

**SYMPATHETIC LOADING IN
CRITICAL TASKS**

A Dissertation

Presented to

the Faculty of the Department of Computer Science

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Malcolm Dcosta

December 2015

SYMPATHETIC LOADING IN CRITICAL TASKS

Malcolm Dcosta

APPROVED:

Advisor Dr. Ioannis Pavlidis
Dept. of Computer Science, UH

Dr. Christoph Eick
Dept. of Computer Science, UH

Dr. Carla Sharp
Dept. of Psychology, UH

Dr. Tammy Tolar
Dept. of Educational Psychology, UH

Dr. Nikolaos Tsekos
Dept. of Computer Science, UH

Dr. Ricardo Vilalta
Dept. of Computer Science, UH

Dean, College of Natural Sciences and Mathematics

This research was supported by grants from the Defense Academy for Credibility Assessment (DACA), National Science Foundation (NSF) under grant # IIS-1249208 entitled ‘The Effect of Stress and the Role of Computer Mediation on Exam Performance’ and Toyota Motor Engineering & Manufacturing, Inc. The views expressed in this work do not necessarily reflect those of the funding agencies.

I am very appreciative of Dr. Ioannis Pavlidis who has been guiding my research for the past six years. He has always recommended interesting approaches to research as well as provided me numerous opportunities to improve my leadership skills. The financial support has also been instrumental to pursuing this degree.

I have deep gratitude towards Dr. Dvijesh Shastri who mentored me and trained me in the ways of the lab. He has always shared his views and experience in several of the projects that I worked upon. Similarly, I also acknowledge the insights and efforts of Dr. Panagiotis Tsiamyrtzis for helping me with statistical analysis during my research.

I also want to thank my other committee members, Dr. Christoph Eick, Dr. Carla Sharp, Dr. Tammy Tolar, Dr. Nikolaos Tsekos and Dr. Ricardo Vilalta for their advice, analysis and feedback making this research possible.

I also have to thank all the lab members, who have been colleagues and friends and provided a healthy learning environment.

Most of all I have to thank God for the abundantly blessing he has bestowed upon me. I also am very thankful to my parents and sister for always loving and supporting me. Additionally, as an international student studying abroad, I am grateful to the many friends I made at the Catholic Newman Center and International Christian Fellowship who have been a second family to me.

SYMPATHETIC LOADING IN CRITICAL TASKS

An Abstract of a Dissertation
Presented to
the Faculty of the Department of Computer Science
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Malcolm Dcosta
December 2015

Abstract

In this dissertation I developed or perfected unobtrusive methods to quantify sympathetic arousals. Furthermore, I used these methods to study the sympathetic system's role on critical activities, arriving at intriguing conclusions. Sympathetic arousals occur during states of mental, emotional, and/or sensorimotor strain resulting from adverse or demanding circumstances. They are key elements of human physiology's coping mechanism, shoring up resources to a good effect. When the intensity and duration of these arousals are overwhelming, however, then they may block memory and disrupt rational thought or actions at the moment they are needed the most.

Arousals abound in three types of critical activities: high-stakes situations, challenging tasks, and critical multitasking. Accordingly, my research was based on three studies representative of these three activity types: 'Subject Screening', 'Educational Exam', and 'Distracted Driving'. In the first study I investigated the association of sympathetic arousals with deceptive behavior in interrogations. In the second study, I investigated the relationship between sympathetic arousals and exam performance. In the third study, I investigated the interaction between sympathetic arousals and driving performance under cognitive, emotional, and sensorimotor distractions.

In the interrogation study, I used for the first time a contact-free electrodermal activity measurement method to quantify arousals. The method detected deceptive behavior based on differential sympathetic responses in well-structured interviews. In the exam study, I documented that sympathetic arousals positively correlate with students' exam performance, dispelling the myth of 'easy going' super achievers. Finally, in the driving study, my results revealed that not only apparent sensorimotor

stressors (texting while driving) but also hidden stressors (cognitive or emotional) could have a significant effect on driving performance.

