Università Politecnica delle Marche
Doctoral School of Engineering Sciences
Industrial Engineering - curriculum in Mechanical Engineering

# Towards Objectivization of Acoustic Jury Tests: Physiological and Biometrical Measurement of Human Response to Acoustic Stimuli

**Reza Jamali**

Advisors:
**Prof. Paolo Castellini**

Co-Advisors:
**Prof. Milena Martarelli**
**Prof. Paolo Chiariotti**
**Dr. Gianmarco Battista**

Università Politecnica delle Marche
Department of Industrial Engineering and Mathematical Sciences
Via Brecce Bianche - 60131 - Ancona, Italy

# Abstract

In the evolving field of acoustic engineering, the subjective assessment of sound quality plays a pivotal role. Traditional methodologies predominantly rely on direct feedback from human jurors, a process often challenged by its subjective nature and potential for inconsistency. This thesis introduces an innovative approach to augment sound quality jury testing by integrating physiological and biometric measurements, specifically focusing on electroencephalography (EEG) and facial expression analysis. By measuring these responses simultaneously with the jurors' sound quality assessments, this research aims to provide a more objective, reliable, and nuanced understanding of human auditory perception.

To substantiate this approach, the thesis conducts two preliminary tests and two jury tests. The preliminary tests involved 16 and 40 participants, respectively, who were exposed to various audio and audio-visual stimuli. Their facial expressions were recorded to examine the emotional impact of these stimuli, highlighting the efficacy of using facial expression analysis to predict emotions elicited by acoustic inputs. The tests also demonstrated that visual components alongside audio stimuli can enhance emotional response. For analyzing facial expression, a key aspect of this research involved utilizing a computer vision-based software tool and a Convolutional Neural Network for assessing the level of attention in participants. Eventually, results showed that facial expression analysis is able to play an important role in jury testing, mainly in assessing the level of jurors' involvement, thus identifying more reliable responses among other responses provided by jurors.

In the EEG-focused jury test, forty-three participants were involved, with EEG signals recorded using wearable sensors. The analysis of power spectral densities (PSDs) was pivotal in identifying features correlated with acoustic sensations induced by the stimuli. The tests revealed statistically significant differences in responses to different audio stimuli. These findings underscore the potential of integrating wearable EEG sensors in jury test assessments, offering a novel perspective on how physiological measurements can enhance the reliability and depth of sound quality evaluations.

Subsequently, this thesis demonstrates that the integration of physiological and biometric measurements—specifically EEG and facial expression analysis—into sound quality jury testing can somehow enrich the assessment process. These methods provide a more objective basis for understanding auditory perception, offering promising

avenues for future research in acoustic engineering and related fields.

# Sommario


Nel campo in evoluzione dell'ingegneria acustica, la valutazione soggettiva della qualità del suono gioca un ruolo fondamentale. Le metodologie tradizionali si basano prevalentemente sul feedback diretto di giurati umani, un processo spesso messo in discussione dalla sua natura soggettiva e dal potenziale di incoerenza. Questa tesi introduce un approccio innovativo per incrementare l'affidabilità dei jury test per la valutazione soggettiva della qualità del suono integrando misure fisiologiche e biometriche, in particolare concentrandosi sull'elettroencefalografia (EEG) e sull'analisi dell'espressione facciale. Misurando queste risposte contemporaneamente alle valutazioni dei giurati sulla qualità del suono, questa ricerca mira a fornire una comprensione più oggettiva, affidabile e sfumata della percezione uditiva umana.

Per corroborare questo approccio, la tesi conduce due test preliminari e due jury test acustici. I test preliminari hanno coinvolto rispettivamente 16 e 40 partecipanti, che sono stati esposti a diversi stimoli audio e audiovisivi. Le loro espressioni facciali sono state registrate per esaminare l'impatto emotivo di questi stimoli, evidenziando l'efficacia dell'uso dell'analisi delle espressioni facciali per prevedere le emozioni suscitate dagli input acustici. I test hanno anche dimostrato che la risposta emotiva è più evidente quando gli stimoli audio sono accompagnati da componenti visive. Per l'analisi dell'espressione facciale, un aspetto fondamentale di questa ricerca è stato l'utilizzo di uno strumento software basato sulla computer vision e di una rete neurale convoluzionale per valutare il livello di attenzione dei partecipanti. Alla fine, i risultati hanno dimostrato che l'analisi dell'espressione facciale è in grado di svolgere un ruolo importante nei jury test, soprattutto nel valutare il livello di coinvolgimento dei giurati, identificando così risposte più affidabili tra le quelle fornite dai giurati.

Nei jury test acustici condotti in simultanea con l'acquisizione della risposta filologica dei giurati (mediante EEG) sono state coinvolte 43 persone, i cui segnali EEG sono stati registrati con sensori indossabili. L'analisi delle densità spettrali di potenza (PSD) è stata fondamentale per identificare le caratteristiche correlate alle sensazioni acustiche indotte dagli stimoli. I test hanno rivelato differenze statisticamente significative nelle risposte ai diversi stimoli audio. Questi risultati sottolineano il potenziale dell'integrazione di sensori EEG indossabili nelle valutazioni dei jury test, offrendo una prospettiva nuova su come le misurazioni fisiologiche possano migliorare l'affidabilità e l'accuratezza delle valutazioni della qualità del suono.

Di conseguenza, questa tesi dimostra che l'integrazione di misure fisiologiche e biometriche - in particolare l'EEG e l'analisi delle espressioni facciali - nei jury test per la valutazione della qualità del suono può in qualche modo arricchire il processo di valutazione. Questi metodi forniscono una base più oggettiva per la comprensione della percezione uditiva, offrendo strade promettenti per la ricerca futura nell'ingegneria acustica e nei campi correlati.

# Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Paolo Castellini, for his invaluable support throughout my PhD journey. and I am equally thankful to my advisors, Professor Milena Martarelli, Professor Paolo Chiarotti, and Dr. Gianmarco Battista, for their invaluable contributions to my academic development. The knowledge and expertise they have imparted have been instrumental in my learning and success.

My heartfelt thanks go to my family, specifically my mother and sister, whose unwavering support and encouragement have been my constant source of strength and motivation. Their belief in me has been a guiding light in my journey.

I am also grateful to my colleagues and friends, as well as my fellow researchers in the Eco-drive project. Their collaboration, camaraderie, and shared wisdom have greatly enriched my PhD experience.

A special thanks to Professor Jerome Antoni from INSA University, Lyon. My time at INSA was not only academically rewarding but also personally fulfilling. Working with Professor Antoni and our colleague, Muhammad Albezzawi, on signal processing was a great experience that significantly enhanced my research skills and knowledge.

I extend my gratitude to Dr. Rafael Dias, my colleague in the same lab and consortium. His assistance, advice, and willingness to share his expertise were always invaluable to me.

Finally, I am honored to have had the opportunity to work in the Laboratory of Measurement in the Department of Industrial Engineering and Mathematical Sciences at the Università Politecnica delle Marche. This experience has been a significant milestone in my academic career, providing me with a platform to grow and excel.

This project has received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Marie Sklodowska-Curie Grant Agreement nº 858018.

Reza Jamali
Ancona, 30/11/2023

# Contents


**Abstract** **iii**

**Sommario** **v**

**Acknowledgements** **vii**

**Nomenclature** **xvii**

**1 Introduction** **1**

1.1 Background and Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

1.1.1 The Impact of Auditory Experience in Product Design . . . . . . 1

1.1.2 Introduction to Sound Quality Metrics . . . . . . . . . . . . . . . . 1

1.1.3 Jury Testing in Product Sound Quality Assessment . . . . . . . . 3

1.1.4 Ensuring Accuracy and Reliability in Jury Tests . . . . . . . . . 5

1.2 Research Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

1.3 Thesis Outline . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

**2 Literature Review** **9**

2.1 Acoustic Assessment, and Jury Testing . . . . . . . . . . . . . . . . . . . . 9

2.2 Facial Expression Analysis in Emotion Detection . . . . . . . . . . . . . 11

2.3 EEG as a Tool in Acoustic Evaluation . . . . . . . . . . . . . . . . . . . . 14

2.4 Research Gaps . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

**3 Jury Test Combined with Facial Expression Measurement** **17**

3.1 Theoretical Framework: Emotion and Facial Expression . . . . . . . . . 18

3.2 Device Specifications . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

3.3 Preliminary test 1: Stimulating Facial Expression by Video/Sounds . . . 21

3.3.1 Participants' Demographics . . . . . . . . . . . . . . . . . . . . . . 21

3.3.2 Design and Procedure . . . . . . . . . . . . . . . . . . . . . . . . . 21

3.3.3 Sequence . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

3.4 Preliminary Test 2: Emotion Response to Acoustic Stimuli . . . . . . . 25

3.4.1 Objective of the Test Campaign . . . . . . . . . . . . . . . . . . . 25

3.4.2 Test Sequence . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25

3.4.3 Acoustic Perception Feedback . . . . . 25
3.5 Experimental Study of Combining Facial Expression with Traditional Jury Test . . . . . 26
3.5.1 Test Environment and Procedure . . . . . 27
3.5.2 Methodology . . . . . 27
3.5.3 Jury Testing Process and Sequence . . . . . 29
3.5.4 Interface app and input device . . . . . 30

**4 Jury Test Combined with Physiological Response 33**
4.1 Theoretical Framework . . . . . 33
4.1.1 Standard EEG measurement . . . . . 33
4.1.2 EEG Recording Devices . . . . . 35
4.1.3 Standardized Systems for EEG Electrode Placement . . . . . 36
4.1.4 Characterization and Analysis of EEG signals . . . . . 37
4.1.5 EEG Signals Oscillations . . . . . 40
4.2 Device Specifications . . . . . 42
4.2.1 Equipment Used for EEG Signal Acquisition . . . . . 42
4.2.2 Earphones . . . . . 43
4.2.3 Data Acquisition Process . . . . . 44
4.3 EEG Test Campaign for Acoustic Evaluation . . . . . 44
4.3.1 Experimental Protocol . . . . . 45
4.3.2 Test Contents and Sequence . . . . . 48
4.3.3 Estimation of Effective Sample Size . . . . . 49
4.3.4 Participants Demographics . . . . . 49
4.3.5 EEG Signal Pre-processing . . . . . 50
4.3.6 Quantification of Individual Sound Perceptions . . . . . 50
4.3.7 Quantifying Measurement Uncertainty . . . . . 51

**5 Results and Discussions 53**
5.1 Results of Experiments on Facial Expression in Acoustic Perception . . 53
5.1.1 Results of Preliminary Test 1 . . . . . 53
5.1.2 Results of Preliminary Test 2 . . . . . 56
5.1.3 Results of Jury Test Campaign enhanced by Facial Expression Measurement . . . . . 57
5.2 Results of Physiological Measurement . . . . . 62
5.2.1 Subjective Acoustic Perception Distribution . . . . . 63
5.2.2 EEG-Acoustic Perception Correlation . . . . . 63
5.2.3 Theta Waves and Annoyance . . . . . 64
5.2.4 Alpha Waves and Auditory Attentiveness . . . . . 64
5.2.5 Interplay Between Auditory Context and Brain Activation . . . . 64

**6 Conclusion 69**

**References** **73**

# List of Figures


1.1 An overview of objective sound quality metrics categorized into modulation, strength, and spectral content [1]. . . . . . . . . . . . . . . . . . . . 3
1.2 Subjective sound quality metrics (performing jury testing). . . . . . . . 3
1.3 An overview of the objectives of the thesis . . . . . . . . . . . . . . . . . 6

3.1 Dlib facial landmarks used for attention evaluation [2] . . . . . . . . . . 19
3.2 (a): face recording by a Logitech HD C925E webcam. (b): AKG K702 headphone to playback of sounds to the participants . . . . . . . . . . . 21
3.3 View of the performing the preliminary tests 1 . . . . . . . . . . . . . . 24
3.4 Sequence of the first preliminary test. (a): sequence of audio, between [0 250] seconds. (b): sequence of audio plus video, between [210 460] seconds. (c): sequence of validated video, from 420 seconds until the end 24
3.5 View of the performing the preliminary tests 2 . . . . . . . . . . . . . . 26
3.6 Test Sequence for the second preliminary test . . . . . . . . . . . . . . . 26
3.7 Questionnaire capturing the jurors' self-declaration of their involvement level performed in Google Forms. The jurors filled out this questionnaire at the end of the test . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
3.8 Experimental setup overview of sound quality jury assessment test . . . 30
3.9 A screenshot of the graphical user interface used to interact with the jurors 31
3.10 Using the external mini keyboard used to collect input answers from the jurors . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

4.1 Generation of extracellular voltage fields from graded synaptic activity. Relationship between polarity of surface potentials and site of dendritic postsynaptic potentials [3]. . . . . . . . . . . . . . . . . . . . . . . . . . 34
4.2 A view of different EEG recording devices with different electrode types. (a) dry, (b) saline solution, (c) gel-based [4] . . . . . . . . . . . . . . . . 36
4.3 Electrode positions and labels in the 10-20 and 10-10 system. Black circles indicate positions of the original 10-20 system, gray circles indicate additional positions introduced in the 10-10 extension [5] . . . . . . . . . 38
4.4 Selection of 100-5 electrode positions in a realistic display [5] . . . . . . 39

4.5 A view of the Interaxon MUSE headband and the locations of its electrodes in the 10-20 illustration system . . . . . 42
4.6 View of the performing the test . . . . . 46
4.7 Experiment sequence . . . . . 46
4.8 Sound quality assessment questionnaire created by GoogleForms. (a), (b), (c), and (d) respectively illustrate the first, second, third, and fourth pages of the questionnaire. . . . . . 47
4.9 Spectrograms illustrating the spectral content of three sound stimuli over time. (a) engine noise, (b) soothing music, and (c) road noise . . . . . 52

5.1 Engagement values averaged over each stimulus for Subject No.4 . . . . 54
5.2 Engagement values averaged over each stimulus for Subject No.3 . . . . 55
5.3 Mean values and 95% confidence intervals for the means of the 3 macro-categories used in preliminary test 1 . . . . . 56
5.4 Changes in anger values extracted for the periods of times when the jurors chose audio B as the more annoying audio . . . . . 58
5.5 Changes in valence values extracted for the periods of times when the jurors chose audio B as the more annoying audio . . . . . 58
5.6 Changes in engagement values extracted for the periods of times when the jurors chose audio B as the more annoying audio . . . . . 59
5.7 Radar chart to represent the jurors' concentration levels throughout the test campaign . . . . . 60
5.8 Bar chart to illustrate the quantitative jurors' concentration . . . . . 60
5.9 Boxplot charts of attention values for different concentration statuses. Note that the attention values normalized on a scale from 0 to 1 . . . . 62
5.10 Comparative boxplot charts depicting attention values across various concentration statuses, with "fully concentrated" as the baseline . . . . 62
5.11 Subjective questionnaire results overview . . . . . 63
5.12 Histograms of Dwass-Seel-Critchlow-Fligner Pairwise Comparing Audio 1 and Audio 2 in terms of delta wave. The ranges above the bars represent the standard deviation. (a): related to the electrode AF8. (b): related to the electrode TP10. . . . . . 65
5.13 Histograms of Dwass-Seel-Critchlow-Fligner Pairwise Comparing Audio 1 and Audio 2 observed in the electrode TP10. The ranges above the bars represent the standard deviation. (a): relative alpha wave. (b): theta wave. . . . . . 66
5.14 Histograms of Dwass-Seel-Critchlow-Fligner Pairwise Comparing Audio 2 and Audio 3 observed in terms of TA alpha. The ranges above the bars represent the standard deviation. . . . . . 66

# List of Tables


4.1 Correlation of EEG Bands with Brain States [6] . . . . . . . . . . . . . . . 40
4.2 Spectral Characteristics of the Sounds. The presented values are median values derived from the temporal distribution of frequency components. 49
4.3 G*Power3 statistical setting parameters. . . . . . . . . . . . . . . . . . . . 49

5.1 Correlations coefficient between the emotional valence and the questionnaire answers given by the jurors . . . . . . . . . . . . . . . . . . . . . . . . 56
5.2 Repeated measured ANOVA - Kruskal-Walli's test results . . . . . . . . 65

# Nomenclature

**Acronyms**

ANN Artificial Neural Network

BCI Brain-Computer Interfaces

BLE Bluetooth Low Energy

CNN Convolutional Neural Network

CPU Central Processing Unit

CSV Comma-Separated Values

DSP Digital Signal Processing

EEG Electroencephalography

ERP Event-Related Potential

FA Frontal Asymmetry

FACS Facial Action Coding System

FER Facial Expression Recognition

FFT Fast Fourier Transform

GUM Guide to the Expression of Uncertainty in Measurement

HAP-RM Human Auditory Perception-based Roughness Metric

IQR Interquartile Range

PAD Pleasure, Arousal, Dominance model

PC Personal Computer

PSD Power Spectral Density

qEEG Quantitative Electroencephalography

SPL Sound Pressure Level

TA Temporal Asymmetry

VR Virtual Reality

**Roman letters**

B The power of the EEG signal

C Central brain cortex

dB Decibels

F Frontal brain cortex

Fp Frontopolar brain cortex

I Periodograms

L Signal length

O Occipital brain cortex

P Parietal brain cortex

P Total EEG power

p Probability value in statistical tests

t Time

w Window function

Z Z-score in statistical tests

**Greek letters**

$\alpha$ Alpha, EEG wave

$\beta$ Beta, EEG wave

$\delta$ Delta, EEG wave

$\gamma$ Gamma, EEG wave

$\theta$ Theta, EEG wave

$\bar{f}$ Arithmetic mean

$\chi^2$ -square statistic

$u_f$ Standard uncertainty of the mean

# Chapter 1

# Introduction

## 1.1 Background and Motivation

### 1.1.1 The Impact of Auditory Experience in Product Design

The auditory experience in product design is a crucial element that goes beyond the simple reaction to sound. It is an intricate process where the sound produced by a product influences the users' perceptions and feelings. Users often associate the sound of a product with characteristics like reliability and high quality. This association affects users' emotions, the relationship they build with the product, and their decision to purchase it[7].

Every sound a product makes interacts with the users' senses, creating an environment that influences their perception and experience. These sounds are linked with different attributes of the product, affecting how users view its value and effectiveness. For example, a smooth and quiet engine sound in a car can make the car seem more luxurious and well-designed. In today's market, customers pay close attention to the sound quality of a product, along with its functionality and performance. The noise level and quality of a product are important factors that help shape the brand image and customer loyalty. For example, products like cars, household appliances, and architectural spaces are often designed with a focus on creating pleasing and effective sounds [8].

Sound design is integrated into various products to enhance their appeal and performance. It is used in cars to improve the driving experience, in household appliances to make them seem more efficient, and in architectural spaces to create a better environment. In conclusion, sound design plays a vital role in product design, affecting user satisfaction, brand image, and the overall success of a product in the market [9].

### 1.1.2 Introduction to Sound Quality Metrics

Sound quality assessment is a multifaceted field that is classified primarily into two distinct types of metrics: subjective and objective. Subjective metrics which commonly will be performed by jury testing are related to the personal, human experience of sound, which encompasses the perceptions, preferences, and responses of listeners in various acoustic scenarios. Figure 1.2 represents a view of performing subjective sound quality metrics. Objective metrics, which are schematically shown in Figure 1.1, on the other hand, entail the quantification of sound quality using precise, measurable parameters that can be systematically analyzed and replicated. These metrics facilitate the consistent assessment of sound quality across different environments and systems [10].

This section explores the three primary categories of objective sound quality metrics: modulation-based, strength-based, and spectral-based metrics.

### Modulation-Based Metrics

Modulation encompasses the temporal variations in a sound signal. Modulation metrics evaluate how the fluctuations of a sound signal influence the listener's perception. Fluctuation strength and roughness are two well-known examples of modulation-based metrics [11].

### Strength-Based Metrics

The strength of a sound relates to its power or energy, typically measured in decibels (dB). Strength-based metrics relate to the perceived power [12].

- **Loudness:** A psychoacoustic metric reflecting the perceived strength of a sound.
- **Sound Pressure Level (SPL):** While SPL is a physical measure, it correlates with perceived loudness, with A-weighting providing a more accurate representation of perception.
- **Dynamic Range:** The ratio between the softest and loudest parts of a sound, with a wide dynamic range often associated with high fidelity.

### Spectral Content-Based Metrics

The spectral content of a sound determines its timbre. Metrics in this category analyze frequency distribution and energy to assess quality. Tonality, sharpness, and Signal-to-Noise Ratio are some examples of spectral content-based metrics [13].

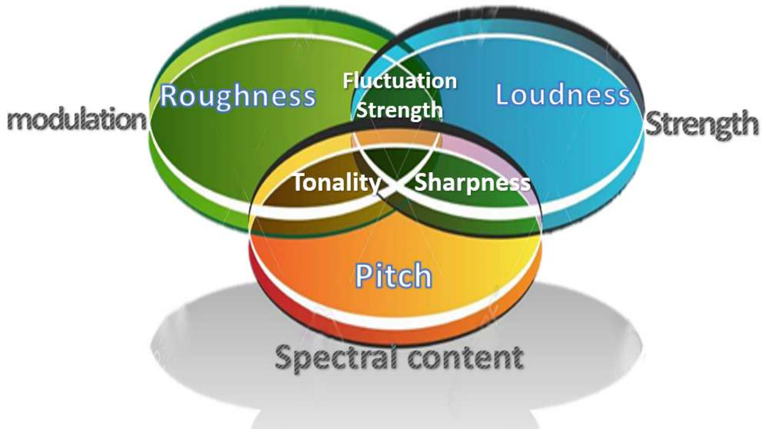



**Figure 1.1:** An overview of objective sound quality metrics categorized into modulation, strength, and spectral content [1].

**Figure 1.2:** Subjective sound quality metrics (performing jury testing).

### 1.1.3 Jury Testing in Product Sound Quality Assessment

Jury testing is a pivotal methodology employed in the realm of product development, focusing predominantly on evaluating the perceived quality of sounds emanated by various products. It involves a systematic approach where a selected group of individuals, referred to as jurors, participate actively in assessing and rating the sound quality of a product under multiple operational scenarios. These jurors play a crucial role by providing invaluable feedback, reflecting potential customer reactions and perceptions towards the auditory aspects of new products[1].

The fundamental objective of jury testing lies in garnering insights into how a collective group of users perceive and evaluate the sounds produced by a product. The sound samples utilized in these experiments are meticulously recorded to capture the product's auditory performance under a diverse array of operating conditions. Jurors are then exposed to these sounds, and their opinions and ratings are solicited through a series of structured questions. Their feedback establishes a nuanced understanding of the correlation between the sound attributes of a product and its perceived quality. In ensuring the applicability of the jury testing process, a statistically representative sample of jurors is selected, often embodying the characteristics of potential future customers. Through this strategic selection process, the obtained results are rendered more authentic and reflective of genuine customer perceptions and preferences. The jurors' responses and ratings serve as a critical feedback mechanism, facilitating informed modifications and enhancements in the sound design of the investigated products[14][15].

Jury testing is a pivotal experimental activity in the development of new products, focusing explicitly on evaluating the perceived quality of sounds emanating from the products. This methodology involves conducting a series of tests where a group of individuals, referred to as jurors, are asked specific questions about a product's sound quality. The essence of jury testing lies in understanding and analyzing how a collective group of users perceive, interpret, and rate the auditory aspects of a product. These evaluations subsequently serve as crucial feedback, instrumental in guiding the future modifications and enhancements of the investigated products [16].

In a typical jury test, sounds emanating from a product under various operational conditions are recorded and presented to a group of jurors. These jurors are generally selected based on their representation of the potential future customer base of the product. Jurors are exposed to each sound or group of sounds, followed by a series of questions aimed at garnering their perceptions, feelings, and ratings of the sound quality. These interactions establish a profound relationship between the sound emanations and the perceived quality of the product in question, providing a rich dataset for further analysis and interpretation [17][18][19].

