[283] Ray D Kent and Houri K Vorperian. Speech impairment in down syndrome: A review. 2013.

[284] Minkyung Kang and Yong-Seok Lee. The impact of rasopathy-associated mutations on cns development in mice and humans. *Molecular brain*, 12(1):96, 2019.

[285] Geciane Xavier Torres, Emerson de Santana Santos, Carla Patrícia Hernandez Alves Ribeiro César, Roxane de Alencar Irineu, Isabel Ribeiro Rocha Dias, and Alice Fontes Ramos. Clinical orofacial and myofunctional manifestations in an adolescent with noonan syndrome: a case report. *Revista CEFAC*, 22:e16519, 2020.

[286] Elizabeth I Pierpont, Susan Ellis Weismer, Amy E Roberts, Erica Tworog-Dube, Mary Ella Pierpont, Nancy J Mendelsohn, and Mark S Seidenberg. The language phenotype of children and adolescents with noonan syndrome. *Language*, 2010.

[287] Voice characteristics of children with noonan syndrome, author=Türkyilmaz, Meral and Tokgöz Yılmaz, SUNA and Özcebe, Esra and Yüksel, Selcen and Süslü, Nilda and Tekin, Mustafa, journal=Turkiye Klinikleri Journal of Medical Sciences, volume=34, number=2, year=2014.

[288] Olga Madej, Rebecka Kiff, and Haytham Kubba. How common are ear, nose and throat disorders in children with noonan syndrome and other rasopathies? *International Journal of Pediatric Otorhinolaryngology*, 166:111480, 2023.

[289] Karen W Gripp, Lindsey A Morse, Marni Axelrad, Kathryn C Chatfield, Aaron Chidekel, William Dobyns, Daniel Doyle, Bronwyn Kerr, Angela E Lin, David D Schwartz, et al. Costello syndrome: clinical phenotype, genotype, and management guidelines. *American Journal of Medical Genetics Part A*, 179(9):1725–1744, 2019.

[290] Heather L Thompson. Speech-language pathology in the rasopathies. In *The RASopathies: Genetic Syndromes of the RAS/MAPK Pathway*, pages 653–678. Springer, 2024.

[291] Andrea L Gropman, Wallace C Duncan, and Ann CM Smith. Neurologic and developmental features of the smith-magenis syndrome (del 17p11. 2). *Pediatric neurology*, 34(5):337–350, 2006.

[292] Irene Hidalgo-De la Guía, Elena Garayzábal-Heinze, Pedro Gómez-Vilda, Rafael Martínez-Olalla, and Daniel Palacios-Alonso. Acoustic analysis of phonation in children with smith–magenis syndrome. *Frontiers in human neuroscience*, 15:661392, 2021.

[293] Berardo Rinaldi, Roberta Villa, Alessandra Sironi, Livia Garavelli, Palma Finelli, and Maria Francesca Bedeschi. Smith-magenis syndrome - clinical review, biological background and related disorders. *Genes*, 13(2):335, 2022.

[294] Juli G Pausas and Jon E Keeley. Wildfires and global change. *Frontiers in Ecology and the Environment*, 19(7):387–395, 2021.

[295] Projected changes in fire danger. `https://www.eea.europa.eu/en/analysis/maps-and-charts/projected-meteorological-forest-fire-danger`. Accessed: 25-02-2025.

[296] Global Wildfire Information System. Annual area burnt per wildfire. `https://ourworldindata.org/wildfires`. Accessed: 25-02-2025.

[297] Global Wildfire Information System. Severe wildfires in portugal in september 2024. `https://www.copernicus.eu/pt-pt/node/68750`. Accessed: 25-02-2025.

[298] City of Beverly Hills. Home hardening. `https://www.beverlyhills.org/554/Home-Hardening`. Accessed: 25-02-2025.

[299] City of Maple Ridge. Firesmart program. `https://www.mapleridge.ca/community-safety/fire-rescue/fire-prevention/firesmart-program`. Accessed: 25-02-2025.

[300] Washington State Department of Natural Resources. Wildfire prevention. `https://www.dnr.wa.gov/WildfirePrevention`. Accessed: 25-02-2025.