Contents

1	Introduction	1
2	Background	3
2.1	Sympathetic Arousals	3
2.2	Thermal Imaging	6
2.3	Q Sensor	11
2.4	Zephyr BioHarness	12
3	Experiments	14
3.1	Deception Detection Study	17
3.1.1	Overview	17
3.1.2	The Interrogation	20
3.2	Exam Study	23
3.2.1	Overview	23
3.2.2	Measurements	24
3.2.3	Experiment	25
3.3	Driving Study	28
3.3.1	Overview	28
3.3.2	Experiment	28

4	Methods	38
4.1	Deception Detection Study	38
4.1.1	Interview Segmentation	38
4.1.2	Tracking Region of Interest	42
4.1.3	Perspiration Signal Extraction	44
4.1.4	Signal Processing	45
4.2	Exam Study	60
4.3	Driving Study	63
5	Results	66
5.1	Deception Detection Study	66
5.2	Exam Study	74
5.3	Driving Study	81
6	Conclusion	91
6.1	Deception Detection Study	91
6.2	Exam Study	94
6.3	Driving Study	96
7	Appendix	99
7.1	Overview	99
7.2	Experimental Design	102
7.3	Methods	105
7.4	Results	107
7.5	Discussion	111
	Bibliography	120

List of Figures

2.1	Cross section of the skin.	5
2.2	The electromagnetic spectrum. The mid-wave infrared spectrum is highlighted by the arrow.	6
2.3	Black body spectrum. The graph shows a significant amount of energy radiation change in the thermal infrared band (300-1400 nm) due to change in temperature.	7
2.4	Custom thermal imaging system, ATHEMOS. The custom thermal imaging system developed by our group integrates a computer, a thermal camera, and several peripheral hardware components.	9
2.5	Thermal image. A sample thermal image of a subject. The bar on the right side shows the mapping between color and temperature in degrees centigrade.	10
2.6	The thermal imaging cameras used in the experiments	10
2.7	The Q Sensor worn on the palm, with the palm strap.	11
2.8	The Zephyr BioHarness worn on the chest.	13
3.1	Diagram of the experimental setup.	19
3.2	A view of the study in progress. 4 students take their examination while being monitored with wearable sensors, on the palm, ankle and a chest strap. Their facial expressions and exam booklets are recorded with video cameras.	26
3.3	The timeline of administration of the various measurements throughout the course of the semester.	27

3.4	The driving simulator experimental setup. A thermal imaging camera mounted above the center TV screen records the subjects face. The virtual car is controlled by an accelerator pedal and a brake pedal, and steered with a steering wheel.	29
3.5	The practice drive scenario.	31
3.6	The relaxing drive scenarios.	33
3.7	The four loaded drive scenarios.	35
4.1	Audio segmentation of the interview questions.	39
4.2	The tracking region of interest is depicted in 4.2(a). Within the tracking region of interest the measurement region is selected as shown in 4.2(b) from which the signal is extracted.	42
4.3	Extraction of the perspiration signal.	44
4.4	Figure 4.4(a) (top) and 4.4(b) (top) represents two perspiration signals whose initial intensities are different by an order of magnitude. 4.4(a) (bottom) and 4.4(b) (bottom) show the normalized perspiration signals within the same intensity range.	46
4.5	Symmetrically extended signal.	47
4.6	Effect of symmetric extension of signal on wavelet energy.	48
4.7	Figure (a) represents the original perspiration signal, figure (b) represents impulse cleaning applied on (a). Figure (c) and (d) represent normalization applied on (a) and (b) respectively.	50
4.8	Effect of impulse on computed wavelet energy.	51
4.9	Raw, Impulse cleaned and Noise reduced signal.	53
4.10	Effect of noise cleaning on wavelet energy.	53
4.11	Noise-reduced signal superimposed with the wavelet fitting (top). Wavelet energy graph (bottom). The peak at the lower wavelet scale corresponds to high-frequency nose while the peak at the higher wavelet scale corresponds to true perspiration phenomenon.	55