The questions which will be asked during the jury tests can span a wide spectrum, ranging from the pleasantness or annoyance of the sounds, to more subtle inquiries such as which sounds are perceived as more luxurious or robust. Central to the design of the questions is their alignment with the overarching objectives of the test, ensuring that they elicit responses that are both relevant and insightful. Responses garnered during jury tests are inherently subjective, rooted in the personal opinions and perceptions of the jurors. Thus, there are no correct or incorrect answers, allowing for a diversity of opinions and perceptions to be captured and analyzed. Post-testing, the responsibility resides with the test moderators to meticulously analyze the collected responses. A comprehensive analysis often involves calculating average ratings, identifying prevalent trends and patterns in jurors' responses, and distilling these findings into actionable insights. This meticulous approach to analysis ensures that the extracted insights are both representative and robust, providing a solid foundation upon which future product modifications and enhancements can be judiciously informed and implemented [20][21].

### 1.1.4 Ensuring Accuracy and Reliability in Jury Tests

Ensuring accuracy and reliability in the results derived from jury tests is an aspect of paramount importance. This has captivated significant attention within the scientific acoustics community, driven by a necessity to affirm the objectivity of the subjective outputs garnered from jury tests. A central challenge intrinsic to jury tests revolves around ensuring that the measurements obtained are not only reliable and repeatable but also encapsulated with a degree of objectivity that minimizes the subjective influences exerted by both the moderator and the jurors. Traditionally, increasing the size of the participant pool has been viewed as the most effective method to achieve this. However, this approach often encounters practical limitations. Expanding the size of the jury entails significant efforts, particularly in terms of human resources, which is a major constraint in conducting sound quality tests. The need for a large number of participants can make the process resource-intensive and less feasible in many scenarios [22].

As an alternative, the integration of physiological and biometrics measurements alongside traditional jury testing presents a promising solution. This approach can significantly enhance the reliability and repeatability of the tests. By leveraging physiological data, it is possible to gain deeper insights into the participants' responses to sound stimuli without the need to substantially increase the number of jury members.

The use of physiological and biometrics measurements offers several advantages. Primarily, it simplifies the testing process as it does not require additional human resources in the form of extra participants. Moreover, this method can provide more nuanced data that may not be explicitly captured through conventional jury testing methods. The integration of these measurements can lead to more comprehensive and reliable assessments of sound quality, making it a valuable tool in both research and practical applications.

## 1.2 Research Objectives

The primary objective of this research is to enhance the reliability and predictive accuracy of subjective sound quality jury testing results by incorporating biometric and physiological parameters measurement. This study explores the feasibility of utilizing Electroencephalogram (EEG) data and facial expression analysis as innovative tools to refine the outcomes of the jury tests in assessing sound quality. as it is shown in Figure 1.3, specific objectives include:

- Investigating the correlation between EEG data and facial expression analysis with the jurors' subjective responses, aiming to establish a methodology that allows for a more accurate prediction of subjective sound quality ratings.

- Assessing the level of involvement of jurors during the evaluation process by analyzing their facial expressions. This assessment aims to determine the reliability of jurors' responses, enabling the extraction of more dependable data for a comprehensive understanding of subjective sound quality assessments.

In pursuing these objectives, this study tries to present a novel approach in the subjective evaluation of sound quality, striving for more robust, reliable, and enriched data to enhance the accuracy and integrity of sound quality jury testing methodologies. Through a series of meticulously conducted experiments, this research aims to unveil significant insights into the potential of biometric and physiological parameters as pivotal contributors to improving traditional sound quality evaluation practices.

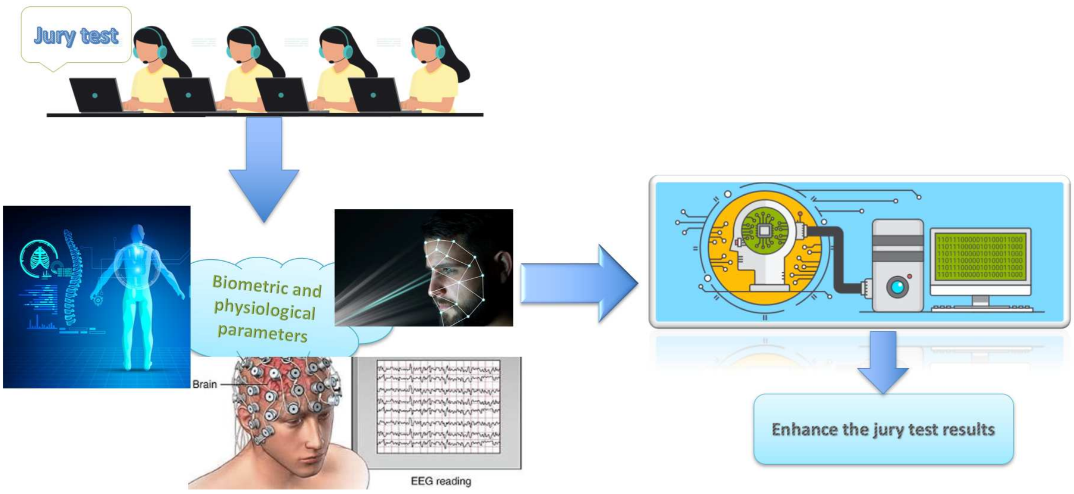



**Figure 1.3:** An overview of the objectives of the thesis

## 1.3 Thesis Outline

This thesis is structured into six chapters, each focusing on a specific aspect of the study on sound quality metrics and the integration of facial expression and EEG measurements in jury testing.

This thesis is structured into six chapters, each focusing on a specific aspect of the study on sound quality metrics and the integration of facial expression and EEG measurements in jury testing.

**Chapter 1: Introduction to Sound Quality Metrics** This chapter introduces the phenomena under investigation and discusses the importance of having reliable metrics in subjective sound quality jury testing.

**Chapter 2: Review of Related Research** This chapter delves into previous research related to sound quality assessment, providing a comprehensive background and context for the study.

**Chapter 3: Methodology for Facial Expression Measurement in Jury Testing** Discusses the methodologies utilized to assess the integration of facial expression measurement in jury testing, aimed at enhancing the accuracy and reliability of the test results.

**Chapter 4: Methodology for EEG Measurement in Jury Testing** This chapter outlines the methodologies used to assess the application of EEG measurement in jury testing, focusing on improving the accuracy and reliability of the outcomes.

**Chapter 5: Results and Analysis** Discusses the results of the experiments and analysis, presenting the findings from the application of facial expression and EEG measurements in sound quality jury testing.

**Chapter 6: Conclusion** The final chapter concludes the discussions of the thesis, summarizing the key insights and contributions of the study.

# Chapter 2

# Literature Review

In this chapter, we scrabble around in the existing body of knowledge surrounding Acoustic Sound Quality Jury Testing and explore how emerging technologies, specifically EEG (Electroencephalography) and facial expression analysis can help the processing step and the accuracy of results obtained. The synthesis of relevant literature will serve as the foundation for understanding the current state of research, and afterwards, identifying the potential gaps which will be presented in the last section.

## 2.1 Acoustic Assessment, and Jury Testing

Acoustic Sound Quality Jury Testing has been an essential method for evaluating the perceived quality of audio and sound systems for a long time. Jurors, typically comprising experts or individuals with relevant experience, are tasked with assessing various acoustic properties. This subjective evaluation is crucial for product development, as it reflects how end-users will experience sound. Accordingly, the domain of sound quality assessment is taking considerable attention within the scientific acoustics field. The work by Dedene et al. [23] stands as an example to this interest, where Dedene et al. introduced a parametric model based on multiple regression techniques. Their model tried to forecast the subjective sound quality assessments performed by jury testing, using the objective sound quality metrics. moreover, Bergman et al. [24] argue that auditory pleasantness is mainly governed by the perceived loudness of a sound, whereas arousal is influenced predominantly by its perceived sharpness. Moreover, the insertion of highly pleasant sounds can modulate the pleasantness of noise. In a similar vein, Wang and Subic [25] embarked on an exploration of the sound quality associated with vehicle side mirrors. They attempted to draw parallels between subjective assessments and objective measurements, employing two distinct mathematical constructs. Notably, they focused on several key acoustic quality indicators, such as sound pressure level, roughness, and tonality, as the basis of their correlation algorithms.

Further advancing the field, Ma et al. [26] demonstrated the efficacy of Artificial Neural Networks (ANNs) in the prediction of subjective sound quality metrics from objective ones. Their method stood out, particularly due to its average error rate of just 3.97%, supplemented by an insightful weight analysis.

Complementing these approaches, Wang et al. [27] took the concept of roughness in vehicle interior noise and refined it to better align with human auditory perception. Their proposed metric, the Human Auditory Perception-based Roughness Metric (HAP-RM), showed promising results in deciphering both stationary and transient sound signals.

Lastly, the study by Parizet et al. [28] studied the acoustics of car door closures. They employed psychoacoustic metrics to determine sound quality and concluded that two timbre-related attributes—the frequency balance and the cleanness of the sound emerged as predominant factors in evaluating the auditory impact of this specific noise event.

Critical explorations and advancements have been made in pursuit of bolstering the reliability of jury tests. An exemplary manifestation of this is the work undertaken by Kim et al. [29], who ingeniously employed a decision error model as a strategy to enhance the accuracy and reliability of subjective acoustic evaluations within jury tests. Their study, particularly focused on the acoustic evaluations of laser printers, unveiled that there exists a negative correlation between the likelihood of decision error and the normalized variations discerned in the perceived acoustic stimuli. By implementing the decision error model, they were able to discern insights that are instrumental in deciphering and augmenting the reliability of jurors' responses. Their findings underscore the potential of such models to act as pivotal tools, aiding in the systematic identification and reduction of decision errors, thereby fostering an environment where jury test results are characterized by enhanced objectivity and dependability.

A novel approach that has gained traction is the application of virtual reality (VR). The study by Robotham et al. [30] explored the realm of sound quality assessment within VR platforms. It compares several evaluation methods, including conventional jury testing, paired comparison tests, and rating scales, underlining the necessity to tailor sound quality evaluation methods to the distinct attributes and challenges presented by VR environments.

In addition to the innovative use of VR, the refinement of statistical analysis techniques has markedly improved the reliability and uniformity of results derived from jury testing. [31] underscored the critical role of proper sample size determination and statistical modelling in minimizing variability, thereby bolstering the precision of jury testing outcomes.

Another emerging trend is the integration of wearable devices and physiological monitoring tools in the evaluation of sound quality. Such devices have the capability to track real-time physiological and biometric data, including heart rate, skin conductance, and brain activity, offering a window into the physiological reactions of listeners to auditory stimuli. Correlating these physiological responses with subjective assessments allows researchers to explore more deeply the innate responses and physiological underpinnings related to sound quality perception [32]. These topics will be examined in further detail in subsequent sections.

## 2.2 Facial Expression Analysis in Emotion Detection

An intriguing dimension to the evaluation of sound quality is the study of facial expressions as they naturally respond to different auditory stimuli. Hu et al. [33] conducted a seminal investigation into this phenomenon, where they exploited advanced deep learning techniques for the nuanced analysis of facial expressions. Their findings revealed a high degree of congruence between the emotions detected through facial expression analysis and those reported by subjects via questionnaires. Remarkably, in certain instances, the facial expression analysis proved even more effective than traditional questionnaires. In contrast, Huang et al. [34] took a more generative approach by correlating specific facial expressions with noise annoyance thresholds. Engaging in an extensive social survey involving over seven thousand participants in a Chinese city, they captured the public's noise perception through questionnaires while simultaneously measuring noise levels with an analyzer. Subsequently, using a free-form deformation technique, they constructed facial expressions reflective of the annoyance levels gleaned from the survey data. Hadinejad et al. [35] investigated the emotional impact induced by a tourism advertisement. By recording the facial expressions of individuals from different cultural backgrounds as they viewed the advertisement, and using the FaceReader software for emotion recognition, they could pinpoint a range of emotions, including seven distinct categories alongside valence and arousal, thus showcasing the tool's adeptness at real-time emotion detection. Moreover, the study by Meng et al. [36] explores urban sound perception and its manifestation through facial expressions. Subjects were exposed to a variety of typical urban sounds, including traffic, natural, and community noises. The research juxtaposed findings from a sound perception questionnaire against facial expression analysis, concluding that the latter is a potent tool for investigating sound perception. It was observed that emotions such as happiness, sadness, and surprise were indicative of the participants' reactions to the acoustic stimuli.

The investigation into the interplay between soundscapes and psychophysiological responses has been furthered by the work of Park et al. [37]. Park et al. assessed the influence of various soundscape stimuli on parameters such as heart rate, electrodermal activity, respiratory rate, and facial electromyography. Their laboratory experiments, which were carried out in both VR and non-VR settings, indicated that rural soundscapes facilitated better psychophysiological recovery compared to urban soundscapes. Notably, while VR and non-VR conditions did not yield significant differences in overall recovery, certain physiological responses were notably affected under VR conditions. Further contributing to the field, Özseven [38] studied emotion recognition through the processing of speech spectrogram images. By converting speeches into spectrograms and employing texture analysis methods, emotions were detected and their recognition rates were benchmarked. Using support vector machines, they found that this approach outperformed traditional acoustic analysis in recognizing emotions from speech. Liu et al. [39] utilized FaceReader software to examine the emotional responses elicited in elderly individuals with dementia when exposed to various sounds. Their findings affirmed FaceReader's capability to distinguish between emotional changes corresponding to different auditory stimuli. It was noted that music, in particular, was a powerful emotional trigger, outstripping other sounds such as birdsong and the sound of a stream in its effect. Moreover, Frescura and Lee [40] probed into the emotional responses elicited by common residential noise sources, such as footsteps, speech, and music, within wooden buildings. The study combined the Self-Assessment Manikin with physiological measurements, including facial electromyography of the corrugator supercilii and zygomaticus major muscle groups, heart rate, and electrodermal activity. Their results presented a noteworthy concordance between physiological responses and self-reported emotion assessments.

A multitude of research has been undertaken to gauge the level of user engagement elicited by soundtracks. This body of work advocates for the assessment of both the pleasantness and the semantic associations of products, commencing with the work of Zampini and Spence [41]. They found that sharper sounds can amplify the perceived crispness of potato crisps, thereby enhancing enjoyment. Conversely, the coarse sounds emitted by an epilator may induce fear. [42]. It is therefore imperative, especially in Jury Testing contexts, to scrutinize the pleasantness and arousal associated with products. These constructs not only help dissociate the test context from jurors' preconceived notions but also shed light on instances when jurors may become inattentive during evaluations.

The PAD model (Pleasure, Arousal, Dominance), proposed by Russell and Mehrabian [43], delineates the experiential and communicative facets of affective attributes related to objects, events, and people. Within this framework, pleasure represents the degree of positive or negative valuation, similar to the hedonic tone engendered by one's state. Arousal reflects the intensity and stimulatory impact of an emotion, denoting the level of activity and responsiveness within a given context. Bergman et al. [24] further explicate the interrelation between pleasantness and arousal with respect to auditory experiences. Research by Generosi et al. [44] extends this understanding by implementing facial expression analysis to decipher emotional responses. This approach utilizes an emotion and valence model, with valence representing the inherent attractiveness or averseness of a situation, as outwardly expressed by an individual. This is closely related to the emotional state being conveyed, and engagement epitomizes a composite measure of discrete emotions, such as anger, disgust, fear, happiness, sadness, and surprise, as characterized by Ekman [45]. It signals the degree of a person's active participation in a scenario. Accordingly, valence corresponds to the pleasure component of the PAD model, and engagement parallels arousal.

These findings underscore the significance of exploring the emotional bond in user-product interactions. Notably, valence and engagement emerge as key constituents of this interplay, with observational techniques like facial expression recognition serving as viable methods for the examination [44].

The quest to encapsulate human emotions in an objective and quantifiable form has long fascinated the scientific community. Researchers have explored methods for classifying and categorizing human emotions based on discernible external signals. These signals, which include facial expressions, vocal intonations, and various physiological indicators, are the mediums through which humans, often subconsciously, communicate their emotional states [46]. At the forefront of this research stands Paul Ekman, whose seminal work in 1971 laid the groundwork for the modern understanding of emotions. Ekman postulated that there exists a finite set of fundamental emotions, each associated with a unique and universally recognized facial expression. This concept was encapsulated in what is now referred to as "Ekman's Universal Facial Expressions". The emotions identified by Ekman—anger, disgust, fear, happiness, sadness, and surprise—are expressed similarly around the globe, emphasizing the universal nature of these emotional experiences [45].

Facial Expression Recognition (FER) systems have become a cornerstone in interpreting human emotions, with a profound impact on the development of human-computer interaction technologies. The utilization of deep learning methods, particularly Convolutional Neural Networks (CNNs), has been extensively documented in the literature, illustrating their capacity to process and analyze complex visual data with remarkable accuracy [47]. CNNs have shown exceptional efficacy in classifying the six fundamental emotions identified by Ekman – happiness, surprise, sadness, anger, disgust, and fear [46]. Recent studies have explored the recognition of emotions even under the constraints of personal protective equipment, demonstrating the versatility of FER systems [48]. Moreover, the application of neural networks extends beyond emotion recognition to the prediction of psychological states such as depression, anxiety, and stress by analyzing facial expressions through the Facial Action Coding System (FACS), offering a nonintrusive and real-time solution for mental health monitoring [49]. The continuous refinement of CNN architectures has led to broader applications, enhancing user experience in human-computer interaction, providing valuable customer feedback, aiding in elderly care, and reinforcing law enforcement in smart cities [50]. These advancements underscore the shift towards noninvasive emotion recognition technologies. Unlike biofeedback tools like EEG, ECG, or GSR, which, despite their accuracy, demand extensive setup time, intrusive attachment to the body, and costly equipment, camera-based systems provide a more practical and user-friendly alternative for real-world implementations [51, 52].

## 2.3 EEG as a Tool in Acoustic Evaluation

The pursuit of enhancing sound quality jury testing is a multifaceted challenge that encapsulates the reliability, accuracy, and repeatability of subjective evaluations. The inherent complexity of sound quality necessitates a comprehensive approach that balances human perception with technical rigour [53, 1]. In response to these challenges, there has been a concerted effort within the research community to refine and advance the methodologies used in jury testing.

Further progress in this domain has been marked by the development of standardized test protocols, utilizing objective sound quality metrics, statistical tools, and virtual reality to enhance the reliability of jury testing outcomes [54, 55, 56]. As an example, Zhang et al. [57] have showcased the pivotal role of careful sample size determination and robust statistical modelling. These statistical approaches are crucial in mitigating variability and improving the precision of jury test outcomes, thereby enhancing sound quality metrics. Moreover, the advent of wearable technology and physiological monitoring tools marks a significant leap forward in jury test methodologies. These devices are capable of monitoring and recording real-time biometric data such as heart rate, skin conductance, and brainwave patterns. Such physiological data are invaluable as they offer an objective glimpse into the listener's immediate response to auditory stimuli. When these physiological markers are analyzed in tandem with subjective evaluations, a richer and more detailed picture of sound quality perception emerges. This correlation not only sheds light on the listener's sensory experience but also on the complex interplay between sound and its physiological impact. The integration of these wearable devices in jury testing assessment presents a unique opportunity to bridge the gap between subjective auditory perception and its objective physiological underpinnings, potentially augmenting the reliability and accuracy of the sound quality assessment process [58].

The use of EEG has become increasingly prevalent within the field of acoustic perception, particularly due to its capability to track and record the brain's physiological response to auditory stimuli in real time. The brain's electrical activity is known to fluctuate in response to the perception and cognitive processing of environmental sounds. EEG technology captures these fluctuations by measuring the voltage changes generated by ionic currents flowing within the brain's neurons [59]. The EEG sensors, placed on the scalp following the internationally recognized 10-20 system, detect a spectrum of brainwave frequencies, each associated with different states of consciousness [60]. Budak et al. [61] explored the novel application of EEG by focusing on the detection of driver drowsiness. By processing EEG signals, they were able to differentiate between states of wakefulness and drowsiness in drivers. Their approach involved comparing EEG data from these two states and developing a model to assess drowsiness levels effectively. This study highlights the potential of EEG in enhancing road safety by monitoring drivers' alertness levels. In another intriguing exploration, Zuk et al. [62]

ventured into the classification of natural sounds using EEG signal processing. Their study revealed unique EEG patterns associated with speech and music, demonstrating the distinct brain responses elicited by different types of auditory stimuli. This research underscores the versatility of EEG in distinguishing between various acoustic experiences.

The exploration of human brain physiology in response to external stimuli, particularly visual and acoustic, has yielded profound insights into the simultaneous occurrence of brain activities and external stimulations [63, 64]. Trimmel's work [65] has highlighted that exposure to noise can significantly alter the central nervous system's activity. The extent of this impact varies depending on the type of sound and its Sound Pressure Level (SPL). This finding underscores the intricate relationship between auditory stimuli and the physiological responses they evoke.

Further investigations into EEG power spectral densities (PSDs) have revealed a direct correlation with individual acoustic perceptions. Among these, some studies explored the interplay between emotional states and EEG responses elicited by acoustic stimuli. Notably, the research conducted by Schmidt and Trainor [66] presented a fascinating observation. Their study involved playing musical excerpts characterized as happy or sad and examining the resultant EEG activity. They discovered a decrease in alpha power at the left frontal electrodes during exposure to happy music, while sad music was linked to a more pronounced decrease in alpha power at the right frontal leads. Kabuto et al. [67] expanded on this by analyzing the changes in PSDs induced by pleasant music. They found that alpha power is heightened in states described as "pleasant" and "calm," suggesting a strong association between increased alpha power and a state of relaxation.

Moreover, Lee et al. [68] studied the relationship between subjective sound quality ratings and EEG responses. In addition to establishing a correlation between subjective ratings and sound metrics, they discovered a significant association between these ratings and the brain's EEG responses. This finding is particularly significant as it bridges the gap between subjective auditory perceptions and objective physiological measures, offering a comprehensive approach to understanding and evaluating sound quality.

Additionally, Walker [69] reported findings that establish a connection between self-reported inattentiveness to music and the production of high theta- and high delta-waves. This correlation provides an intriguing insight into how the brain's electrical activity can reflect varying levels of engagement with auditory stimuli, even in the absence of conscious attention.

These studies collectively contribute to a more nuanced understanding of how EEG can serve as a potent tool for assessing the impact of sound on human brain activity and emotional states. This intersection of acoustic stimuli and EEG responses offers invaluable perspectives for advancing sound quality metrics and enhancing auditory experiences.

## 2.4 Research Gaps

In the previously mentioned research on enhancing sound quality through jury testing, an important aspect that has often been overlooked is the degree of involvement of the jurors. Considering the level of engagement of jurors is crucial because it allows a more accurate assessment of the reliability of the answers provided by the jurors. By giving more weight to the responses provided by highly involved jurors and discounting or excluding responses from distracted jurors, the overall quality of the evaluation can be improved.

Furthermore, the existing research on sound quality assessments has not explored the potential of utilizing physiological and biometric measurements as a valuable tool for predicting the outcomes of subjective sound quality jury testing and assisting the moderators to get an insight into the level of involvement of the jurors during the sound quality assessment. These supplementary measurements can provide additional insights into individuals' emotional and perceptual responses to sound stimuli that can be correlated to human preference and perception, rather than potentially biased surveys-only data sources coming from traditional jury testing processes. For example, by incorporating EEG measurement alongside the conventional jury testing approach, a complementary and corroborative assessment of sound quality can be achieved. This integration of prediction results derived from EEG and facial expression analysis with the outcomes of jury testing can offer a dual assurance mechanism, strengthening the reliability and validity of the overall evaluation process.

Considering the degree of juror involvement and integrating physiological and biometric measurements into subjective sound quality assessments, allows us to enhance the accuracy and robustness of jury testing evaluations. This approach provides a more comprehensive understanding of individuals' subjective experiences and preferences regarding sound quality, leading to more reliable and informed decisions in various audio-related applications and industries.