4.12	The figure compares the wavelet energies of the two mother wavelets and it is observable that while using Morlet mother wavelet, the global maxima is clearly distinguishable.	56
4.13	The figure points out correspondence of low and high wavelet scales, with high-frequency noise and global trends in the perspiration signal. (a) top: Perspiration signal with dominant high-frequency noise. (a) bottom: Wavelet energy corresponding to the (a) top. (b) top: Perspiration signal containing a very long cycle. (b) bottom: Wavelet energy corresponding to the (a) bottom.	57
4.14	Translation of wavelet scale at maximum wavelet energy to perspiration frequency values. (top) Wavelet energy graph with maximum energy at scale 227. (bottom) Table containing perspiration frequency values for each segment for all subjects.	59
4.15	Video annotation of subjects actions at every instance when they start and end an action.	60
4.16	(a) - Declarative Thinking Question. (b) - Critical Thinking Question	61
4.17	Physiological signal segmentation and annotation.	62
4.18	A screenshot of SubjectBook showing Subject 29, LD_E . The signal on the top is the perinasal perspiration and the signals on the bottom are the output of the various simulator parameters. Between the signals, we see a synchronized playback of videos corresponding the the black vertical marker. The highlighted section annotates the loaded part of the drive.	65
5.1	Distribution of SAI scores before (Pre) and after (Post) the exams for all subjects.	74
5.2	Figure 5.2(a) represent scatterplots of exam scores vs. student GPA for all 5 exams attempted by the 23 subjects. Figure 5.2(b) represent the scatterplots of the mean exam scores vs. student GPA.	76
5.3	5.3(a) represent the percent score distribution of all students in the class. 5.3(b) represent the percent score of the 23 students who participated in the study.	77

5.4	Mean physiological and kinetic indicators of sympathetic arousal per student.	78
5.5	Distribution of mean palm EDA per exam per GPA.	80
5.6	Mean palm EDA per student per GPA.	80
5.7	Mean palm EDA per student against mean percent score.	80
5.8	Paired tests for the explanatory (perinasal perspiration) and response I (steering) variables in each phase of the cognitively, emotionally, and sensorimotorically loaded drives.	89
5.9	Paired tests for the response II (lane departure) variable in each phase of the cognitively, emotionally, and sensorimotorically loaded drives.	90
7.1	Experiment and sample outcome. A, Experimental setup, demonstrating all seven nodes. B, Raw EDA signal on the fingers of subject S003, captured via the conventional sensing device. The dotted lines mark the occurrence of the three stimuli. Only the last minute of the baseline period is depicted to economize space. C, The signal after noise filtering. There are multiple arousals after each stimulus; circles mark onsets, triangles mark peaks, and crosses mark offsets; t_{on} , denotes the time of Onset occurrence, t_p the time of Peak occurrence, and t_{off} the time of Offset occurrence; A stands for the arousals amplitude.	104
7.2	C and M EDA responses on the fingers per subject. From the four minutes of baseline only the last minute is depicted in the graphs to economize space. Vertical dotted lines identify stimuli times and inverted triangles denote peaks, facilitating arousal comparison. The M EDA signal for subject S009 was not collected due to technical reasons. The graphs confirm qualitatively the high responsiveness of the fingers location and the agreement between the two sensor modalities.	113