# Chapter 3

# Jury Test Combined with Facial Expression Measurement

In this Chapter, we aim to explore the practicality of including assessments of facial expressions to ensure consistent jurors' feedback. The underlying idea here is that by adding facial expression recognition, we might reduce the personal opinions that can vary among jurors and lessen the need for intervention by a moderator. This study focused on examining the various sounds in car noise, recorded under different conditions for research purposes. Using facial expression recognition methods could help counteract biases and personal opinions in evaluations, since facial expressions are often natural and automatic. This approach evaluates the possibility of having a more objective measure of jurors' emotional responses. Also, it's important to consider the context in which the responses are given [58][70].

This chapter of the thesis is dedicated to evaluating the feasibility of incorporating facial expression analysis into sound quality jury testing. To thoroughly examine this integration, two preliminary tests were conducted prior to the main jury test campaign.

The following sections will delve into the specifics of these preliminary tests, outlining their design, execution, and the rationale behind them. These tests were crucial in establishing a foundational understanding of how facial expressions correlate with auditory stimuli in the context of sound quality evaluation.

Additionally, a comprehensive discussion on the jury test campaign, which combined traditional sound quality assessment with facial expression analysis, will be presented. This campaign was instrumental in exploring the potential synergies and insights gained from this multidimensional approach.

## 3.1 Theoretical Framework: Emotion and Facial Expression

Facial expression analysis merges the disciplines of psychology, biometry, and computer science. The core principle of facial expression analysis lies in the interpretation of physical manifestations of emotions on the face. This involves identifying specific muscle movements and categorizing them into recognizable emotions. Historically, the study of facial expressions as a means of conveying emotions dates back to Charles Darwin, who first suggested their evolutionary importance. This field gained substantial traction with Paul Ekman's pioneering research, which identified universal emotions expressed through distinct facial expressions [71, 45].

To assess emotion factors, one of the most promising approaches in this domain involves the utilization of Deep Learning-driven software, specifically leveraging models designed for recognizing facial expressions. While techniques demanding direct user input tend to introduce significant subjectivity and bias, the analysis of facial parameters offers an objective avenue for rating the degree of user appreciation.

In the context of this research, a tool founded on a Convolutional Neural Network (CNN), developed using the Python programming language and harnessed through the Keras and TensorFlow frameworks, has been employed [72]. The primary function of this CNN is to perform the recognition of Ekman's six universal emotions—namely, happiness, surprise, sadness, anger, fear, and disgust—using facial images as input. The output generated by the network consists of a percentage-based probability assigned to each of these emotions.

Ekman's emotion scores, predicted from every camera video frame, are normalized to a percentage value of 100. Valence can then be calculated at each moment by distinguishing between positive and negative emotions according to Ekman. Positive emotions are generally associated with expressions like happiness, joy, contentment, and love, which tend to be characterized by features such as smiling, bright eyes, and a relaxed facial posture. On the other hand, negative emotions include sadness, anger, fear, and disgust, often reflected in facial expressions through frowning, furrowed brows, narrowed eyes, and tense mouth. As a result, valence ranges from -100 to 100, indicating the total positivity or negativity expressed by the participants [73].

To enhance the dependability of the assessment regarding valence and engagement, a facial expression recognition system has been integrated with software capable of discerning the user's attentiveness during the evaluation. This system combines several parameters, including the degree of head rotation in relation to the camera (positioned on top of the screen), the direction of gaze, and the proportion of eyelid aperture. The underlying assumption is that instances where the test subject displays inattentiveness might lead to less reliable emotional data derived from the facial expression recognition CNN. This could be attributed to emotions expressed through facial expressions that may have been triggered by events or memories unrelated to the ongoing evaluation

[74]. In order to gauge the level of attention, the third-party library Dlib has been exploited. This library facilitates the extraction of a comprehensive mapping of the user's facial features, enabling a comprehensive assessment of their attention status [2].

In particular, the distances between six couples of points (i.e, 17-31, 1-31, 13-55, 5-49, 17-27, 1-18 have been considered to estimate the head orientation with respect to the screen figure 3.1. In fact, based on the usual symmetric characteristic of a face, such distances should remain constant between the left and right sides. Consequently, the division ratio between the distances constructed in the left part of the face and those in the right part should remain in the neighborhood of 1. If the calculated ratio remains in such a neighborhood the user is considered to be attentive. The more the ratio is far from the desired value the lower the user attention level will be. Also in this case a threshold is defined; if the user's attention level is below that threshold the user is considered to be distracted [75].

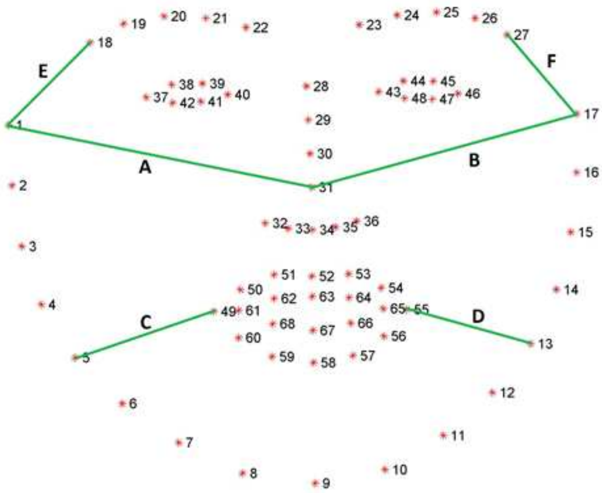



**Figure 3.1:** Dlib facial landmarks used for attention evaluation [2]

## 3.2 Device Specifications

The measurement campaigns employed a variety of tools and software to facilitate the collection, management, and analysis of data.

**Face Recording Setup:** For capturing video streams, the Windows Camera app was utilized, storing the footage taken by a Logitech HD C925E webcam connected to a PC. This setup ensured high-quality video capture, essential for accurate analysis (refer to Figure 3.2a).

**Audio/Video Stimuli Management:** A software developed in Matlab played a crucial role in managing and casting the audio/video stimuli. It also handled the synchronization of these stimuli with the recording software, ensuring precise timing across all data streams.

**Video Analysis Software:** The analysis of the recorded videos was conducted using software developed in Python, as described in paragraph 3. This software was instrumental in processing the video data to extract meaningful insights.

**Data management and Formatting:** The outputs from the video analysis, including the probability percentages associated with each frame for Ekman's emotions, along with timestamps and attention values (boolean true/false), were stored in CSV format. This format facilitated easy access and manipulation of the data for further analysis.

**Computational Resources:** The image processing tasks were executed on a desktop PC equipped with an Intel i9 10900 CPU, Nvidia Quadro P2200 video card, and 128GB of RAM. This high-performance setup was critical for handling the intensive computational demands of the image processing algorithms.

**Display:** To maximize the immersiveness of the subjects in the test, a ProjectionDesign F10 AS3D ZOOM projector, laptop screen, and a 49-inch Samsung Odyssey G9 curved screen were used in order to display visual content, respectively, for the first and second preliminary test and the jury test campaign. The high-quality displays were chosen for thier ability to deliver clear and engaging visuals, thereby enhancing the overall experience and effectiveness of the test.

**Headphone:** For the efficient playback of sounds to the participants during the test, Reference studio headphones, specifically the AKG K702 model (Figure 3.2b), were utilized. The AKG K702 headphones, belonging to the renowned AKG company, were selected for their high-quality sound reproduction capabilities. These over-ear headphones are known for their comfort and sound fidelity, making them an ideal choice for tests requiring precise and clear audio delivery. Key specifications of the AKG K702 headphones that contributed to their selection include a rated impedance of 60 ohms and a sensitivity of 105 dB SPL/V @ 1 kHz. These technical features ensure that the headphones provide accurate and consistent sound levels, which is crucial for maintaining the integrity of the auditory stimuli presented to the participants.


(a) (b)

**Figure 3.2:** (a): face recording by a Logitech HD C925E webcam. (b): AKG K702 headphone to playback of sounds to the participants

## 3.3 Preliminary test 1: Stimulating Facial Expression by Video/Sounds

The primary objective of this experiment was to investigate whether variation in auditory and visual stimuli can influence the emotional reactions of participants detected by facial expression analysis. To this end, a comprehensive test has been devised in which participants are exposed to a variety of auditory and visual stimuli.

### 3.3.1 Participants' Demographics

The experimentation took place at Università Politecnica delle Marche, where the test participants primarily consisted of diverse individuals, including students and professionals from varying age groups associated with the university. Their involvement encompassed exposure to a variety of acoustic and visual stimuli while a camera, positioned in front of their faces, captured their facial expressions. The participants contained a total of 16 subjects. The gender distribution among the participants was nearly equal, maintaining a balance between males and females. The average age of subjects was 25 years.

### 3.3.2 Design and Procedure

The test was designed with the specific goal of assessing variations in facial expressions in response to a range of video and sound stimuli.

### Environment Setup

For this test, a carefully designed environment was prepared to ensure that the participants' concentration was fully focused on the content presented.

The test environment was deliberately set up to be a dimly lit room, creating a conducive atmosphere for focus and minimizing distractions. Ambient noise was significantly reduced, which was crucial for the participants' concentration. This was achieved by providing the participants with headphones, which also served to isolate them from any external noise.

To further enhance the participants' focus and to create an enclosed environment, all external stimuli and sources of potential distraction were eliminated. The environment was designed to be as controlled and isolated as possible, ensuring that the participants' concentration remained solely on the test.

### Test Procedure

Upon arrival at the testing site, subjects underwent a process of acclimatization to the environment and received comprehensive instructions regarding the test protocols.

The initial step involved acquainting each participant with the testing environment. This was crucial for ensuring that subjects felt comfortable and at ease, which could significantly influence their responses and the overall quality of the data collected.

Additionally, clear and detailed instructions regarding the test procedures were provided to all participants. This briefing was essential to ensure that the subjects understood their roles and the test's objectives. The importance of following the protocols accurately was emphasized to maintain the integrity of the experiment.

Furthermore, an Informed Consent Form was presented to each participant. This form contained all necessary information about the experiment, including its purpose, potential risks, and benefits. Subjects were required to read and sign this form, confirming that they were participating voluntarily, fully informed, and aware of what their involvement entailed.

Figure 3.3 illustrates the test environment, providing a visual representation of the setting in which the subjects participated in the experiment.

### Design of the Test

A carefully designed test was implemented in which participants were exposed to a variety of auditory and video stimuli. The objective of this test was to elicit a range of emotions in the participants, corresponding to those identified in Ekman's expressions as described by the Facial Action Coding System (FACS) method.

The stimuli used in the test were categorized into three macro-categories to facilitate a structured and comprehensive emotional response analysis. These categories were:

- **Sound:** This category, depicted in Figure 3.4a, consisted solely of auditory stimuli designed to evoke specific emotional responses based on their characteristics and content.

- **Sound with Video:** Represented in Figure 3.4b, this category included stimuli that combined both auditory and visual elements, providing a more immersive experience to potentially elicit a broader range of emotional reactions.

- **Validated Video:** The third category, shown in Figure 3.4c, comprised of validated video stimuli, specifically selected for their proven effectiveness in evoking certain emotional responses in line with Ekman's expressions [76].

### 3.3.3 Sequence

The test is structured as follows: in the first phase (Figure 3.4a), three 30-second samples of various noises are interspersed with identical 40-second soothing sounds of a tranquil spring river. This is intended to restore the participants' emotional state to a neutral baseline.

The subsequent phase (Figure 3.4b) aims to compare the emotional responses evoked by sound and video stimuli. The sequence from the initial phase (Figure 3.4a) is repeated, replacing the auditory stimuli with corresponding videos. This entails exposing participants to videos of a blackboard, clapping, and mosquito noises interleaved with the same relaxing video as before.

The final phase (Figure 3.4c) includes six different videos, each lasting 10 seconds, with the relaxing 40-second video again interspersed between them. These videos are selected from a validated dataset tailored to induce emotional responses [76].

To ensure an environment conducive to capturing participants' attention and focusing it on the screen's content, a purpose-designed setting has been arranged (Figure 3.3). This controlled setup features a dimly lit laboratory equipped with a large screen projector. To minimize ambient noise, participants are provided with headphones. Moreover, the environment is enclosed to eliminate external stimuli or potential sources of distraction.

**Figure 3.3:** View of the performing the preliminary tests 1

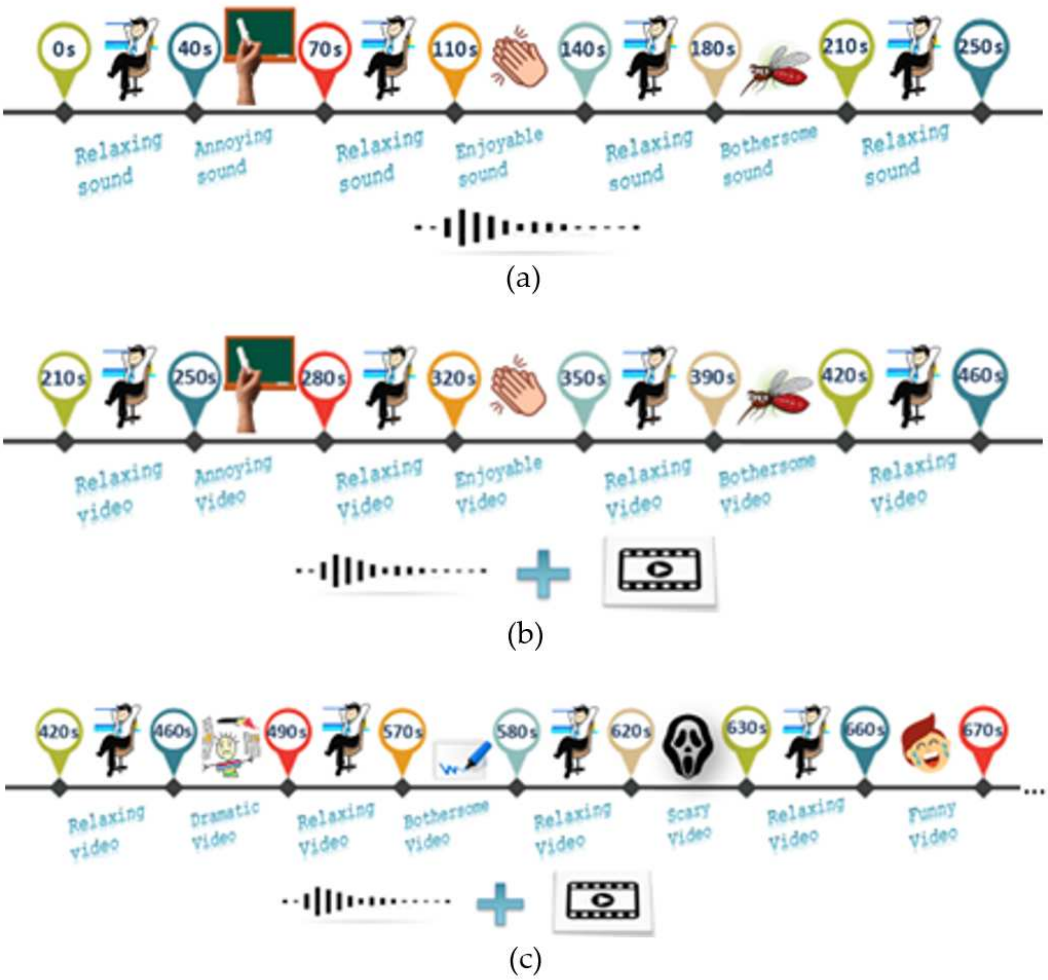



**Figure 3.4:** Sequence of the first preliminary test. (a): sequence of audio, between [0 250] seconds. (b): sequence of audio plus video, between [210 460] seconds. (c): sequence of validated video, from 420 seconds until the end

## 3.4 Preliminary Test 2: Emotion Response to Acoustic Stimuli

This section provides a detailed overview of the test protocols conducted during the campaign aimed at assessing variations in facial expressions in response to different acoustic stimuli. A diverse group of participants, including students and professionals from various age ranges, took part in the research. This study was conducted at the Università Politecnica delle Marche, with all participants having an affiliation with the university, thus ensuring a wide-ranging demographic representation.

The study included a total of 40 subjects, with an almost equal distribution of males and females, thereby maintaining gender balance. The average age of the participants was calculated to be 30.8 years.

The central component of the test involved exposing participants to a blend of acoustic and visual stimuli. A camera, strategically positioned in front of the participants, was tasked with capturing the nuances of their facial expressions in response to these stimuli (refer to Figure 3.5).

### 3.4.1 Objective of the Test Campaign

The primary aim of this test campaign was to scrutinize how different acoustic stimuli impact facial expression variations. To achieve this, participants were presented with three distinct auditory stimuli, alongside a carefully crafted questionnaire. This approach was designed to probe the correlation between auditory stimuli and facial expression responses.

### 3.4.2 Test Sequence

As illustrated in Figure 3.6, the test sequence was structured in which participants were subjected to three varied soundtracks, each representing different acoustic stimulation. Specifically, three types of soundtracks were employed engine noise, relaxing sound, and exterior car noise. The procedure involved a 60-second silent interval preceding each soundtrack to establish a consistent auditory baseline for the subjects. Following the playback of each 60-second soundtrack, an answering session was conducted where participants responded to a set of questions. This structure ensured a controlled and uniform testing environment.

### 3.4.3 Acoustic Perception Feedback

After exposure to each sound, participants were asked a series of three questions. These inquiries were crafted to evaluate the subjects' perceptions of each sound in terms of annoyance or pleasantness, relaxation or stress, and quietness or loudness. The purpose of these questions was to gain insights into the subjective auditory experience of each participant, thereby linking acoustic stimuli with observed facial expressions. The data obtained from these responses were integral in understanding the emotional and psychological impact of different acoustic environments on individuals.

**Figure 3.5:** View of the performing the preliminary tests 2

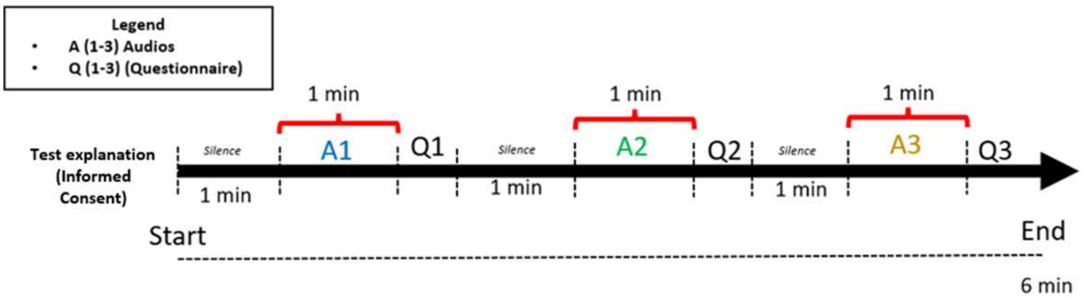



**Figure 3.6:** Test Sequence for the second preliminary test

## 3.5 Experimental Study of Combining Facial Expression with Traditional Jury Test

The subjective sound quality jury test campaign was conducted to investigate the perception of various interior car sounds. The experiment utilized facial expression analysis to gauge reactions and assess the impact of different auditory stimuli. The pool of the jurors consisted of 40 individuals in a diverse range of genders and ages, specifically, 21 females and 19 males between the ages from 19 to 54 years. Notably, a substantial 75% of the jurors fell within the age of $28 \pm 6$ years, ensuring a youthful perspective in the assessment process.

### 3.5.1 Test Environment and Procedure

The Virtual Reality laboratory at Università Politecnica delle Marche, Italy, facilitated the experiment in June 2023. Sessions were structured in two distinct slots: morning (9 AM to 1 PM) and afternoon (2:30 PM to 6:30 PM), to accommodate jurors' availability and maintain a consistent testing environment.

Jurors, upon arrival, were acclimated to the testing environment and furnished with clear instructions regarding the test protocols and their roles. This also included reading carefully and signing an Informed Consent Form, ensuring that all participants were well-informed, willing to partake in the experiment voluntarily, and fully aware of what they are agreeing to, including potential risks and benefits.

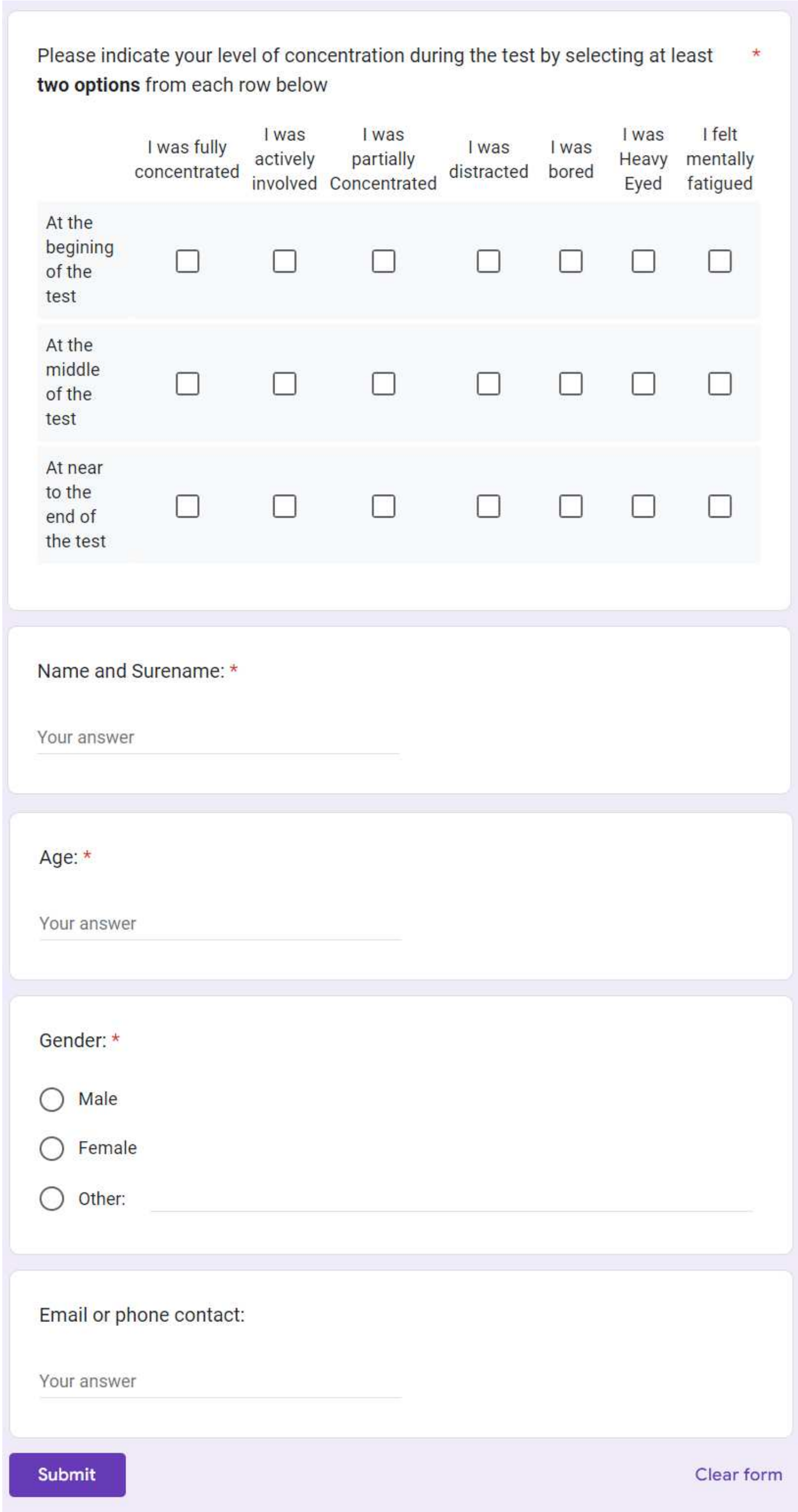
### 3.5.2 Methodology

Employing an AB comparison approach augmented by AB BA repetition, the study aimed to analyze jurors' subjective assessments of seven different interior car sounds. Jurors interacted with a user-friendly interface developed in Matlab. This design consideration aimed to enhance the jurors' comfort and concentration during the experiment, ensuring that their responses were as accurate and reflective as possible to the comparison queries.

In tandem with the sound evaluations, the experiment incorporated continuous facial expression recordings, providing real-time emotional metrics to enhance the depth of the analytical findings. Jurors were advised to maintain a relaxed demeanor, minimizing extraneous movements and focusing on the acoustic assessments, which is indeed necessary for the facial expression assessment as well. Figure 3.8 is depicting a view of the jury test campaign, during which a juror is performing the sound quality assessment.

The main objective of the experiment was to assess jurors' acoustic perceptions and differentiate them under various sound exposures. Two primary questions guided the comparison: "Which sound is more annoying?" and "Which one appears to be from a higher quality car?" Through this approach, the study aimed to uncover insightful data regarding the subjective acoustic quality of different car interiors.

Subsequent to the jury testing, jurors completed a questionnaire, accessed via a QR code. This questionnaire aimed at gauging the jurors' focus and involvement throughout the experiment, acting as a reflective feedback tool. Figure 3.7 represents the questionnaire.



Please indicate your level of concentration during the test by selecting at least **two options** from each row below *

| | I was fully concentrated | I was actively involved | I was partially Concentrated | I was distracted | I was bored | I was Heavy Eyed | I felt mentally fatigued |
|---|---|---|---|---|---|---|---|
| At the begining of the test | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| At the middle of the test | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| At near to the end of the test | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

Name and Surename: *

Your answer

Age: *

Your answer

Gender: *

- Male
- Female
- Other:

Email or phone contact:

Your answer

Submit

Clear form

**Figure 3.7:** Questionnaire capturing the jurors' self-declaration of their involvement level performed in Google Forms. The jurors filled out this questionnaire at the end of the test

### 3.5.3 Jury Testing Process and Sequence

This section outlines the comprehensive procedure adopted for the jury testing, which involved a total of seven distinct sounds as testing material. These sounds were specifically chosen to represent a variety of interior and exterior vehicle noises, offering a broad spectrum of auditory experiences.

The sounds selected for the jury testing included:

- One exterior noise sample from an electric vehicle.
- Three exterior noise samples from internal combustion engine vehicles.
- Three interior noise samples from internal combustion engine vehicles.

For the purpose of conducting an AB comparison jury test, the sounds were paired into 21 unique sound couples, ensuring that each sound was compared against every other. To reinforce the robustness of the test, an additional 21 sound couples were included, inverting the initial pairings. This resulted in a total of 42 sound sequences, which were then shuffled to maintain an unbiased testing environment.

On the other hand, to familiarize the jurors with the test mechanism and expectations, a primary trial part was incorporated at the commencement of the test. This trial segment comprised only one sequence, which entailed a single sound couple comparison. This initial step was instrumental in acclimating the jurors to the test environment and process, ensuring they had a clear understanding of how to engage with and respond to the comparisons effectively. The inclusion of this preliminary trial sequence was crucial in promoting a smooth transition for jurors into the main test.

The jury test was structured into several distinct stages:

1. **Initial Instruction Stage:** Jurors were first provided with instructions and briefed on the test's objectives. They were also required to read and sign an Informed Consent Form, ensuring they were fully aware of the test's nature and their role.

2. **Trial Part:** This stage allowed jurors to familiarize themselves with the test in a practical sense. It was also an opportunity for them to ask any questions and clarify any doubts about the process.

3. **Main Test Stage:** During this stage, jurors were presented with the 42 sound couples in sequence. They were instructed to focus intently on the test, minimizing movements and refraining from speaking to ensure the accuracy of their responses.

4. **Questionnaire Completion:** The final stage involved the jurors filling out a questionnaire, which served to gather their subjective assessments and feedback on the sound samples.

### 3.5.4 Interface app and input device

A user interface app has been developed, in Matlab App Designer platform, to facilitate the jurors in performing the experiment, offering graphical instructions and an interactive platform. Figure 3.9 represents a screenshot of the App used for the test.

During the experiment, the app automated the presentation of sound pairs, termed as A and B, to the jurors. The app was instrumental in managing and collecting the jurors' responses to comparison questions by an external Bluetooth mini keyboard which is presented in Figure 3.10. The external Keyboard was used as the primary device for jurors to input their responses. The decision to use this device stemmed from its convenience advantages. As it is depicted in Figure 3.8, jurors could easily hold the keyboard in their hands, allowing for a more natural and relaxed posture during the experiment. This setup minimized eye or hand movements, enabling jurors to focus intently on the sounds and respond effortlessly by pressing the keyboard buttons corresponding to their choices. The prevention of eye and head movements strengthens the accuracy of facial expression analysis as well, since it can be slightly affected by fast movements.

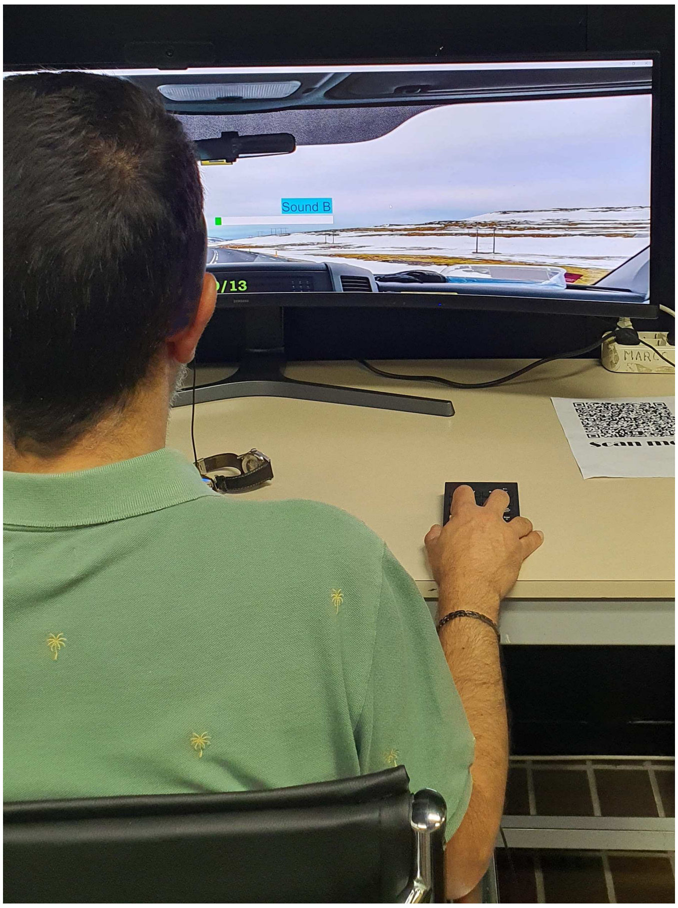



**Figure 3.8:** Experimental setup overview of sound quality jury assessment test


**Figure 3.9:** A screenshot of the graphical user interface used to interact with the jurors


**Figure 3.10:** Using the external mini keyboard used to collect input answers from the jurors

# Chapter 4

# Jury Test Combined with Physiological Response

## 4.1 Theoretical Framework

In the 21$^{st}$ century, a major challenge in brain science is developing mathematical models to understand and predict how brain activities relate to EEG (electroencephalography) signals. EEG is a method that captures the brain's electrical activity, providing valuable insights into various health conditions and cognitive processes. This technology has diverse applications, such as detecting seizures, diagnosing epilepsy, identifying abnormal EEG patterns, recognizing brain activity related to Alzheimer's disease, detecting consciousness levels, and facilitating Brain-Computer Interfaces (BCI) [77].

### 4.1.1 Standard EEG measurement

EEG works by monitoring the electrical activity in the brain. This electrical activity is complex but holds key information about how we think and process information. To measure EEG, electrodes are placed on the scalp. These electrodes detect differences in electrical potential caused mainly by the synchronized activity of certain brain cells (pyramidal neurons) arranged in cortical columns. These potential differences give us a glimpse into the brain's intricate workings. Figure 4.1 provides a simplified illustration of how EEG electrodes detect the brain's electrical activity [78].

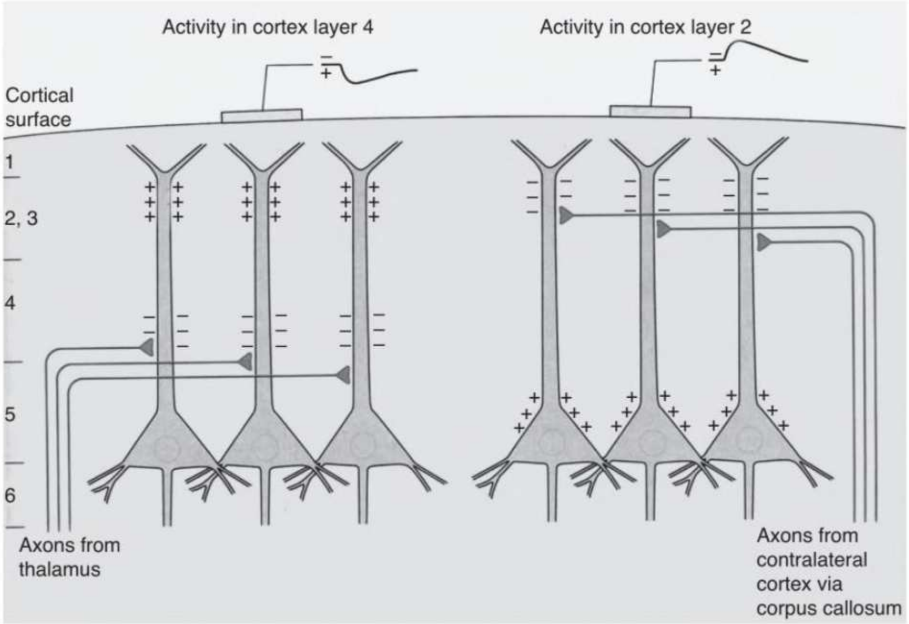



**Figure 4.1:** Generation of extracellular voltage fields from graded synaptic activity. Relationship between polarity of surface potentials and site of dendritic postsynaptic potentials [3].

Neurotransmitters stimulate neurons, opening sodium channels on the neuron's postsynaptic dendrites. This action allows positive $Na^+$ ions to flow in, creating a negative charge near the dendrites compared to other areas of the neuron. This charge difference forms an electric dipole field. When neurotransmitters inhibit these dendrites, the dipole's polarity, and consequently, the polarity at the EEG electrode, reverses. The electric field created travels to the EEG electrode through two processes: volume conduction in the brain's fluid and capacitive conduction through biological tissues and the skull. However, this journey weakens the field significantly. To detect a measurable signal at the electrode, a large number of neurons (around $10^8$) must be activated simultaneously so their electric fields can combine [3].

EEG signals are crucial in both clinical and research applications, offering deep insights into brain activity. Captured using sensors on the scalp, they measure electrical activity from large neuron groups in the brain. Analyzing these signals helps in understanding cognitive and emotional states, monitor alertness, study chronic conditions, and even develop biofeedback and assistive technologies. A major benefit of EEG is its ability to provide multi-dimensional information. By examining signals in time, frequency, or spatial domains, we can decode complex neural patterns and understand brain activities from various angles. EEG's rapid signal capture also allows for real-time brain function analysis. Practically, EEG is advantageous because it's reliable, portable, non-invasive, affordable, and accessible. These qualities make it widely used in both research and healthcare [6].

### 4.1.2 EEG Recording Devices

EEG recording devices, vital in neuroscience and clinical medicine, are designed to capture the brain's electrical activity. These devices use electrodes placed on the scalp to detect signals from brain neurons. These electrodes are linked to an amplifier, which strengthens and filters the signals for accurate recording. The signals are then digitized and stored for analysis and interpretation. EEG devices vary in design, from traditional wired systems with multiple electrodes to more modern, portable, and wireless versions, offering ease of use and flexibility. These devices are key in diagnosing and monitoring neurological conditions like epilepsy, sleep disorders, and brain injuries. They also serve as important tools in research, enabling the study of brain activity and the exploration of cognitive processes and disorders. EEG recording devices have undergone significant advancements, leading to improvements in signal quality, patient comfort, and data analysis capabilities. These advancements have solidified their role as essential tools in both neuroscience research and clinical applications [6].

#### Classification of EEG Devices Based on communication type

EEG headsets, for data transmission, come in two primary forms: wired and wireless, each with distinct features and implications for data integrity and user experience.

- **Wired Communications:** Wired EEG headsets use cables to connect to a computer, providing stable data transmission and the ability to handle large volumes of data effectively. However, it's crucial to acknowledge that the movement of cables and electrodes might introduce artifacts into the EEG signals. This is mainly due to the disturbance in electrode-skin connections, potentially affecting the accuracy of the data [79].

- **Wireless Communications:** In contrast, wireless EEG headsets utilize technologies like Wi-Fi or Bluetooth for connectivity. This design grants users greater mobility, reducing constraints imposed by wired connections. Nevertheless, a significant limitation of wireless systems is the risk of connectivity loss during data acquisition. Such interruptions can lead to gaps in data recording, posing challenges for consistent and reliable data capture [79].

#### Classification of EEG Devices Based on Electrode Type

The classification of EEG recording devices can also be based on the type of electrode used for scalp connection. Each type presents unique characteristics and application methods which are described in the following [4].

- **Soft Gel-based Electrodes:** These electrodes require the application of a conductive gel to establish a connection with the scalp. Post-experiment, it is imperative to clean the headset by removing the gel and thoroughly cleansing the electrodes, typically using alcohol for its evaporative properties.

- **Saline Solution Electrodes:** Some EEG headsets use electrodes that are connected by applying a saline solution. This method is employed to facilitate a low-impedance electrical connection between the skin and the sensor electrode, providing an alternative to conductive gel.

- **Dry Electrodes:** Dry EEG devices eliminate the need for conductive gel or saline, thereby streamlining the EEG data recording process. These devices do not require a trained technician for setup, and they offer a significantly reduced setup time compared to wet headsets.

- **Other Types:** There are EEG devices that utilize unique connection methods, not fitting squarely into the aforementioned categories. For instance, some use conductive solid gel materials to establish electrode connections in EEG recording.

EEG recording devices vary significantly in their design and functionality, especially concerning the types of electrodes they use. Each type of electrode offers distinct advantages and may be suited to different applications.

Figure 4.2 represents Three samples of three primary categories of EEG recording devices based on their electrode types.


(a) (b) (c)

**Figure 4.2:** A view of different EEG recording devices with different electrode types. (a) dry, (b) saline solution, (c) gel-based [4]

### 4.1.3 Standardized Systems for EEG Electrode Placement

Accurate electrode placement is vital for obtaining reliable and meaningful EEG data. It ensures comprehensive coverage of brain regions while minimizing potential signal distortions or artifacts.

Standardized electrode placements, such as the 10-20, 10-10, and 10-5 Systems, facilitate data comparison and exchange across different studies. They enable precise localization of brain activities and provide a uniform framework for EEG researchers and clinicians [80].

- **The International 10-20 System** The 10-20 System is grounded in anatomical landmarks and precise measurements, ensuring systematic and reproducible electrode positioning across different participants and sessions. It categorizes the scalp into distinct regions and assigns electrode positions based on calculated percentages of distances between different electrodes. The labeling of electrodes, such as Fp (frontopolar), F (frontal), C (central), P (parietal), and O (occipital), coupled with numbers, provides a clear indication of their location in relation to the left/right hemisphere and specific areas within each region (refer to Figure 4.3).

- **The international 10-10 System** As an extension of the 10-20 System, the 10-10 System offers an increased number of electrode positions. This results in a higher spatial resolution, which can be particularly beneficial in detailed brain mapping studies (refer to Figure 4.3). Owing to its enhanced accuracy in identifying specific brain regions, the 10-10 System has gained popularity in both research and clinical domains.

- **The international 10-5 System** The 10-5 System further elevates the electrode density by adding even more positions between those of the 10-10 System. This provides an even finer spatial resolution. It is especially beneficial in high-density EEG studies, such as source localization and connectivity analysis, the 10-5 System allows for a more detailed mapping of brain activity (refer to Figure 4.4).

### 4.1.4 Characterization and Analysis of EEG signals

The analysis of EEG data involves categorizing rhythmic activities and transients into various frequency bands, a process integral to understanding the underlying neural processes and mental states.

- **Categorization of EEG Frequency Bands:** While the division into specific bands like alpha (8–12 Hz) can be somewhat subjective, these categorizations are based on observed scalp distribution patterns and biological implications. Spectral analysis methods, such as Welch's method, are commonly used to extract these frequency bands. Software like EEGLAB or the Neurophysiological Biomarker Toolbox facilitate this process, allowing for detailed analysis of EEG data.

- **Quantitative Electroencephalography (qEEG):** qEEG refers to the computational analysis of EEG data, focusing on voltage fluctuations associated with

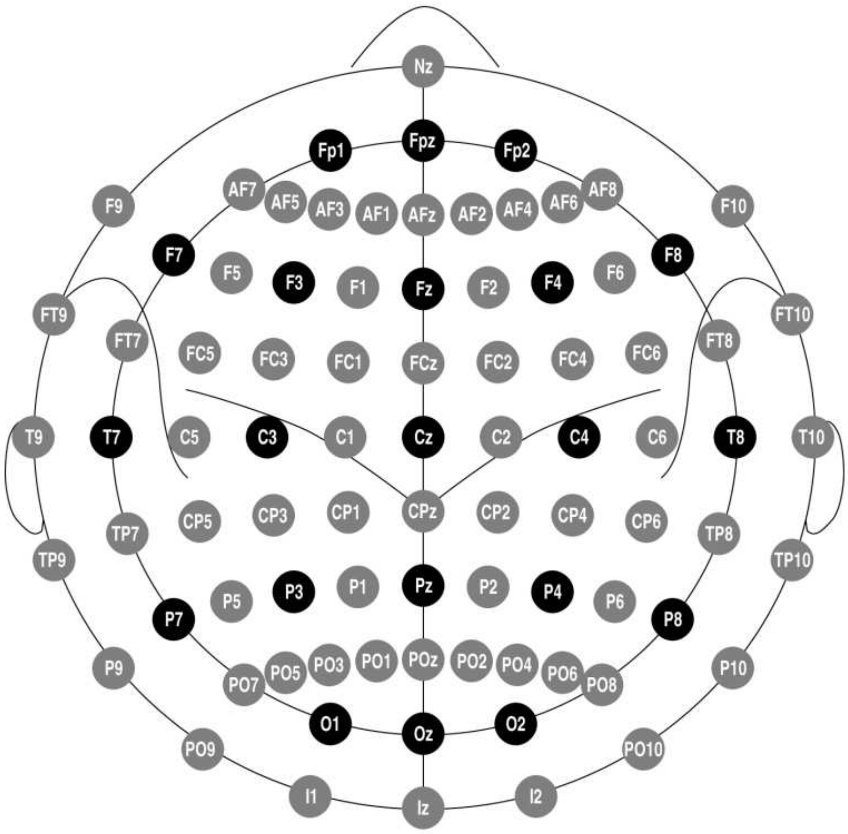



**Figure 4.3:** Electrode positions and labels in the 10-20 and 10-10 system. Black circles indicate positions of the original 10-20 system, gray circles indicate additional positions introduced in the 10-10 extension [5]

different mental states, internal conditions, or pathological disorders. Research indicates a strong correlation between distinct cognitive processes and specific frequency domains.

These waves, generated by synchronized neural activities, vary depending on the participant's internal state. Analysis reveals harmonic frequencies ranging from 1 to 100 Hz, with most informative content lying below 45 Hz. Frequencies beyond this range are often considered artifacts. In standard clinical recordings, EEG waveforms are categorized into alpha, beta, theta, delta, and gamma bandwidths, each playing a crucial role in clinical practice and research and are explained in the Table 4.1.

In the following a detail explanation is performed for the EEG signal bandwidths.

**Delta Waves (0.1 Hz – 4 Hz):** In adults, delta waves are primarily observed during slow-wave sleep. Conversely, in infants, these waves are commonly present, reflecting developmental stages of the brain. The prominence of delta waves in different brain regions varies with age. In adults, they are usually most noticeable in the frontal region, while in children, a more pronounced presence is observed in the posterior region. The occurrence of focal delta waves may indicate the presence of subcortical lesions. In contrast, the widespread distribution of delta waves

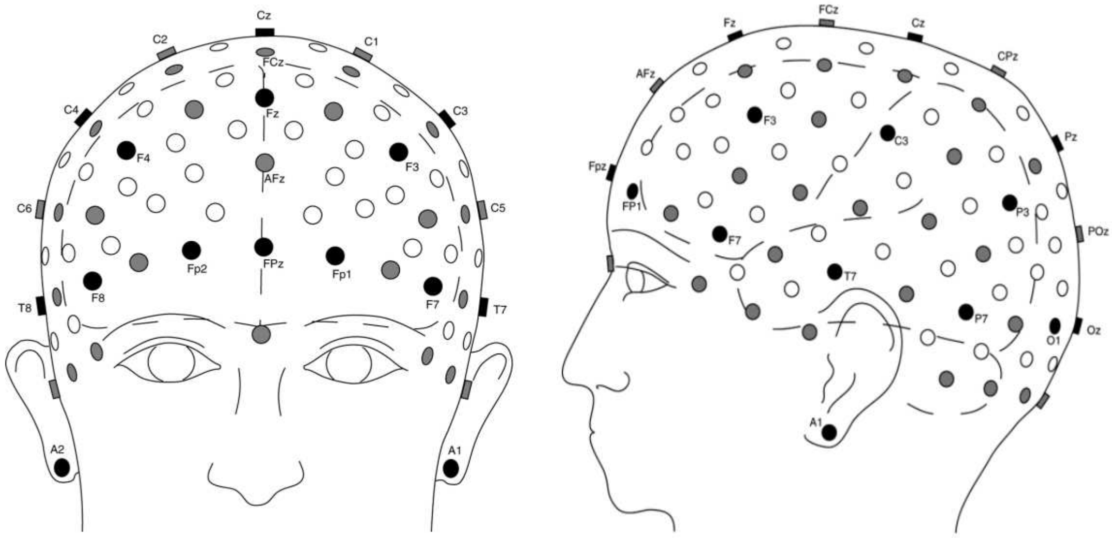



**Figure 4.4:** Selection of 100-5 electrode positions in a realistic display [5]

can be associated with conditions like diffuse lesions, metabolic encephalopathy, hydrocephalus, or deep midline lesions.

**Theta Waves (4 Hz – 7.5 Hz):** In young children, the presence of theta waves is considered normal. In older individuals, these waves are often observed during drowsiness or arousal and are even present during meditation. Apart from their occurrence in sleep states, theta waves are also associated with relaxed, meditative, and creative states. An abnormal increase in theta activity for a particular age group can signify dysfunctional brain function. Focal disruptions in theta wave patterns can be indicative of focal subcortical lesions. Conversely, a generalized distribution of theta waves may point towards conditions like diffuse disorders, metabolic encephalopathy, deep midline disorders, or certain types of hydrocephalus.

**Alpha Waves (7.5 Hz – 12 Hz):** Alpha waves was initially coined by Hans Berger to describe the rhythmic EEG activity he observed, known as the "posterior basic rhythm" or "posterior dominant rhythm" [81]. It is typically observed in the posterior regions of both sides of the head, alpha waves show greater amplitude on the dominant side. They become more prominent when the eyes are closed and during relaxation, but diminish in intensity with eye opening or mental exertion. In young children, the posterior basic rhythm may fall below 8 Hz, technically aligning with the theta range. Abnormal alpha wave patterns can be seen in conditions such as "alpha coma", where diffuse alpha activity is unresponsive to external stimuli.