7.3	<p><i>C</i> and <i>M</i> EDA responses on the palm per subject. From the four minutes of baseline only the last minute is depicted in the graphs to economize space. Vertical dotted lines identify stimuli times and inverted triangles denote peaks, facilitating arousal comparison. The <i>M</i> EDA signal for subject S005 and both <i>C</i> and EDA signals for subject S003 were not collected due to technical reasons. The graphs confirm qualitatively the high responsiveness of the palm location and the agreement between the two sensor modalities.</p>	114
7.4	<p><i>C</i> and <i>M</i> EDA responses on the wrist per subject. From the four minutes of baseline only the last minute is depicted in the graphs to economize space. Vertical dotted lines identify stimuli times and inverted triangles denote peaks, facilitating arousal comparison. A, Signals manifesting sympathetic responses on the wrist. B, Signals manifesting the absence of sympathetic responses on the wrist, which were excluded from further processing and analysis. The graphs confirm qualitatively the low responsiveness of the wrist location, a result that was documented quantitatively in the text.</p>	115
7.5	<p><i>M</i> EDA responses on the sole per subject. From the four minutes of baseline only the last minute is depicted in the graphs to economize space. Vertical dotted lines identify stimuli times and inverted triangles denote peaks, facilitating arousal comparison. The graphs confirm qualitatively the responsiveness of the sole location, a result that was documented quantitatively in the text. However, amplitude correlations with the finger and palm locations are not strong, suggesting that precise EDA measurements on the sole may be challenging. . . .</p>	116
7.6	<p>Responsiveness of subjects per node. Each bar chart depicts the distribution of the number of subjects over different levels of response (i.e., number of recorded peaks in the node). Red bars indicate the number of completely nonresponsive subjects for the specific node. The maximum likelihood estimates (MLE, $\hat{\theta} = X/n$) for getting at least one peak appear on the upper right corner of the node's panel.</p>	117

7.7	Arousal timing relationships between nodes. BELOW THE DIAGONAL: Scatterplots of Onset (t_{on}), Peak, (t_p), and Offset (t_{off}) times for the various node pairs. ABOVE THE DIAGONAL: The correlation coefficients that correspond to the strength of the linear relationships depicted in the scatter-plots below the diagonal (all are significant, $p < 0.01$).	118
7.8	Arousal intensity relationships between nodes. BELOW THE DIAGONAL: Scatterplots of amplitude (A) for the various node pairs. ABOVE THE DIAGONAL: The correlation coefficients that correspond to the strength of the linear relationships depicted in the scatter-plots below the diagonal (coefficients in bold are significant, $p < 0.01$).	119

List of Tables

- 3.1 Reward Table 19
- 4.1 Constitution of segments 41
- 5.1 Prediction on the validation set 67
- 5.2 Prediction of Test Set 1 70
- 5.3 Prediction of Test Set 2 72

Chapter 1

Introduction

'*Stress*' is a state of mental, emotional or motoric strain which results from adverse of demanding circumstances. Its influence is seen both in the psychological as well as the physiological realm. Its effects on physiology is a result of the bodys mechanism to cope with the imminent threat by activating the sympathetic nervous system (*SNS*), which in turn triggers the adrenergic and cholinergic receptors. The former result in elevation of the cardiovascular operation, while the latter result in activation of sweat glands on the fingers, perinasal area and other peripheral parts of the body.

A lot is known and researched about the impact of stress on performance [32] on critical tasking. These effects extend to cognitive functions too. Stress is present, and part of most daily activities, however its manifestation is seen especially during challenging or engaging tasks and during situations where the stakes are high. It is extremely difficult to isolate specific causes of stress on cognitive functions as the

effects of stress usually occur as a combination of several factors, sometimes triggered by specific tasks, and sometimes external [22]. Sandi and Pinelo-Nava propose that three factors related to delivery of stress have distinct effects on cognitive functions [23]. These are a) the source of the stress, whether internal or external, b) the duration of the stressor, whether acute or chronic and c) the intensity of the stressor.

Measurement of stress and its quantification is another delicate operation which needs to be carefully considered in developing psychological experiments. Care must be taken to ensure that the measurement process does not act as an additional source of stress to the subjects. Ideal conditions for achieving this goal is to use unobtrusive or minimally obtrusive devices. Thermal imagery is one such measurement device capable of measuring at a distance. In [29], Tsiamyrtzis et al. use thermal imagery of the face in a deception-detection application in an interrogation. This method is unlike traditional means of measuring physiology with various contact sensors probing the subject. Pavlidis et al. [18] used a similar methods to extract perinasal perspiration. This was indicative of emotional perspiration modulated by stress as surgeons performed training exercises while being evaluated on their precision and speed in completing the task. In the recent years the introduction of the Q-Sensor [20] has allowed measurement of emotions in settings outside the laboratory environment. In [15], Mc Duff et al. continuously logged users emotional state over long periods of time, allowing them to reflect upon it.