**Beta Waves (12 Hz – 30 Hz):** Beta waves typically appear symmetrically on both sides of the brain, especially noticeable in the frontal region. Beta activity is closely related to motor behavior and is observed to diminish during active movements. Low-amplitude beta waves with multiple and varying frequencies are often associated with active or anxious thinking and focused concentration. Beta waves are predominantly seen in individuals who are alert, anxious, or have their eyes open. Rhythmic beta activity characterized by dominant frequencies can be indicative of conditions like Dup15q syndrome or the effects of certain medications, particularly benzodiazepines. In areas of cortical damage, beta waves may be absent or exhibit reduced activity.

**Gamma Waves (30 Hz – 45 Hz):** Gamma waves, which occupy the higher frequency range in EEG, are essential for understanding the synchronization of neuronal populations and their role in cognitive and motor functions. These waves are believed to represent the synchronization of different neuronal populations, forming networks that are responsible for specific cognitive or motor functions. Gamma rhythms are associated with the process of binding neural circuits together, indicating their role in complex brain activities and functionalities.

Understanding these EEG bandwidths and their corresponding mental states is crucial for interpreting EEG data and its application in both clinical and research settings [6].

**Table 4.1:** Correlation of EEG Bands with Brain States [6]

| Band | Frequency [Hz] | Brain State |
|---|---|---|
| Delta ($\delta$) | [1,4] | Deep sleep |
| Theta ($\theta$) | [4,8] | Meditation |
| | | Emotional stress |
| | | Creative inspiration |
| Alpha ($\alpha$) | [8,13] | Closed eyes wakeful state |
| | | Wakeful relaxation |
| | | Mental stress |
| Beta ($\beta$) | [13,30] | Strong mental activity |
| | | Problem solving |
| | | Concentration |
| Gamma ($\gamma$) | [30,100] | Cognitive activity |
| | | Motor activity |

### 4.1.5 EEG Signals Oscillations

EEG signals, characterized by rhythmic activity arising from neuron population excitability, provide insights into neural oscillations and cognitive functions. Alterations in the rhythmic patterns of neural oscillations can indicate the neurophysiological manifestation of cognitive functions. Changes in these patterns are often reflective of different states of the brain.

The strength of neural oscillations is quantified by the amount of energy conveyed by their electric fields per unit of time. This correlates with the power of the EEG signal. The relationship can be expressed through following equations: [78].

$$P = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x^*(t)x(t)\,dt \tag{4.1}$$

or for discrete signals:

$$P = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x^*[n]x[n] \tag{4.2}$$

Understanding the distribution of power across various frequency domains in EEG signals is crucial, and the power spectral density (PSD) serves as a key tool for this purpose.

PSD is used to specify how power is distributed across different frequency domains in an EEG signal. The PSD is defined as the Fourier transform of the signal's auto-correlation function, providing a means to analyze the signal in the frequency domain. The mathematical representation of this relationship is represented as following:

$$\text{PSD} = S_{xx}(\omega) = \int_{-\infty}^{+\infty} r_{xx}(\tau)e^{-i\omega\tau}\,d\tau \tag{4.3}$$

and:

$$r_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x^*(t)x(t+\tau)\,dt \tag{4.4}$$

Welch's method provides a reliable approach for estimating the PSD of time-discrete signals, crucial for EEG analysis. The initial step involves segmenting the signal into $K$ segments $x_k[j]$ of length $L$. These segments may overlap to ensure comprehensive analysis. Each segment is then weighted with a window function $w[j]$. This step is designed to reduce the leakage effect, enhancing the accuracy of the PSD estimation.

Following the application of the window function, the Fourier transform of these segments is computed. This process is mathematically represented by:

$$A_k[n] = \frac{1}{L} \sum_{j=0}^{L-1} x_k[j] W_L^{2kji} \tag{4.5}$$

In a next step, K periodograms $I_k$ are calculated:

$$I_k[f_n] = \frac{L}{U} |A_k[n]|^2 \tag{4.6}$$

$$f_n = \frac{n}{L}, \quad n = 0, \ldots, \frac{L}{2} \tag{4.7}$$

and:

$$U = \frac{1}{L} \sum_{j=0}^{L-1} w^2[j] \tag{4.8}$$

The estimated PSD is the average of the periodograms (eq. 4.9):

$$\text{PSD} = \hat{S}_{xx}[f_n] = \frac{1}{K}\sum_{k=1}^{K} I_k[f_n] \tag{4.9}$$

To obtain the band power, the PSD is integrated over the frequency intervals:

$$P_{\text{band}} = \int_{\text{lower limit}}^{\text{upper limit}} \text{PSD}(f)\, df \tag{4.10}$$

## 4.2 Device Specifications

This section outlines the methodology adopted for acquiring the data and performing the tests in this study, focusing on the equipment used.

### 4.2.1 Equipment Used for EEG Signal Acquisition

The EEG signals in this study were acquired using the Interaxon MUSE headband, a commercially available wearable device known for its efficacy in recording EEG data.

The details of the Interaxon MUSE headband and the location of its electrode on the scalp are illustrated in Figure 4.5.

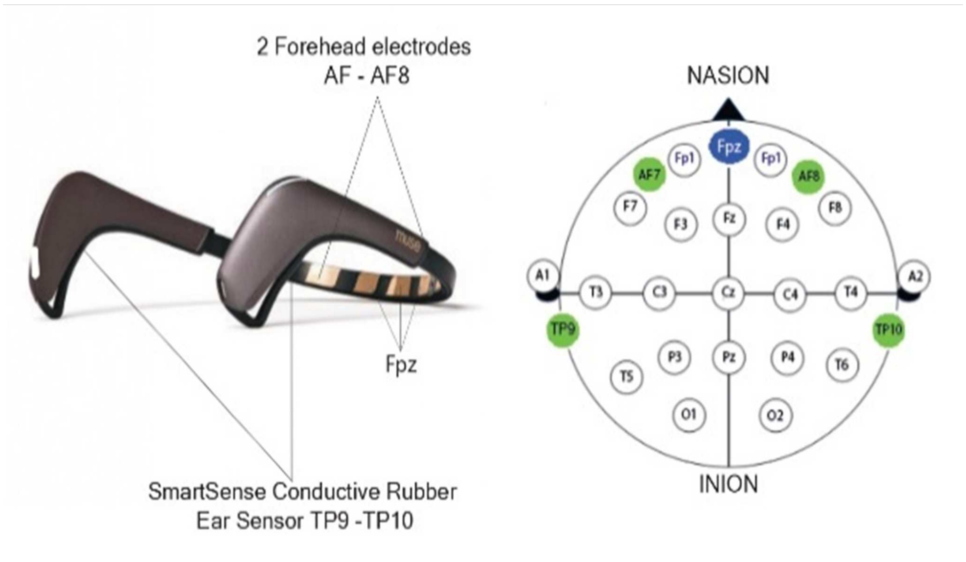



**Figure 4.5:** A view of the Interaxon MUSE headband and the locations of its electrodes in the 10-20 illustration system

#### Electrode Placement

In the Interaxon MUSE headband, the placement of electrodes followes an optimized EEG signal acquisition and is shown in Figure 4.5. The reference electrode FPz is positioned on the forehead, while the input electrodes consisted of two frontal electrodes (AF7 and AF8) located on the left and right sides of the reference, respectively, and are made of silver material. Additionally, two posterior electrodes (TP9 and TP10) are placed above each ear, using conductive silicone-rubber material.

#### Effectiveness of the MUSE Device in Research

Despite the challenges associated with EEG signal acquisition, the MUSE device has been demonstrated to be effective in various research settings. This is evidenced by several studies that have successfully utilized the MUSE device for different research purposes.

For instance, Krigolson et al. [82] employed the MUSE device for Event-Related Potential (ERP) studies with notable success. In their research, they assessed the reliability of ERP data collected using the MUSE device. Through resampling analysis, they were able to obtain reliable ERP components, particularly the N200, even with a limited number of participants. This finding underscores the potential of the MUSE device in capturing critical neural responses in ERP studies.

Another significant study conducted by Youssef et al. [83] focused on lie detection using the MUSE device. Their experiment achieved notable success, demonstrating the applicability of the MUSE device in experimental objectives beyond traditional EEG applications. These studies collectively highlight the versatility and effectiveness of the MUSE device in various research domains.

A significant comparison study was conducted by Ratti et al. [84]. provides insight into the performance of consumer-grade MUSE portable devices relative to traditional EEG medical devices. In their study, the researchers focused on comparing the power spectral densities (PSDs) obtained from MUSE with those from medical-grade EEG systems. The findings indicated that the PSDs of the MUSE device were similar to those of medical-grade systems. However, there was a slightly higher variation in the MUSE data, with power spectral ratios ranging from 1.125 to 1.225, compared to the 0.975 to 1.025 range for medical equipment. This broader spectrum increase in MUSE data could potentially be attributed to artifacts in the recordings made by the dry electrodes of the MUSE device.

Despite these variations, the simplicity and ease of setup of the MUSE device, along with its quick applicability (less than 10 minutes for setup), make it highly convenient for self-help applications and rapid deployment in various settings. Such findings underscore the utility and potential of consumer-grade EEG devices in both research and practical applications.

### 4.2.2 Earphones

The selection of appropriate headphones for use with the Muse 2 EEG headband was a critical aspect of the experimental setup, with the aim of ensuring optimal electrode adhesion and participant comfort.

Due to compatibility issues with over-ear headphones, which can interfere with the proper adhesion of the EEG electrodes, a more suitable type of headphone was necessary. For this purpose, the Sony MDR-E9LP compact earphones were chosen. These earphones were preferred because they do not disrupt the positioning or functioning of the Muse 2 EEG headband's electrodes.

Additionally, in-ear headphones were not considered for this study as they may not be comfortable for all participants. The comfort of the participants was a priority, and it was essential to select a headphone type that would be universally acceptable and comfortable for all individuals involved in the test. This consideration was crucial to ensure participant compliance and the quality of EEG data recorded.

### 4.2.3 Data Acquisition Process

The MUSE device used for recording the EEG signals operated at a sampling frequency of 256 Hz, and the data acquisition process was carried out using the MUSE application, a key component of the EEG signal recording setup. This application was paired with a smartphone via Bluetooth Low Energy (BLE) technology, enabling efficient and wireless communication. The impedance check, an essential step for ensuring signal quality, was facilitated by the application. This was visually confirmed in real-time by observing the raw signal on the smartphone screen.

#### Challenges in EEG Signal Acquisition with Frontal Electrodes

The acquisition of EEG signals, particularly with frontal electrodes, poses specific challenges that can impact the accuracy and reliability of the data collected. Frontal electrodes are notably more susceptible to capturing artifacts caused by eye blinks and movements, which can significantly interfere with the accurate measurement of actual brain waves [85]. This susceptibility necessitates careful consideration in both the placement of electrodes and the interpretation of the recorded data.

Moreover, the use of dry electrodes, often employed for their ease of setup and non-invasiveness, can lead to discomfort over extended periods of use. Additionally, there is a heightened risk of misplacement on the forehead with dry electrodes, potentially resulting in decreased signal accuracy. Factors such as the participant's head shape, size, and hairstyles can further complicate the data collection process. Insufficient contact with the scalp, often arising from these individual differences, may impede proper signal acquisition, thereby affecting the overall quality of the EEG data [86].

## 4.3 EEG Test Campaign for Acoustic Evaluation

The effectiveness of EEG measurements in the context of sound quality jury testing was a key focus of this study. To this end, comprehensive EEG measurements were performed in tandem with acoustic sound quality assessments in September 2022.

The acoustic laboratory at the ''Università Politecnica delle Marche" in Italy served as the venue for these experiments. The choice of this location was driven by its suitability for conducting detailed acoustic and EEG analysis.

### 4.3.1 Experimental Protocol

This section explains the experimental protocol in detail, from initial preparations to the conclusion of the listening test.

The participants were equipped with the Muse headband and headphones, as depicted in Figure 4.6. They were comfortably seated in a chair positioned approximately 70 cm away from a screen. Throughout the course of each trial, the participants were instructed to maintain a relaxed posture, minimize muscle tension, and try to limit eye movements as much as possible. These precautions were essential to reduce potential artifacts in the EEG data and to ensure the accuracy and reliability of the recordings.

**Adjustment and Communication:** The Muse headband was carefully adjusted for each participant's comfort. The experimental protocol and data management procedures were clearly communicated to all volunteers, ensuring they were well-informed about the study's nature and objectives.

**Information Sheet and Consent:** To maintain confidentiality, the data collected from the volunteers underwent a process of anonymization. Each volunteer was required to sign a privacy information sheet and a consent form explaining the purpose of the study. Subsequently, participants were briefed on the test procedure, outlining what to expect and the do's and don'ts during the experiment.

**Questionnaire and Personal Information Collection:** Volunteers were instructed to use their smartphones to scan a QR code leading them to an online questionnaire created with Google Form (See Figure 4.8). In the questionnaire, participants provided personal information such as first name, last name, age, gender, country of origin, and the volume value set on the computer for the test. Additionally, they answered a series of questions related to the sounds they would be listening to during the experiment.

The questions were structured into three sets of three questions, to be filled out immediately after listening to each audio segment. Specifically, participants were asked to rate how annoying or pleasant, relaxing or stressful, and quiet or loud they found each audio piece. This feedback was crucial for assessing the subjective experience of each participant regarding the sounds played during the test.

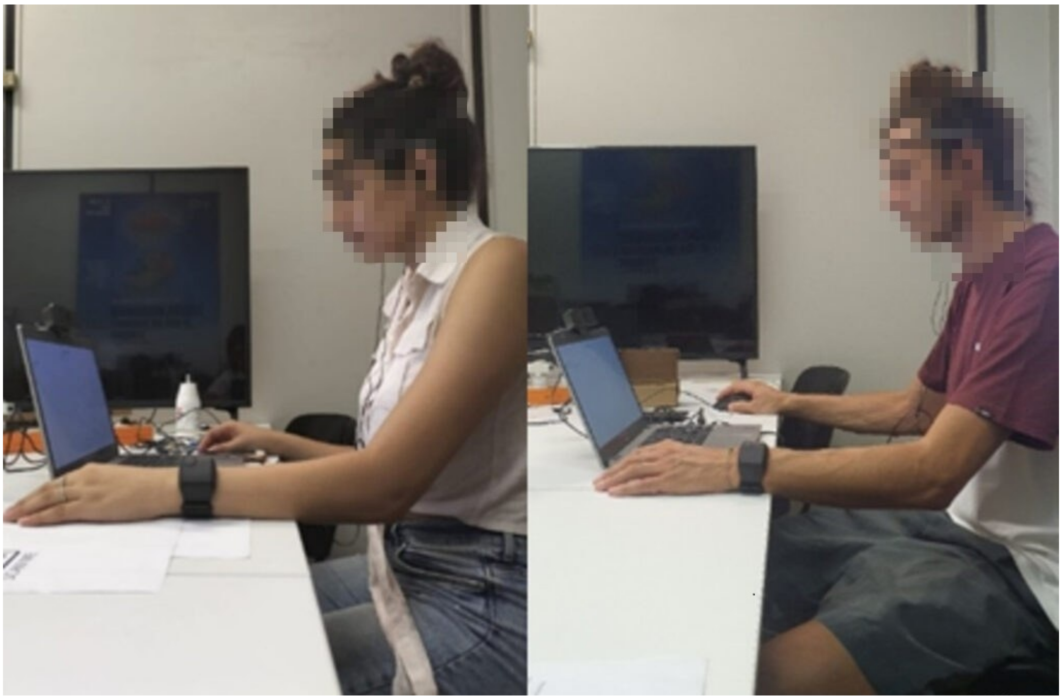


**Figure 4.6:** View of the performing the test

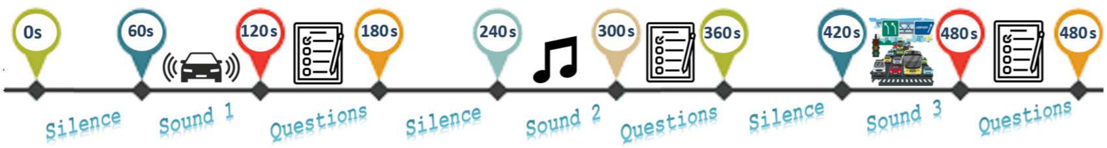



**Figure 4.7:** Experiment sequence

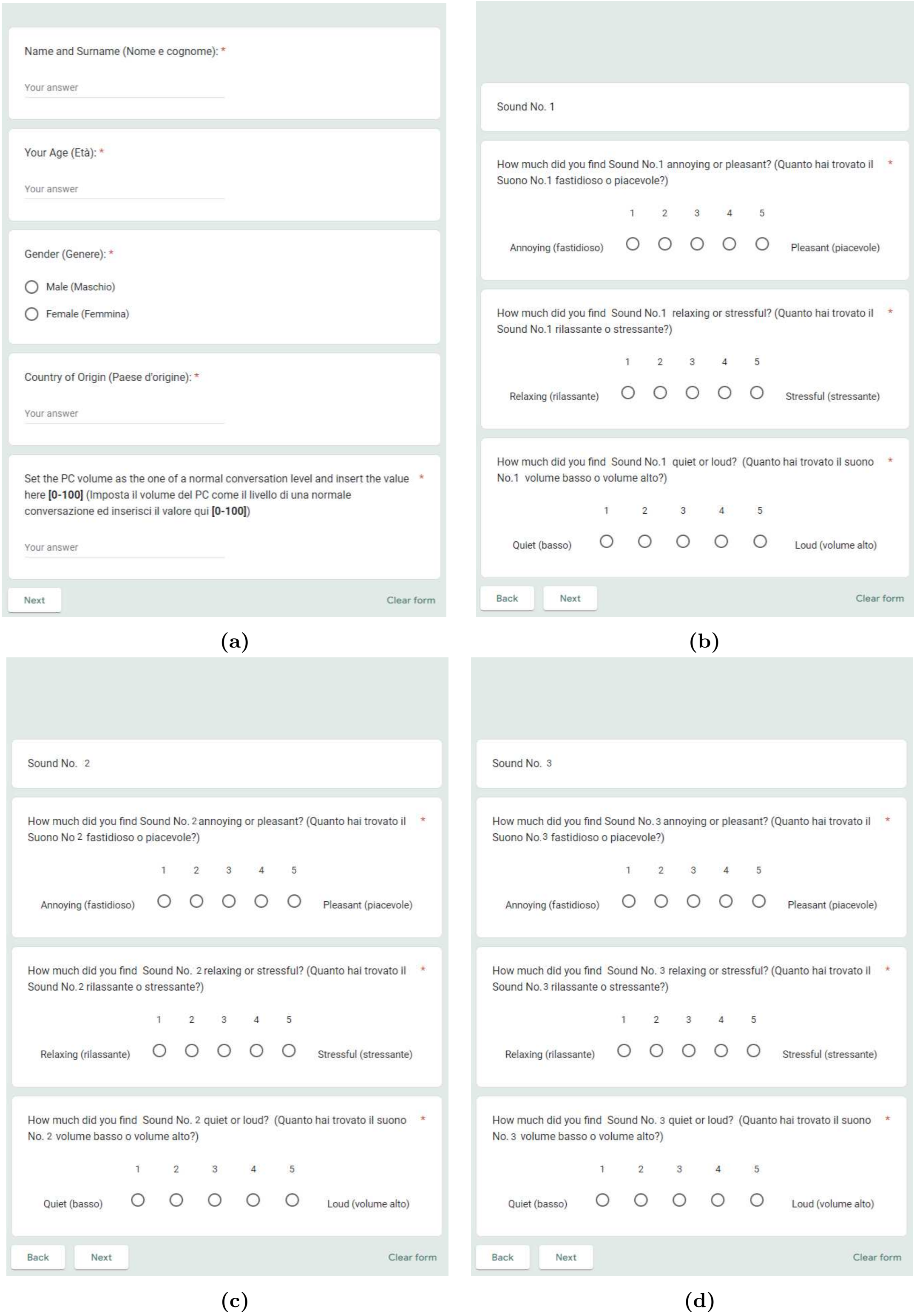



**Figure 4.8:** Sound quality assessment questionnaire created by GoogleForms. (a), (b), (c), and (d) respectively illustrate the first, second, third, and fourth pages of the questionnaire.

### 4.3.2 Test Contents and Sequence

The test consisted of three one-minute audio segments, each preceded by a one-minute silence interval (refer to Figure 4.7). The silence served to relax the participant between the different audio stimuli, mitigating any residual effects from the previous audio.

Participants initiated the test by clicking a point on the PC desktop, which triggered a “beep” sound. This sound served both to inform the participants that the test had begun and to aid in synchronizing the data for later processing. During the test, the proper coupling between the electrodes and the participant’s skin, crucial for the quality of EEG data, was monitored live through the Muse Headband app. The use of Bluetooth Low Energy (BLE) technology enabled communication between the EEG Muse Headband and a paired smartphone. After approximately six minutes, encompassing the listening of the three audios, the test was concluded by an operator stopping the recording. Subsequently, the captured data was automatically transferred to a Google Drive folder, allowing for real-time review by all operators involved in the project.

#### Audio Description

In this study, three distinct auditory stimuli were utilized to evoke emotional responses in participants. The selection of each sound was based on its distinct acoustic properties, including spectral attributes. Figure 4.9 visually presents spectrograms for the three chosen stimuli, providing a comprehensive visualization of their spectral content. Moreover, Table 4.2 serves as a crucial reference, presenting the spectral characteristics obtained from the temporal distribution of frequency components for each sound. Systematic analysis of these metrics can offer valuable insights into the nuanced acoustic differences within the sounds. Building on the observations from Figure 4.9 and Table 4.2, the subsequent section provides a detailed exposition of each sound, along with their corresponding spectral metrics.

Sound No. 1 encompasses the noise generated by an internal combustion engine, evoking associations with power, mechanical precision, and intensity. This sound exhibits a centroid value of 341 Hz, indicating a moderate concentration of frequencies around the mid-range. The entropy value of 0.48 signifies a relatively balanced distribution of frequencies. A crest value of 1152 points to a dynamic sound characterized by pronounced peaks. Additionally, a kurtosis value of 148 suggests an elevated concentration of spectral energy around the mean frequency, while a skewness value of 10.2 implies substantial asymmetry in the spectral distribution.

Sound No. 2 features soothing music, eliciting sensations of tranquility, relaxation, and harmony. The sound showcases a centroid value of 203 Hz, indicative of a lower frequency concentration. An entropy value of 0.36 implies a relatively predictable frequency distribution. A crest value of 1617 underscores a highly dynamic sound with prominent peaks. Furthermore, a kurtosis value of 6 indicates a more uniform distribution of spectral energy, while a skewness value of 1.0 points to a relatively symmetrical spectral distribution.

Sound No. 3, represented by road noise, conjures associations of a bustling ambient environment. This sound is characterized by a higher centroid value of 538 Hz, indicating a preponderance of frequencies in the higher range. An entropy value of 0.65 suggests a more intricate and diverse frequency distribution. A crest value of 564 denotes a moderately dynamic sound with discernible peaks. Additionally, a kurtosis value of 36 suggests a moderate concentration of spectral energy around the mean frequency, while a skewness value of 4.5 hints at a moderate degree of asymmetry in the spectral distribution.

**Table 4.2:** Spectral Characteristics of the Sounds. The presented values are median values derived from the temporal distribution of frequency components.

| | **Audio description** | **Centroid** | **Entropy** | **Crest** | **Kurtosis** | **Skewness** |
|---|---|---|---|---|---|---|
| Audio 1 | Engine noise | 341 Hz | 0.48 | 1152 | 148 | 10.2 |
| Audio 2 | Music | 203 Hz | 0.36 | 1617 | 6 | 1 |
| Audio 3 | Road noise | 538 Hz | 0.65 | 564 | 36 | 4.5 |

### 4.3.3 Estimation of Effective Sample Size

To estimate the effective sample size needed for the experimental campaign, a G*Power3 statistical analysis was performed [87]. The priori analysis setting, as outlined in Table 4.3, provided the basis for this estimation. The analysis yielded a sample size of 36 for the parametric tests and 38 for the non-parametric tests.

**Table 4.3:** G*Power3 statistical setting parameters.

| **Parameters** | **Values** |
|---|---|
| required power level (1- $\beta$) | 0.95 |
| prespecified significant level ($\alpha$) | 0.05 |
| the effect size $f$ | 0.25 |
| Number of measurements | 3 |
| Number of groups | 1 |
| statistical test | repeated measures - within factor ANOVA and test<br>Wilcoxon signed-rank test |

### 4.3.4 Participants Demographics

The experimental campaign ultimately included a total of 43 participants, comprising 21 females and 22 males. The participants were primarily young students, with ages ranging from 19 to 61 years old. Notably, 80 percent of the participants were within the age range of 28 $\pm$ 7 years, highlighting the study's focus on a younger demographic.

### 4.3.5 EEG Signal Pre-processing

The Muse 2 headband is equipped with an integrated Digital Signal Processing (DSP) module that undertakes essential data preprocessing steps. Notably, the DSP module incorporates a bandpass noise filter operating within the frequency range of 0.1 to 45 Hz, effectively eliminating unwanted frequencies. Moreover, a 50 Hz notch filter is implemented to eliminate power line interference. To enhance data quality, the DSP module employs advanced techniques for the removal of artifacts stemming from eye blinks and jaw clenching. To unlock frequency domain insights, a Fast Fourier Transform (FFT) is executed using a window size of 256 samples and a step interval of 22.

#### Feature Extraction

The resultant FFT provides access to PSD within distinct frequency bands. Specifically, five crucial frequency bands are extracted: delta, theta, alpha, beta, and gamma. Within each of the aforementioned frequency bands, comprehensive feature extraction is performed across four input channels: TP9, AF7, AF8, and TP10. The following essential features are computed for each channel within these frequency bands:

- Relative Power:

$$\frac{\sum_{i=0}^{n} B}{\sum_{i=0}^{n} P} \tag{4.11}$$

  where $B$ is the power of the signal in a specific frequency band and $P$ is the total power.

- Frontal Asymmetry (FA):

$$B_{AF8} - B_{AF7} \tag{4.12}$$

- Temporal Asymmetry (TA):

$$B_{TP10} - B_{TP9} \tag{4.13}$$

- Band Ratio in Each Channel:

$$\frac{alpha}{beta} \tag{4.14}$$

### 4.3.6 Quantification of Individual Sound Perceptions

The frequencies of individual responses regarding personal sound perceptions were computed. Employing a tailored Python script, a robust statistical analysis was undertaken to establish intrinsic correlations connecting individual acoustic perceptions with dynamic changes in EEG features. This exploration was conducted across three distinct audio stimuli.

### Statistical Validation

In the pursuit of statistical rigor, the Shapiro test and Bartlett test were harnessed to assess the underlying assumptions of normality and variance homogeneity, respectively. Notably, within the cohort of features, a solitary exception emerged ($-DeltaAF8$), demonstrating a distribution distinct from the Gaussian one [88][89].

To discern the nuances of EEG feature means across distinct variables under differing conditions, two distinct methodologies were exploited. The first, the repeated measures ANOVA model, was chosen given the satisfaction of both normality and homogeneity assumptions. The second, the Kruskal-Wallis non-parametric test, was selectively employed when these assumptions were not upheld [90].

### In-depth Post-hoc Examination

The discriminative potential of EEG features across different auditory conditions (A1-A2, A1-A3, and A2-A3) was subjected to in-depth post-hoc analysis. This assessment was conducted using the Dwass-Steel-Critchlow-Fligner pairwise comparison test [91].

## 4.3.7 Quantifying Measurement Uncertainty

Drawing inspiration from the Guide to the Expression of Uncertainty in Measurement (GUM) **??**, a detailed uncertainty analysis was conducted on the EEG features obtained from the experiments.

In the process of analyzing these EEG features, the arithmetic mean and the uncertainty of the mean were key metrics of interest. The arithmetic mean, denoted as $\bar{f}$, provides a central value of the data, representing the typical EEG feature value. Meanwhile, the standard uncertainty of the mean, denoted as $u_f$, offers insight into the variability and reliability of the EEG measurements.

The calculation of these two statistical parameters for each unique EEG feature allowed for a more nuanced understanding of the data. This approach aligns with the principles outlined in the GUM, ensuring that the analysis adheres to established standards in measurement and uncertainty expression [92].

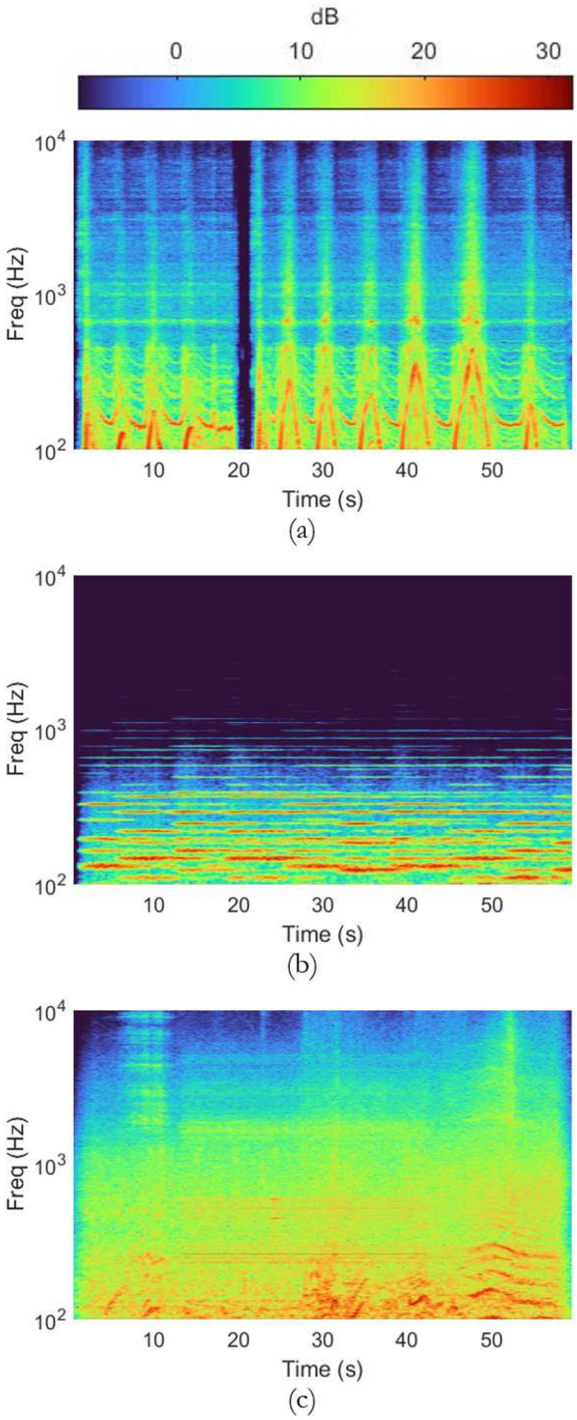



**Figure 4.9:** Spectrograms illustrating the spectral content of three sound stimuli over time. (a) engine noise, (b) soothing music, and (c) road noise

# Chapter 5

# Results and Discussions

## 5.1 Results of Experiments on Facial Expression in Acoustic Perception

This section provides a detailed discussion of the results from the experiments aimed at assessing the role of facial expression analysis in sound quality jury testing. The analysis covers three distinct test campaigns designed to study facial expressions in response to acoustic stimuli.

The first part of the analysis focuses on the two preliminary tests. These tests were crucial in setting the groundwork for understanding how facial expressions vary in response to different acoustic stimuli. The methodology, data collection, and analytical techniques employed in these preliminary tests provided valuable insights, shaping the approach for the subsequent jury test campaign.

Afterward, the jury test campaign which was conducted with an added component of facial expression analysis is discussed. This campaign was pivotal in exploring the practical application of facial expression analysis in a sound quality jury testing context. The results from this campaign are crucial in evaluating the feasibility and effectiveness of integrating facial expression analysis into sound quality assessments.

### 5.1.1 Results of Preliminary Test 1

To evaluate the effect of different stimuli on participant engagement, a non-parametric Friedman test was performed. This choice was necessitated by the lack of normality in the distributions of the engagement data. The stimuli, categorized into three macro-categories, served as the between-subject factor, with engagement as the dependent variable.

The results of the Friedman test revealed a statistically significant difference in engagement across different stimuli, $\chi^2(2) = 23.2, p < 0.001$. This finding indicates that the type of stimulus had a considerable impact on the level of engagement among participants.

Following the Friedman test, post hoc analysis was conducted using Wilcoxon signed-rank tests. To account for multiple comparisons, a Bonferroni correction was applied, setting the significance level at 0.017.

The median (Interquartile Range, IQR) engagement levels for the three macro-categories: "Sound", "Sound with Video", and "Emotional Video", were found to be 7.2 (5.6 to 12.0), 8.8 (5.5 to 13.0), and 10.0 (7.0 to 13.0), respectively. These values indicate varying degrees of engagement elicited by each type of stimulus.

In the pairwise comparisons, no significant difference was found between the "Sound" and "Sound with Video" categories (Z = -1.350, p = 0.177). However, significant differences were observed in engagement levels between the "Sound with Video" and "Emotional Video" categories (Z = -3.236, p = 0.001), as well as between the "Sound" and "Emotional Video" categories (Z = -4.083, p ¡ 0.001).

The discernible distinctions among jurors became readily apparent from a qualitative standpoint, as evidenced in Figure 5.1 and 5.2, where we compare the engagement data of two jurors. It is evident that jurors responded divergently to the presented stimuli.

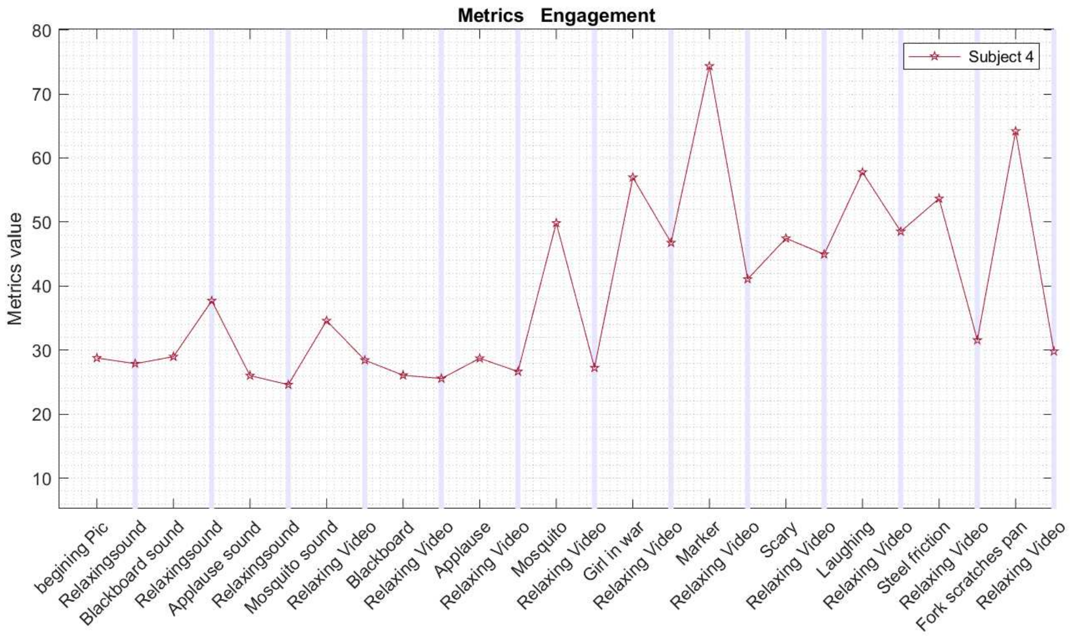



**Figure 5.1:** Engagement values averaged over each stimulus for Subject No.4

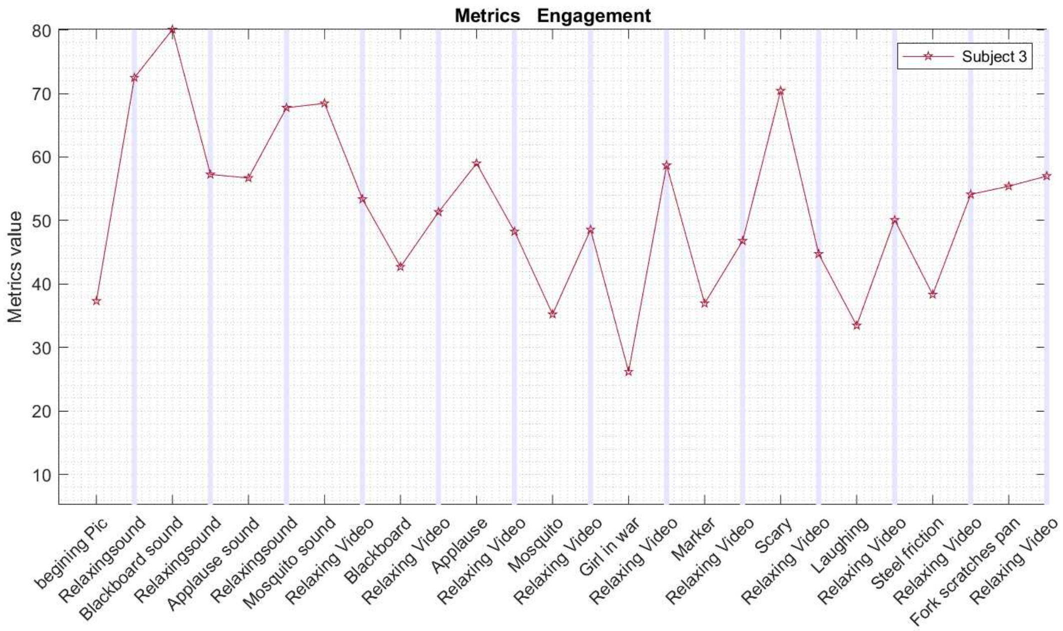



**Figure 5.2:** Engagement values averaged over each stimulus for Subject No.3

While we could discern clear differences in the jurors' responses based on qualitative observations, we aimed to support our findings with quantitative data. We wanted to demonstrate that the different stimuli evoked varying levels of emotional reactions. Specifically, we found that the categories labeled as "Video" and "Video with annoying sounds" elicited stronger emotional responses compared to those categorized as "Sound" or "Sound with video" (please refer to Figure 5.3).

In Figure 5.3, we present boxplots that provide a statistical evaluation of how emotional measures changed when a specific stimulus (in this case, the applause sound) was presented. Jurors were exposed to this stimulus both in isolation and in conjunction with a video featuring the same applause sound. The results clearly show that when accompanied by an animated video background, jurors exhibited more pronounced emotional expressions compared to when they only heard the sounds.

This suggests that visual stimuli, particularly animated video, play a significant role in intensifying the emotional impact of the experience for jurors. It highlights the importance of considering multimedia elements in the presentation of stimuli, as they can significantly influence juror responses.

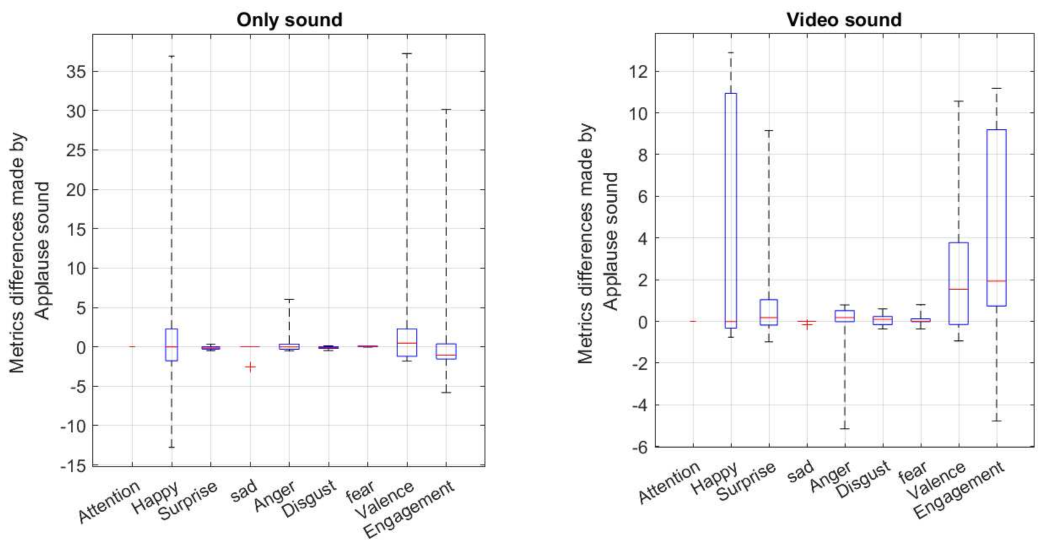



**Figure 5.3:** Mean values and 95% confidence intervals for the means of the 3 macro-categories used in preliminary test 1

### 5.1.2 Results of Preliminary Test 2

Pearson product-moment correlation coefficients were computed to assess the relationship between the emotional valence and the subjective answers provided by the jurors. The results, as shown in Table 5.1, revealed positive correlations between the first audio and the first question ($r = 0.272, p = 0.09$), indicating that jurors found the audio to be pleasing and that the tool effectively captured this interaction. Additionally, negative correlations were observed between the second audio and the second question ($r = -0.333, p = 0.04$), suggesting that while jurors found the audio to be relaxing, the tool predominantly assigned a negative valence. No significant correlations were found between valence and the other audio files and questions. In summary, the tool demonstrated limited capability to capture substantial correlations between valence data and the questionnaires. Achieving a higher precision would be necessary to infer jurors' preferences for one set of videos over another.

**Table 5.1:** Correlations coefficient between the emotional valence and the questionnaire answers given by the jurors

| | **Annoying vs. pleasantness** | **Relaxing vs. stressful** | **Quiet vs. loud** |
|---|---|---|---|
| Audio 1 | 0.272 | -0.135 | -0.112 |
| Audio 2 | 0.055 | -0.333 | -0.052 |
| Audio 3 | -0.076 | -0.205 | -0.092 |

### 5.1.3 Results of Jury Test Campaign enhanced by Facial Expression Measurement

This section is dedicated to presenting the findings from the Jury Test Campaign, which was augmented by the inclusion of Facial Expression Measurement. The results here detail the patterns, variations, and key observations identified through the combination of traditional jury testing methods and advanced facial expression analysis. The analysis of facial expressions offered an additional layer of data, contributing to a more comprehensive understanding of the reliability of participants' subjective experiences and reactions to the test sounds.

#### Correlational Analysis Between Acoustic Perceptions and Facial Expressions

To deepen our understanding of the potential correlation between acoustic perceptions and facial expression analysis, a comprehensive statistical analysis was conducted. This involved dissecting the emotion values, which were extracted from the facial expression analysis, based on the jurors' choices of soundtracks.

The emotion values obtained from the facial expression analysis were segregated into distinct periods, corresponding to the preferences indicated by the jurors. These periods were categorized based on whether the jurors chose sound A, sound B, or expressed an equal preference for both.

The analysis of emotional responses to acoustic stimuli is further elucidated in Figure 5.4, which presents a series of boxplots representing the manifestation of temporal change in anger during audio B compared to the period of audio A and related to the first question, which probes the annoyance level of the sound.

Each boxplot in Figure 5.4 corresponds to an individual juror, depicting the range of changing anger values that emerged when the juror perceived a sound as annoying. The position and spread of each boxplot provide valuable insights. When a boxplot is positioned above zero (above the x-axis), it indicates that the juror exhibited an increase in anger in response to finding the sound annoying. Conversely, a boxplot below zero suggests a decrease in anger, implying a more subdued emotional response to the annoying sound.

The trends observed in Figure 5.4 can be contrasted with those in Figures 5.5 and 5.6, which represent the emotional responses of valence and engagement, respectively. The general observation across these figures is that most boxplots are centered around zero or display random values. This pattern suggests that the jurors did not exhibit a considerable change in anger, valence, or engagement in response to sounds they found more annoying. However, it is important to note the individual variations among jurors, indicating diverse emotional responses to the same acoustic stimuli.

From the collective analysis of Figures 5.4, 5.5, and 5.6, several inferences can be drawn. Despite the majority of the boxplots being clustered around zero or displaying random values, indicating no significant change in anger levels, there are notable exceptions. In Figure 5.4, for instance, 10 boxplots are positioned above zero, suggesting an increase in anger, while only two are below zero, indicating a decrease. The remaining 29 boxplots are balanced at zero, reflecting a neutral response in terms of anger.

This pattern is similarly observed in the valence and engagement analyses. When jurors perceive a sound as annoying, there tends to be a decrease in valence and an increase in engagement, indicating a shift towards negative emotions and heightened attention. Therefore, while responses vary among individual jurors, a general trend can be observed wherein annoying sounds elicit increased engagement and potentially negative emotional states.

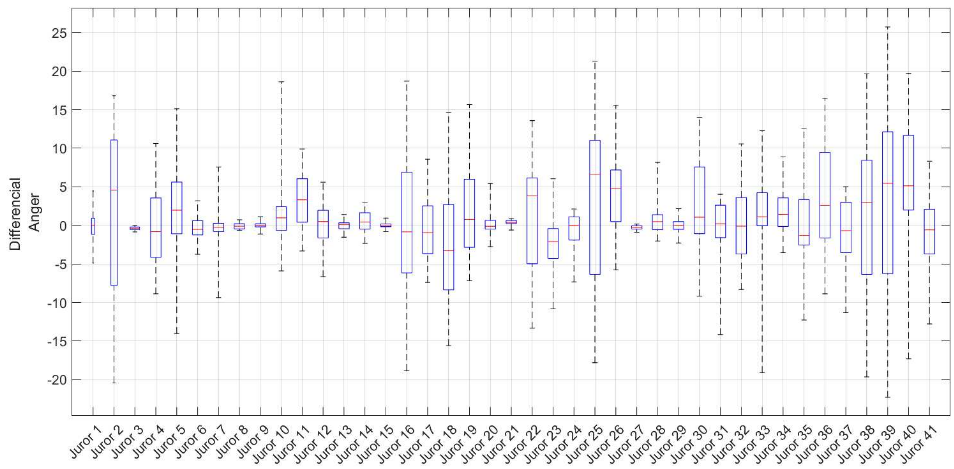



**Figure 5.4:** Changes in anger values extracted for the periods of times when the jurors chose audio B as the more annoying audio

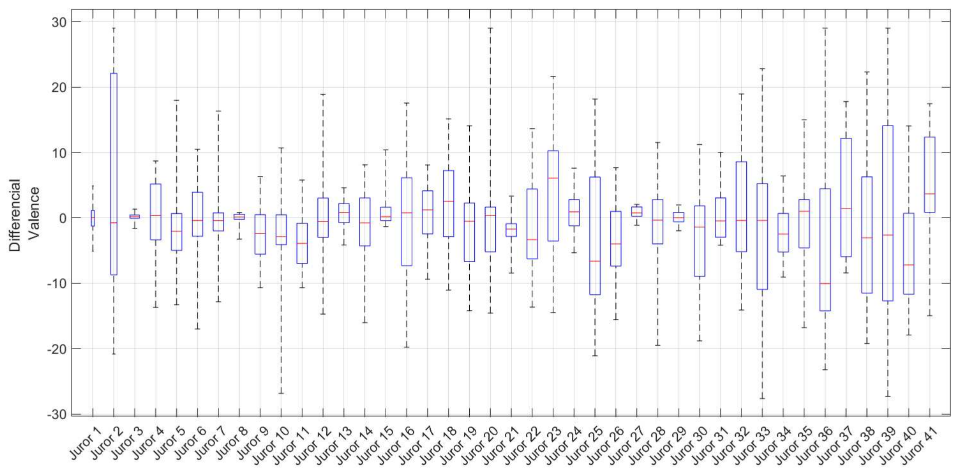



**Figure 5.5:** Changes in valence values extracted for the periods of times when the jurors chose audio B as the more annoying audio

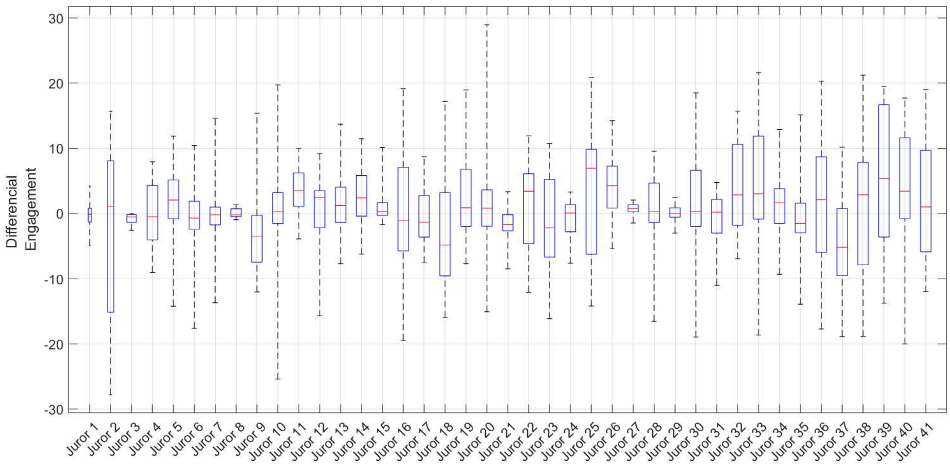



**Figure 5.6:** Changes in engagement values extracted for the periods of times when the jurors chose audio B as the more annoying audio

### Questionnaire Results

At the final stage of the jury test process, after finishing the sound quality evaluation, jurors were asked to complete a questionnaire. In this stage, in addition to providing demographic information such as age and gender, jurors were mainly tasked with expressing their levels of concentration during different stages of the test.

Jurors evaluated their concentration levels at three distinct phases: the beginning, middle, and end of the test. They could select from various categories to best describe their mental state, such as:

- Fully concentrated
- Actively involved
- Partially concentrated
- Distracted
- Feeling bored
- Heavy-eyed
- Feeling mentally fatigued

The results, depicted in Figure 5.7, revealed a trend in jurors' concentration levels throughout the test. Initially, jurors mostly reported being "fully concentrated" or "actively involved." However, as the test progressed, a shift was observed, with jurors indicating lower concentration levels, feeling "partially concentrated" or "bored". Especially after the middle of the test, towards the end of the test, jurors predominantly expressed feelings of being "heavy-eyed" and "mentally fatigued," signifying a decrease in overall alertness and focus. This trend suggests that jurors started the test with a high level of freshness and active involvement but experienced a gradual decline in concentration, attributed partly to the lengthy duration of the test (refer to Figure 5.8).

Understanding jurors' concentration levels is pivotal in interpreting the results accurately. The observed trends could influence the jurors' perception and evaluation of the sound qualities, making this an essential aspect to consider when analyzing and drawing conclusions from the collected data.

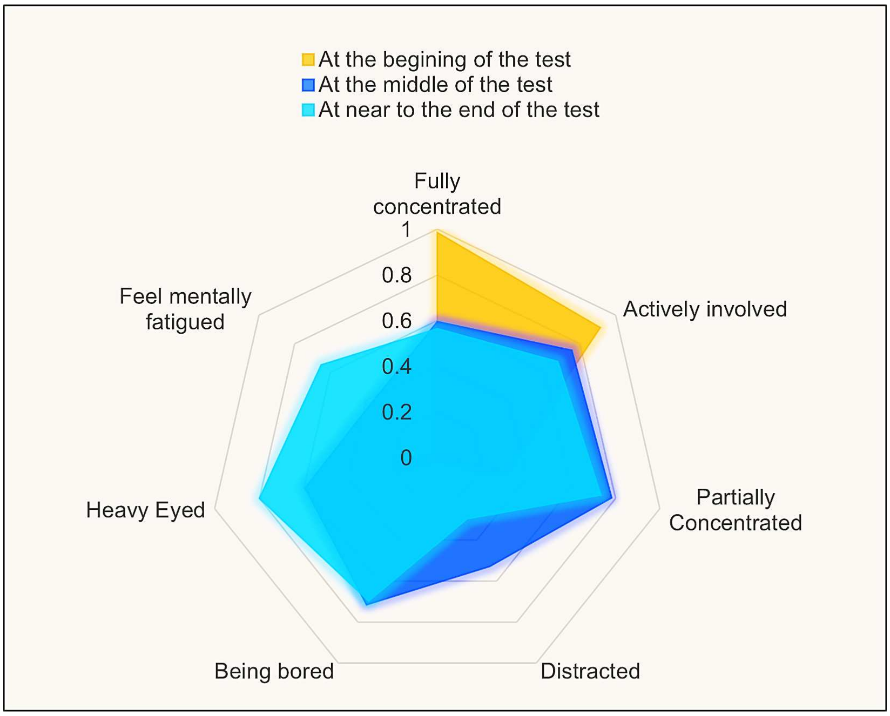



**Figure 5.7:** Radar chart to represent the jurors' concentration levels throughout the test campaign

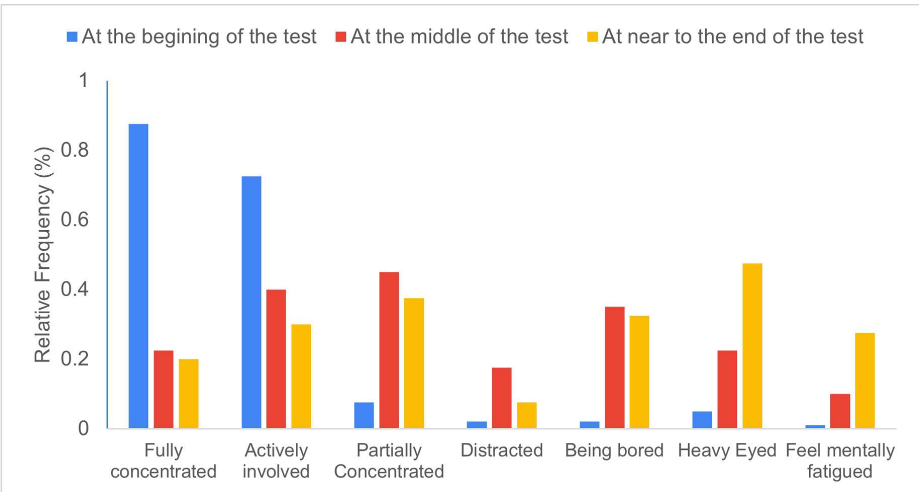



**Figure 5.8:** Bar chart to illustrate the quantitative jurors' concentration

### Correlation between Facial Expression and Concentration Level

Based on the jurors' self-reported mental statuses from the questionnaire, facial expression data were extracted corresponding to the specific time periods when jurors expressed each mental status. This allowed for the assessment of emotional values related to each mental/concentration status from a statistical viewpoint.

The emotional metric of attention was selected as a key indicator to discern differences across various concentration statuses due to the fact that attention value is a representative metric of jurors' levels of concentration and involvement in the experiment.

In Figure 5.9, the attention levels are visualized using three box plots representing three distinct concentration statuses i.e. fully concentrated, distracted, and heavy-eyed. Each box plot symbolizes the range of attention values associated with a specific concentration status, providing a visual representation of the jurors' attention variations corresponding to their self-reported mental states. Figure 5.10 provides a comparative view by subtracting the attention values associated with a "fully concentrated" status from all other attention values linked to each concentration status and setting the "fully concentrated" status as a baseline with fixed zero values. This adjustment made the differences more apparent.

Accordingly, positive values in Figure 5.10 indicate an increase in attention levels when transitioning from a "fully concentrated" status to another, while negative values signify a decrease. An observable shift below zero in the box plots related to the "distracted" and "heavy-eyed" statuses, suggests a reduction in attention levels compared to the "fully concentrated" status. It reveals an overall discernible decrease in attention levels, calculated from facial expression analysis, when jurors felt distracted or heavy-eyed compared to when they were fully concentrated. This insight is crucial for understanding the reliability and validity of the jurors' responses, acknowledging the impact of varying concentration levels on their subjective sound quality assessments.

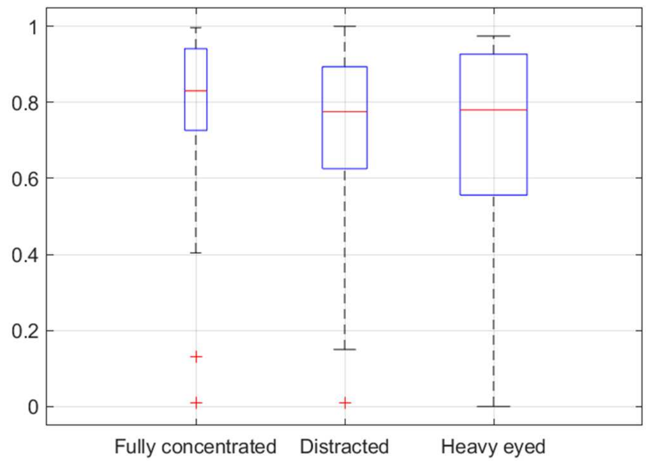



**Figure 5.9:** Boxplot charts of attention values for different concentration statuses. Note that the attention values normalized on a scale from 0 to 1

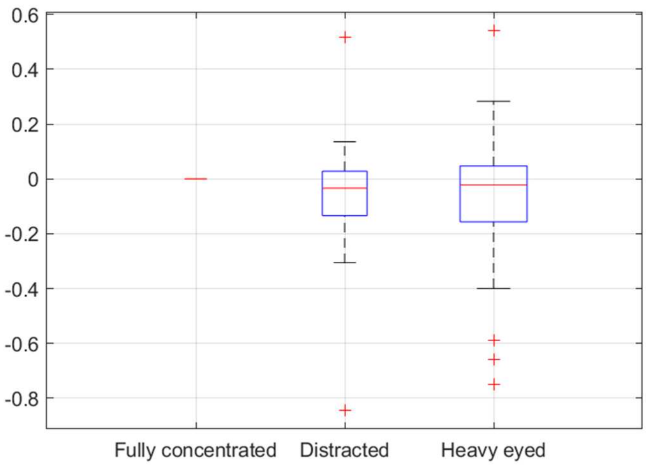



**Figure 5.10:** Comparative boxplot charts depicting attention values across various concentration statuses, with "fully concentrated" as the baseline

## 5.2 Results of Physiological Measurement

This section focuses on the findings from the EEG measurement campaign, specifically examining the effects of acoustic stimuli on the participants' brain activity.

### 5.2.1 Subjective Acoustic Perception Distribution

Upon thorough examination of the averaged ratings derived from the questionnaire responses, noteworthy distinctions emerged across the three auditory stimuli, and depicted in Figure 5.11.

The musical stimulus got the highest scores for both pleasantness and relaxation, signifying its positive influence on participants. In contrast, road noise received the lowest ratings in terms of pleasantness and relaxation, possibly reflecting its disruptive impact on participants' emotional states. Car noise consistently occupied intermediary positions in these dimensions.

Moreover, participants consistently perceived road noise as the loudest among the stimuli, evident from the highest average rating in the quietness/loudness dimension. Conversely, music was perceived as the least loud, aligning well with its calming and enjoyable attributes.

These outcomes underscore the diverse impact of auditory stimuli on participants' subjective experiences, thereby holding implications for comprehending how distinct sounds may elicit varying reactions in individuals.

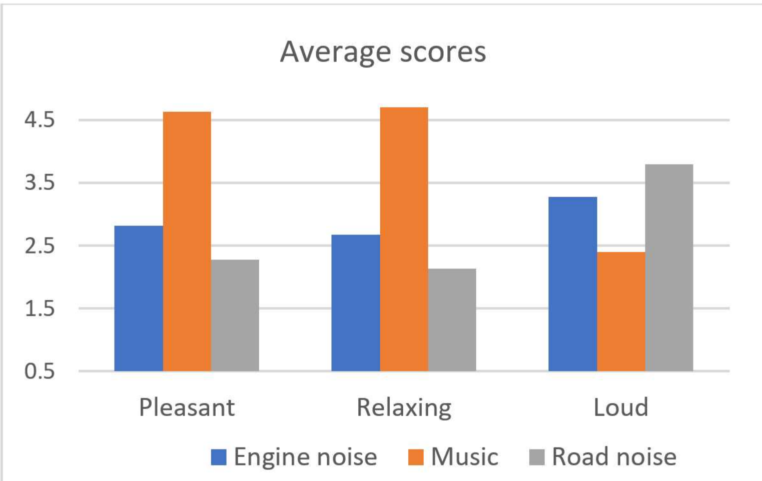



**Figure 5.11:** Subjective questionnaire results overview

### 5.2.2 EEG-Acoustic Perception Correlation

Broadly, the EEG measurements exhibited meaningful correlations with acoustic perceptions, specifically relating to the amplification or attenuation of brain wave power. Elevated delta values were discerned in the right frontal and temporal electrodes (TP10, AF8) as well as theta TP10 under A1 (Audio 1) exposure, distinguishing it from the other stimuli. Post-hoc analysis established statistically significant differences in delta TP10, delta AF8, and theta TP10 between both A1-A2 and A1-A3 conditions (see Table 5.2 and Figures 5.12, 5.13, and 5.14). Moreover, relative alpha TP10 demonstrated higher values during A2 (Audio 2) exposure. Notably, a surge in TA alpha was evident under A3 (Audio 3) exposure, with TA showing statistical contrasts in the A2-A3 condition. The subsequent section provides an in-depth analysis of the results. This comprehensive examination aims to distil key findings, interpret significant patterns, and offer insights derived from these referenced materials.

**Delta Waves and Emotional Experience**

Delta waves, as the lowest recorded brain waves in humans, are intrinsically linked with deep relaxation aspects of restorative sleep. Notably, delta activity has been noted to hold significance in emotional experiences. The findings by J.L. Walker [69] establish a correlation between high-delta and high-theta production and responses to 'unpleasant' music. Additionally, this study unveils a connection between elevated delta activity and a state of 'paying little attention', aligning with the hypothesis that prevalent delta activity is modulated by perceptions of annoying sounds. Our outcomes echo this notion, emphasizing that heightened delta activity aligns with perceptions of annoyance (refer to Figures 5.12 and 5.13b).

### 5.2.3 Theta Waves and Annoyance

Consistent with the work of Zheng-Guang Li et al. [93], our results showcase an increase in theta waves in response to annoying sounds. Their findings demonstrated that theta wave activity escalates as a subjective sense of annoyance intensifies during noise exposure. This congruence highlights the robust association between theta waves and the perception of annoyance (refer to Figure 5.13b).

### 5.2.4 Alpha Waves and Auditory Attentiveness

According to Figure 5.13a, the prominence of alpha activity has been linked to heightened attentiveness to sounds and a state of pleasant relaxation. Notably, emotionally significant stimuli inherently capture attention. Consequently, the possibility that participants directed enhanced attention towards the rhythmic auditory stimuli during musical sound is conceivable [94].

### 5.2.5 Interplay Between Auditory Context and Brain Activation

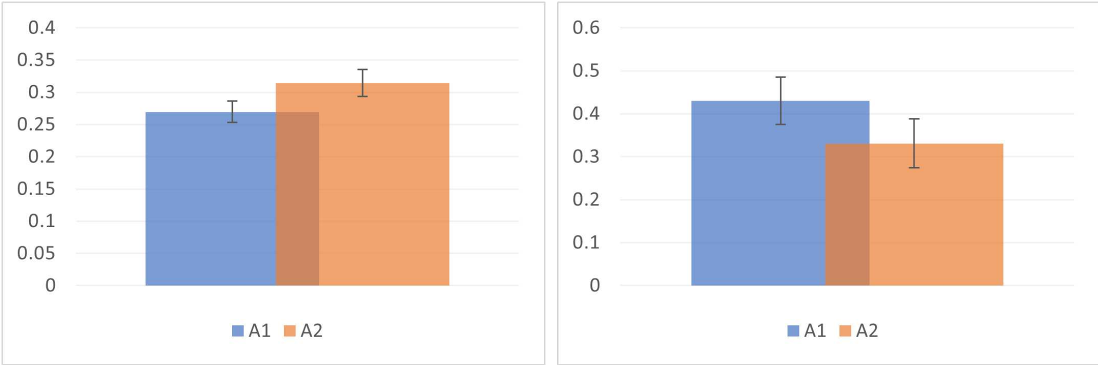
In Figure 5.14, contrasts between A2 and A3 audio exposures were revealed for TA alpha, underscoring a greater right hemisphere brain activation during audio stimulus processing. The study of functional lateralization in subcortical and cortical auditory structures supports the notion that interhemispheric asymmetry is contingent upon acoustic context. However, it is crucial to acknowledge that neither hemisphere can be deemed certainly dominant in processing individual aspects of audio stimuli. This notion aligns with the concept of dynamic functional localization, positing that the entire brain cooperatively engages in every function, due to distributed neuron ensembles [95].

**Table 5.2:** Repeated measured ANOVA - Kruskal-Walli's test results

| **EEG signal features** | $(\bar{f} \pm u_f)A_1$ | $(\bar{f} \pm u_f)A_2$ | $(\bar{f} \pm u_f)A_3$ | **H-statistic** | **p-value** |
|---|---|---|---|---|---|
| $Delta\_TP9$ | $0.6 \pm 0.07$ | $0.5 \pm 0.08$ | $0.5 \pm 0.06$ | 7.5 | 0.001 |
| $Delta\_AF7$ | $0.6 \pm 0.06$ | $0.5 \pm 0.09$ | $0.5 \pm 0.08$ | 5.1 | 0.009 |
| $Delta\_AF8$ | $0.7 \pm 0.06$ | $0.6 \pm 0.08$ | $0.6 \pm 0.09$ | 10.6 | 0.005 |
| $Delta\_TP10$ | $0.7 \pm 0.05$ | $0.5 \pm 0.07$ | $0.5 \pm 0.06$ | 16.7 | $< 0.001$ |
| $Theta\_AF7$ | $0.2 \pm 0.07$ | $0.2 \pm 0.06$ | $0.2 \pm 0.06$ | 3.3 | 0.04 |
| $Theta\_AF8$ | $0.3 \pm 0.07$ | $0.3 \pm 0.07$ | $0.2 \pm 0.07$ | 5.6 | 0.005 |
| $Theta\_TP10$ | $0.4 \pm 0.06$ | $0.3 \pm 0.06$ | $0.3 \pm 0.04$ | 11.1 | $< 0.001$ |
| $Alpha\_AF7$ | $0.4 \pm 0.05$ | $0.3 \pm 0.05$ | $0.3 \pm 0.04$ | 4.9 | 0.01 |
| $Alpha\_AF8$ | $0.4 \pm 0.05$ | $0.4 \pm 0.06$ | $0.3 \pm 0.06$ | 5.5 | 0.006 |
| $Beta\_TP10$ | $0.4 \pm 0.03$ | $0.4 \pm 0.04$ | $0.4 \pm 0.04$ | 4.5 | 0.01 |
| $Gamma\_TP9$ | $0.1 \pm 0.06$ | $0.03 \pm 0.05$ | $0.5 \pm 0.06$ | 3.3 | 0.04 |
| $Gamma\_TP10$ | $0.1 \pm 0.04$ | $0.05 \pm 0.05$ | $0.03 \pm 0.04$ | 4.8 | 0.01 |
| $Relative\_Delta\_TP10$ | $0.3 \pm 0.02$ | $0.3 \pm 0.02$ | $0.3 \pm 0.04$ | 3.6 | 0.03 |
| $Relative\_Alpha\_TP10$ | $0.3 \pm 0.02$ | $0.3 \pm 0.02$ | $0.3 \pm 0.02$ | 11.9 | $< 0.001$ |
| $Relative\_Gamma\_TP10$ | $0.04 \pm 0.02$ | $0.03 \pm 0.02$ | $0.01 \pm 0.02$ | 4.1 | 0.02 |
| $TA\_Theta$ | $-0.04 \pm 0.03$ | $-0.02 \pm 0.03$ | $0.01 \pm 0.02$ | 3.4 | 0.04 |
| $TA\_Alpha$ | $-0.06 \pm 0.03$ | $-0.003 \pm 0.03$ | $0.05 \pm 0.03$ | 8.3 | 0.001 |


**(a)** **(b)**

**Figure 5.12:** Histograms of Dwass-Seel-Critchlow-Fligner Pairwise Comparing Audio 1 and Audio 2 in terms of delta wave. The ranges above the bars represent the standard deviation. (a): related to the electrode AF8. (b): related to the electrode TP10.




(a) (b)

**Figure 5.13:** Histograms of Dwass-Seel-Critchlow-Fligner Pairwise Comparing Audio 1 and Audio 2 observed in the electrode TP10. The ranges above the bars represent the standard deviation. (a): relative alpha wave. (b): theta wave.

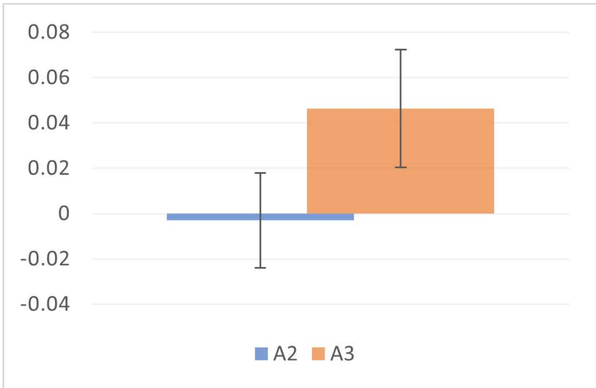



**Figure 5.14:** Histograms of Dwass-Seel-Critchlow-Fligner Pairwise Comparing Audio 2 and Audio 3 observed in terms of TA alpha. The ranges above the bars represent the standard deviation.

### Closure Discussion

The experimental results represent the dynamic nature of EEG responses, which exhibit discernible variations corresponding to different auditory inputs. Notably, the correlation between the acoustic stimuli and the EEG data establishes the feasibility of employing EEG measurements as a predictive tool for discerning jury responses during the listening phase of jury testing.

While the potential of EEG as a diagnostic tool for jury testing is promising, it is imperative to acknowledge certain inherent challenges. Variability in individual responses, influenced by a range of factors including personal predispositions, the intricate complexity of the human brain, and the potential for subjective experiences like daydreaming or mind wandering, introduce certain limitations. These intricacies underscore the necessity for a subtle approach in utilizing EEG data in this context. Despite the current limitations, the potential to enhance the reliability and precision of jury testing results is evident. The objective insights derived from EEG data offer a valuable complement to traditional assessments, providing a deeper understanding of the cognitive processes underlying juror decision-making.

# Chapter 6

# Conclusion

This thesis explored the enhancement of sound quality jury testing results through the integration of physiological and biometric measurements, specifically focusing on EEG and facial expression analysis of jurors during sound quality assessments. The study was structured into four key tests: two preliminary tests and two jury tests.

The preliminary tests, involving 16 and 40 participants respectively, aimed to assess facial expressions in response to audio and audio-visual stimuli. These tests highlighted the feasibility of using facial expression analysis tools to predict emotions induced by acoustic stimuli. Notably, the results demonstrated that visual support could strengthen emotion induction as reflected in facial expressions. The first jury test, fortified by facial expression analysis, revealed a correlation between facial expressions and self-reported engagement levels suggesting that facial expression analysis can enhance the reliability of jury test results. This method could be particularly useful in identifying biases or inconsistencies in jurors' responses, thereby improving the validity of sound quality assessments.

In the second jury test, involving 43 participants, EEG signals were recorded to analyze the correlation between brainwave patterns and acoustic sensations. The study discovered distinct patterns in the power spectral densities (PSDs) correlated with various audio stimuli. Specifically, the delta power in the TP10 and AF8 channels and Theta TP10 were higher under exposure to annoying audio (A1), suggesting these brainwaves are modulated by the perception of annoying audio stimuli. Conversely, relative alpha power was higher during exposure to pleasant sound (A2), and there was a notable right lateralization of the brain in processing sound content associated with negative emotions, particularly in the presence of background noise (A3). Accordingly, the EEG analysis provides insight into the neurological underpinnings of auditory perception. The distinct brainwave patterns associated with different audio stimuli open avenues for more objective and precise assessments of sound quality, offering a potential biomarker for these auditory experiences.

**Limitations of Emotion Recognition Systems from Facial Expressions:** The application of emotional facial expression recognition technologies in this thesis, while innovative, encounters inherent limitations. A critical review by Barrett et al. highlights a significant distinction: these technologies are more adept at detecting facial expressions than deciphering the underlying emotions. This discrepancy arises due to the multifaceted nature of emotional expression, which is influenced by cultural, situational, and individual variations. A universal expression of emotions, such as smiling for happiness or frowning for anger, does not necessarily hold true across different cultures or individuals. This variability challenges the generalization of facial expression recognition results, as noted in the referenced study [56]. Additionally, in the context of sound quality evaluation, jurors may maintain neutral facial expressions, complicating the identification of emotional responses to the sound. This study encountered such challenges, reflecting the findings of asymmetrical emotional responses to different affective media [1]. These limitations underscore the complexity of interpreting emotions from facial expressions and the need for caution in applying these technologies universally.

**Advantages and Limitations of Using EEG in Sound Quality Jury Testing:** The use of wearable EEG devices in this study offered non-invasive and comfortable monitoring of jurors' brain activity, providing an objective complement to subjective assessments. However, this approach is not without its limitations. Wearable EEG devices, while advantageous for their non-invasive nature, suffer from limited spatial resolution compared to more conventional EEG setups. This limitation can affect the precision and reliability of the data collected. External factors like noise and electromagnetic interference can also impact the accuracy of EEG measurements. Furthermore, individual differences in baseline brain activity and the variability in electrode placement add another layer of complexity, making it difficult to establish uniform benchmarks for EEG data interpretation. The practical aspect of integrating EEG with sound quality assessments also posed challenges. Ensuring that the earphone and EEG device were both comfortable and non-intrusive was crucial to maintain the jurors' concentration and deliberation effectiveness. This aspect of the study highlights the need for careful selection and integration of technology in jury testing environments.

## Future Research

This thesis lays the groundwork for the innovative use of EEG technology in jury tests for sound quality assessment. Future research could be done in addressing the identified limitations and challenges of integrating EEG into this context. Continuous refinement of EEG technology, informed by ongoing feedback, research findings, and technological advancements, is essential. This would involve enhancing the spatial resolution of wearable EEG devices, minimizing interference from external factors, and developing more consistent benchmarks for EEG data interpretation. Future studies could also explore ways to integrate EEG data with other biometric measures to provide a more comprehensive understanding of jurors' responses to sound quality.

Another avenue for future research is the improvement of reliability and validity in emotion recognition from facial expressions. This could be achieved by employing more context-specific datasets that take into account cultural, situational, and individual variability in emotional expression. An advanced application of the attention recognition system, currently used as a boolean value, could be developed to provide a more nuanced range between 0 and 1. Additionally, integrating a fuzzy filter to handle uncertainties arising from gaze direction and the degree of face rotation could further refine the accuracy of emotion recognition.

Despite the limitations, this study has demonstrated the potential of combining biometric measurements with jury tests in sound quality assessment. Future research should explore scenarios where a large sample of users is available or where specific sounds elicit distinct emotional states. This would allow for more robust data to objectively evaluate jurors' responses, reducing biases such as distraction. Moreover, this research opens new avenues in integrating sound quality into product design, emphasizing the need to consider the emotional responses of customers. This approach not only enhances the user experience but also offers valuable insights for designers and manufacturers in tailoring products to meet user preferences and expectations.

# References


[1] Reza Jamali, Gianmarco Battista, Milena Martarelli, Paolo Chiariotti, Deepti Shriram Kunte, Claudio Colangeli, and Paolo Castellini. Objective-subjective sound quality correlation performance comparison of genetic algorithm based regression models and neural network based approach. In *Journal of Physics: Conference Series*, volume 2041, page 012015. IOP Publishing, 2021.

[2] Davis E King. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10:1755–1758, 2009.

[3] Nash N Boutros, Silvana Galderisi, Oliver Pogarell, and Silvana Riggio. *Standard electroencephalography in clinical psychiatry: a practical handbook*. John Wiley & Sons, 2011.

[4] Mahsa Soufineyestani, Dale Dowling, and Arshia Khan. Electroencephalography (eeg) technology applications and available devices. *Applied Sciences*, 10(21):7453, 2020.

[5] Robert Oostenveld and Peter Praamstra. The five percent electrode system for high-resolution eeg and erp measurements. *Clinical neurophysiology*, 112(4):713–719, 2001.

[6] Yvonne Tran. Eeg signal processing for biomedical applications, 2022.

[7] Fei Ding, Wutong Xie, and Xiaoping Xie. Research on optimization of car door closing sound quality based on the integration of structural simulation and test. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 237(6):1378–1390, 2023.

[8] Danish Anas, Sharnappa Joladarashi, Ravikiran Kadoli, Pruthviraj Chavan, and Rajesh Bhangale. Development of psychoacoustics index for motorcycle exhaust noise by multiple regression method. *Materials Today: Proceedings*, 72:1197–1205, 2023.

[9] Thomas Carolus. Psychoacoustic assessment of fan noise. In *Fans: Aerodynamic Design-Noise Reduction-Optimization*, pages 131–139. Springer, 2023.

[10] Elif Özcan and Hendrik NJ Schifferstein. The effect of (un) pleasant sound on the visual and overall pleasantness of products. In *Proceedings of the 9th International Conference on Design & Emotion*, pages 601–606, 2014.

[11] Sang-Kwon Lee. Objective evaluation of interior sound quality in passenger cars during acceleration. *Journal of Sound and Vibration*, 310(1-2):149–168, 2008.

[12] Andrew M Willemsen and Mohan D Rao. Characterization of sound quality of impulsive sounds using loudness based metric. In *20th International Congress on Acoustics*, 2010.

[13] Gabriella Cerrato. Automotive sound quality–powertrain, road and wind noise. *Sound and Vibration*, 43(4):16–24, 2009.

[14] Ignazio Dimino, Claudio Colangeli, Jacques Cuenca, Pasquale Vitiello, and Mattia Barbarino. Active noise control for aircraft cabin seats. *Applied Sciences*, 12(11): 5610, 2022.

[15] Claudio Colangeli, Paolo Chiariotti, Gianmarco Battista, Paolo Castellini, and Karl Janssens. Clustering inverse beamforming for interior sound source localization: application to a car cabin mock-up. In *6th Berlin Beamforming Conference*, number 2016-D9, page 17, 2016.

[16] Gianmarco Battista, Paolo Chiariotti, Milena Martarelli, Paolo Castellini, Claudio Colangeli, and Karl Janssens. 3d acoustic mapping in automotive wind tunnel: Algorithm and problem analysis on simulated data. *Applied Sciences*, 11(7):3241, 2021.

[17] Fei Chen, Hongbo Shi, Jianjun Yang, Yu Lai, Jiahao Han, and Yimeng Chen. A new method to identifying optimal adjustment strategy when the car cockpit is uncomfortable: optimal state distance method. *PeerJ Computer Science*, 9:e1324, 2023.

[18] Shuai Zhang, Yipeng Li, Yuntao Cao, Liyou Xu, and Shixin Yuan. Research on active sound control in ev in-vehicle based on engine order sound fitting. *Advances in Mechanical Engineering*, 15(6):16878132231181282, 2023.

[19] Carla Julio da Silveira Brizon and Eduardo Bauzer Medeiros. Combining subjective and objective assessments to improve acoustic comfort evaluation of motor cars. *Applied Acoustics*, 73(9):913–920, 2012.

[20] Claudio Colangeli, Bernardo Lopes, Agnieszka Mroz, Karl Janssens, and Herman Van der Auweraer. Subjective and objective sound quality predictive models for the assessment of a propeller aircraft interior noise.

[21] Daniël Johannes Swart, Anriëtte Bekker, and Jörg Bienert. The subjective dimensions of sound quality of standard production electric vehicles. *Applied Acoustics*, 129:354–364, 2018.

[22] Mohd Jailani Mohd Nor, Mohammad Hosseini Fouladi, Hassan Nahvi, and Ahmad Kamal Ariffin. Index for vehicle acoustical comfort inside a passenger car. *Applied Acoustics*, 69(4):343–353, 2008.

[23] L Dedene, R Valgaeren, M Van Overmeire, and P Guillaume. Correlation of subjective sound perception of exhaust systems with sound quality metrics. In *Society for Experimental Mechanics, Inc, 16 th International Modal Analysis Conference.*, volume 2, pages 1007–1013, 1998.

[24] Penny Bergman, Anders Sköld, Daniel Västfjäll, and Niklas Fransson. Perceptual and emotional categorization of sound. *The Journal of the Acoustical Society of America*, 126(6):3156–3167, 2009.

[25] Xu Wang and Aleksandar Subic. Psychoacoustic modelling of vehicle side mirror power-fold actuator noise characteristics. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 225(6): 1419–1429, 2011.

[26] Conggan Ma, Qing Li, Qinghe Liu, Dafang Wang, Jiajun Gao, Huiyao Tang, and Yanhua Sun. Sound quality evaluation of noise of hub permanent-magnet synchronous motors for electric vehicles. *IEEE Transactions on Industrial Electronics*, 63(9):5663–5673, 2016.

[27] YS Wang, GQ Shen, H Guo, XL Tang, and T Hamade. Roughness modelling based on human auditory perception for sound quality evaluation of vehicle interior noise. *Journal of Sound and Vibration*, 332(16):3893–3904, 2013.

[28] Etienne Parizet, Erald Guyader, and Valery Nosulenko. Analysis of car door closing sound quality. *Applied acoustics*, 69(1):12–22, 2008.

[29] Eui-Youl Kim, Tae Jin Shin, and Sang-Kwon Lee. Design of a decision error model for reliability of jury evaluation and its experimental verification. *Applied acoustics*, 74(6):789–802, 2013.

[30] Thomas Robotham, Olli S Rummukainen, Miriam Kurz, Marie Eckert, and Emanuël AP Habets. Comparing direct and indirect methods of audio quality evaluation in virtual reality scenes of varying complexity. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2091–2101, 2022.

[31] Enlai Zhang, Liang Hou, Chao Shen, Yingliang Shi, and Yaxiang Zhang. Sound quality prediction of vehicle interior noise and mathematical modeling using a back

propagation neural network (bpnn) based on particle swarm optimization (pso). *Measurement Science and Technology*, 27(1):015801, 2015.

[32] Norm Otto, Scott Amman, Chris Eaton, and Scott Lake. Guidelines for jury evaluations of automotive sounds. *Sound and Vibration*, 35(4):24–47, 2001.

[33] Xuejun Hu, Qi Meng, Jian Kang, and Yan-jun Han. Psychological assessment of an urban soundscape using facial expression analysis. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 259, pages 5807–5815. Institute of Noise Control Engineering, 2019.

[34] Baoxiang Huang, Zhenkuan Pan, and Binsen Zhang. A virtual perception method for urban noise: The calculation of noise annoyance threshold and facial emotion expression in the virtual noise scene. *Applied Acoustics*, 99:125–134, 2015.

[35] Arghavan Hadinejad, Brent D Moyle, Noel Scott, and Anna Kralj. Emotional responses to tourism advertisements: the application of facereader™. *Tourism Recreation Research*, 44(1):131–135, 2019.

[36] Qi Meng, Xuejun Hu, Jian Kang, and Yue Wu. On the effectiveness of facial expression recognition for evaluation of urban sound perception. *Science of The Total Environment*, 710:135484, 2020.

[37] Sang Hee Park, Pyoung Jik Lee, Timothy Jung, and Alasdair Swenson. Effects of the aural and visual experience on psycho-physiological recovery in urban and rural environments. *Applied Acoustics*, 169:107486, 2020.

[38] Turgut Özseven. Investigation of the effect of spectrogram images and different texture analysis methods on speech emotion recognition. *Applied Acoustics*, 142: 70–77, 2018.

[39] Ying Liu, Zixuan Wang, and Ge Yu. The effectiveness of facial expression recognition in detecting emotional responses to sound interventions in older adults with dementia. *Frontiers in Psychology*, 12:707809, 2021.

[40] Alessia Frescura and Pyoung Jik Lee. Emotions and physiological responses elicited by neighbours sounds in wooden residential buildings. *Building and Environment*, 210:108729, 2022.

[41] Massimiliano Zampini and Charles Spence. The role of auditory cues in modulating the perceived crispness and staleness of potato chips. *Journal of sensory studies*, 19(5):347–363, 2004.

[42] Maria Rådsten-Ekman, Östen Axelsson, and Mats E Nilsson. Effects of sounds from water on perception of acoustic environments dominated by road-traffic noise. *Acta Acustica united with Acustica*, 99(2):218–225, 2013.

[43] James A Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294, 1977.

[44] Andrea Generosi, José Yuri Villafan, Luca Giraldi, Silvia Ceccacci, and Maura Mengoni. A test management system to support remote usability assessment of web applications. *Information*, 13(10):505, 2022.

[45] Paul Ekman. Universals and cultural differences in facial expressions of emotion. In *Nebraska symposium on motivation*. University of Nebraska Press, 1971.

[46] Paul Ekman et al. Basic emotions. *Handbook of cognition and emotion*, 98(45-60): 16, 1999.

[47] Shan Li and Weihong Deng. Deep facial expression recognition: A survey. *IEEE transactions on affective computing*, 13(3):1195–1215, 2020.

[48] Silvia Ceccacci, Maura Mengoni, Generosi Andrea, Luca Giraldi, Giuseppe Carbonara, Andrea Castellano, and Roberto Montanari. A preliminary investigation towards the application of facial expression analysis to enable an emotion-aware car interface. In *Universal Access in Human-Computer Interaction. Applications and Practice: 14th International Conference, UAHCI 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22*, pages 504–517. Springer, 2020.

[49] Mihai Gavrilescu and Nicolae Vizireanu. Predicting depression, anxiety, and stress levels from videos using the facial action coding system. *Sensors*, 19(17):3693, 2019.

[50] Mohammed F Alsharekh. Facial emotion recognition in verbal communication based on deep learning. *Sensors*, 22(16):6105, 2022.

[51] Szwoch Wioleta. Using physiological signals for emotion recognition. In *2013 6th international conference on human system interactions (HSI)*, pages 556–561. IEEE, 2013.

[52] Nanouk Verhulst, Iris Vermeir, Hendrik Slabbinck, Bart Lariviere, Maurizio Mauri, and Vincenzo Russo. A neurophysiological exploration of the dynamic nature of emotions during the customer experience. *Journal of Retailing and Consumer Services*, 57:102217, 2020.

[53] Bernardo Lopes, Claudio Colangeli, Karl Janssens, Agnieszka Mroz, and Herman Van der Auweraer. Neural network models for the subjective and objective assessment of a propeller aircraft interior sound quality. In *INTER-NOISE and NOISE-CON congress and conference proceedings*, volume 259, pages 4124–4135. Institute of Noise Control Engineering, 2019.

[54] Mara Münder and Claus-Christian Carbon. Howl, whirr, and whistle: The perception of electric powertrain noise and its importance for perceived quality in electrified vehicles. *Applied Acoustics*, 185:108412, 2022.

[55] ZHANG Xiaojuan, LIU Yan, and ZHANG Changbin. Sound quality subjective evaluation analysis of noise inside high-speed trains. In *1st International Conference on Mechanical Engineering and Material Science (MEMS 2012)*, pages 97–99. Atlantis Press, 2012.

[56] ZHANG Xiaojuan, LIU Yan, and ZHANG Changbin. Sound quality subjective evaluation analysis of noise inside high-speed trains. In *1st International Conference on Mechanical Engineering and Material Science (MEMS 2012)*, pages 97–99. Atlantis Press, 2012.

[57] Enlai Zhang, Liang Hou, Chao Shen, Yingliang Shi, and Yaxiang Zhang. Sound quality prediction of vehicle interior noise and mathematical modeling using a back propagation neural network (bpnn) based on particle swarm optimization (pso). *Measurement Science and Technology*, 27(1):015801, 2015.

[58] Norm Otto, Scott Amman, Chris Eaton, and Scott Lake. Guidelines for jury evaluations of automotive sounds. *SAE transactions*, pages 3015–3034, 1999.

[59] J Satheesh Kumar and P Bhuvaneswari. Analysis of electroencephalography (eeg) signals and its categorization–a study. *Procedia engineering*, 38:2525–2536, 2012.

[60] Uwe Herwig, Peyman Satrapi, and Carlos Schönfeldt-Lecuona. Using the international 10-20 eeg system for positioning of transcranial magnetic stimulation. *Brain topography*, 16:95–99, 2003.

[61] Umit Budak, Varun Bajaj, Yaman Akbulut, Orhan Atila, and Abdulkadir Sengur. An effective hybrid model for eeg-based drowsiness detection. *IEEE sensors journal*, 19(17):7624–7631, 2019.

[62] Nathaniel J Zuk, Emily S Teoh, and Edmund C Lalor. Eeg-based classification of natural sounds reveals specialized responses to speech and music. *NeuroImage*, 210:116558, 2020.

[63] W Grey Walter, VJ Dovey, and H Shipton. Analysis of the electrical response of the human cortex to photic stimulation. *Nature*, 158(4016):540–541, 1946.

[64] Alastair Compston. The berger rhythm: potential changes from the occipital lobes in man, by ed adrian and bhc matthews (from the physiological laboratory, cambridge). brain 1934: 57; 355–385. *Brain*, 133(1):3–6, 2010.

[65] M Trimmel, M Kundi, G Binder, E Groll-Knapp, and M Haider. Combined effects of mental load and background noise on cns activity indicated by brain dc potentials. *Environment international*, 22(1):83–92, 1996.

[66] Louis A Schmidt and Laurel J Trainor. Frontal brain electrical activity (eeg) distinguishes valence and intensity of musical emotions. *Cognition & Emotion*, 15 (4):487–500, 2001.

[67] Michinori Kabuto, Takayuki Kageyama, and Hiroshi Nitta. Eeg power spectrum changes due to listening to pleasant musics and their relation to relaxation effects. *Nippon Eiseigaku Zasshi (Japanese Journal of Hygiene)*, 48(4):807–818, 1993.

[68] Young Joon Lee, Tae Jin Shin, and Sang Kwon Lee. Sound quality analysis of a passenger car based on electroencephalography. *Journal of Mechanical Science and Technology*, 27:319–325, 2013.

[69] James L Walker. Subjective reactions to music and brainwave rhythms. *Physiological Psychology*, 5(4):483–489, 1977.

[70] W Krebs, M Balmer, and E Lobsiger. A standardised test environment to compare aircraft noise calculation programs. *Applied acoustics*, 69(11):1096–1100, 2008.

[71] David Matsumoto, Dacher Keltner, Michelle N Shiota, MAUREEN O'Sullivan, and Mark Frank. Facial expressions of emotion. *Handbook of emotions*, 3:211–234, 2008.

[72] Stefan Winkler and Ramanathan Subramanian. Overview of eye tracking datasets. In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 212–217. IEEE, 2013.

[73] Paul Ekman and Wallace V Friesen. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*, 1978.

[74] A Talipu, Andrea Generosi, Maura Mengoni, and Luca Giraldi. Evaluation of deep convolutional neural network architectures for emotion recognition in the wild. In *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)*, pages 25–27. IEEE, 2019.

[75] Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1):18–31, 2017.

[76] Alex Altieri, Silvia Ceccacci, and Maura Mengoni. Emotion-aware ambient intelligence: Changing smart environment interaction paradigms through affective computing. In *Distributed, Ambient and Pervasive Interactions: 7th International Conference, DAPI 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*, pages 258–270. Springer, 2019.

[77] I Velasco, A Sipols, C Simon De Blas, L Pastor, and S Bayona. Motor imagery eeg signal classification with a multivariate time series approach. *BioMedical Engineering OnLine*, 22(1):1–24, 2023.

[78] Jonas Egeler. *Physiological correlates of psychoacoustic annoyance*. PhD thesis, Technische Hochschule Ingolstadt, 2021.

[79] Julia WY Kam, Sandon Griffin, Alan Shen, Shawn Patel, Hermann Hinrichs, Hans-Jochen Heinze, Leon Y Deouell, and Robert T Knight. Systematic comparison between a wireless eeg system with dry electrodes and a wired eeg system with wet electrodes. *NeuroImage*, 184:119–129, 2019.

[80] Uwe Herwig, Peyman Satrapi, and Carlos Schönfeldt-Lecuona. Using the international 10-20 eeg system for positioning of transcranial magnetic stimulation. *Brain topography*, 16:95–99, 2003.

[81] Pierre Gloor. Hans berger on electroencephalography. *American Journal of EEG Technology*, 9(1):1–8, 1969.

[82] Olave E Krigolson, Chad C Williams, Angela Norton, Cameron D Hassall, and Francisco L Colino. Choosing muse: Validation of a low-cost, portable eeg system for erp research. *Frontiers in neuroscience*, 11:109, 2017.

[83] Amira E Youssef, Hebatalla T Ouda, and Mohamed Azab. Muse: a portable cost-efficient lie detector. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 242–246. IEEE, 2018.

[84] Elena Ratti, Shani Waninger, Chris Berka, Giulio Ruffini, and Ajay Verma. Comparison of medical and consumer wireless eeg systems for use in clinical trials. *Frontiers in human neuroscience*, 11:398, 2017.

[85] Silvia Angela Mansi, Giovanni Barone, Cesare Forzano, Ilaria Pigliautile, Maria Ferrara, Anna Laura Pisello, and Marco Arnesano. Measuring human physiological indices for thermal comfort assessment through wearable devices: A review. *Measurement*, 183:109872, 2021.

[86] Emily S Kappenman and Steven J Luck. The effects of electrode impedance on data quality and statistical significance in erp recordings. *Psychophysiology*, 47(5): 888–904, 2010.

[87] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191, 2007.

[88] Hossein Arsham and Miodrag Lovric. Bartlett's test. *International encyclopedia of statistical science*, 2:20–23, 2011.

[89] Michal Brzezinski. The chen–shapiro test for normality. *The Stata Journal*, 12(3): 368–374, 2012.

[90] Paul A Games. Alternative analyses of repeated-measure designs by anova and manova. In *Statistical methods in longitudinal research*, pages 81–121. Elsevier, 1990.

[91] Critchlow E Douglas and Fligner A Michael. On distribution-free multiple comparisons in the one-way analysis of variance. *Communications in Statistics-Theory and Methods*, 20(1):127–139, 1991.

[92] Radosław Wolniak. Performance evaluation in iso 9001: 2015. *Zeszyty Naukowe. Organizacja i Zarzadzanie/Politechnika Śląska*, 2021.

[93] Zheng-Guang Li, Guo-Qing Di, and Li Jia. Relationship between electroencephalogram variation and subjective annoyance under noise exposure. *Applied acoustics*, 75:37–42, 2014.

[94] Margaret M Bradley, Maurizio Codispoti, Bruce N Cuthbert, and Peter J Lang. Emotion and motivation i: defensive and appetitive reactions in picture processing. *Emotion*, 1(3):276, 2001.

[95] Svetlana F Vaitulevich, Ekaterina A Petropavlovskaya, Lidiya B Shestopalova, and Nikolay I Nikitin. Functional interhemispheric asymmetry of human brain and audition. *Human Physiology*, 45:202–212, 2019.