Chapter 2

Background

2.1 Sympathetic Arousals

When faced with stressful events, a variety of physiological changes occur in an individual that are designed to assist in coping with the events. These physiological changes aim to promote survival by activating the necessary attentional and energetic recourses needed to contend with the immediate threat. There are two major physiological systems involved in the stress responses: the Sympathetic-Adrenal-Medullary (SAM) system and the Hypothalamic-Pituitary-Adrenal (HPA) axis. The SAM system is the rapid-response component of stress-system activation, and causes a nearly immediate release of epinephrine (adrenaline) and norepinephrine, and the initial increase in vigilance and arousal that accompany the perception of a potential threat. This immediately produced stress is called *instantaneous stress*. If evaluation of the

threat suggests that it is warranted, the SAM response is reinforced and extended by the activity of the slower but longer acting HPA axis. The HPA response to stress begins in the brain with the release of corticotropin releasing hormone (CRH) from the hypothalamus, which stimulates the pituitary gland to release adrenocorticosteroids (ACTH) into general circulation. ACTH in turn stimulates the release of corticosteroids from the adrenal cortex into the bloodstream. The signal to decrease or shut down further production of cortisol comes from negative feedback of cortisol to the brain, especially to receptors in the hippocampus, hypothalamus and pituitary, with high circulating levels of the hormone suppressing further release of CRH, ACTH and corticosteroids. The long-lasting stress is referred to as *sustained stress*.

Sweat glands cover a large portion of the body and are present under the epidermis. They are coiled tubular structures and perform two vital functions; regulation of the body temperature and providing emotional response to stimuli. Sweat glands are of two kinds; eccrine and apocrine and both are controlled by the sympathetic nervous system. The perspiration due to thermoregulatory effect occurs on most parts of the body, while emotional perspiration/sympathetic arousal usually occurs on the palms or soles of the feet.

It can be very useful to observe the activation of sweat glands in the event of stress in order to measure rate of perspiration which can be linked to emotional stress. Perspiration can be categorized as physical and emotional. Sympathetic arousal occurs as a sudden outburst, while the onset of physical perspiration is gradual.

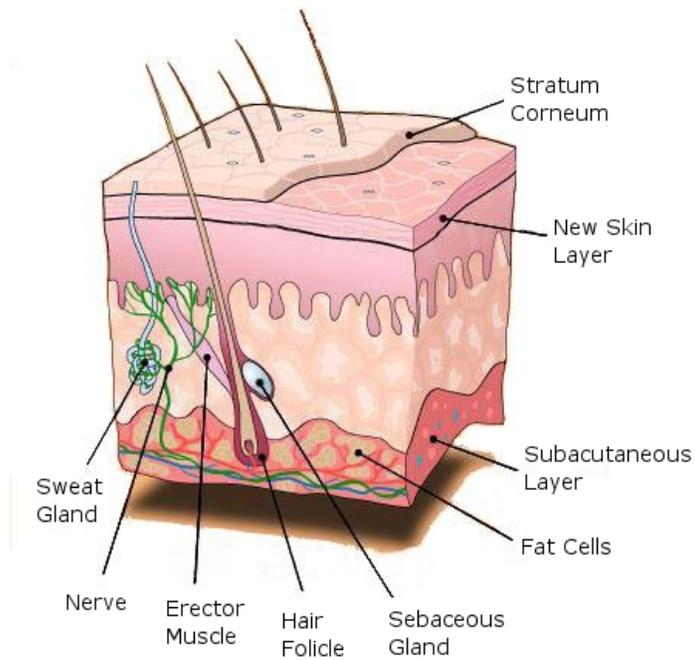


Figure 2.1: Cross section of the skin.

Also, once sympathetic arousal starts, it continues to secrete at a rate that can be correlated to the intensity of stimulation, and lasts no longer than the duration of the stimulation. Another point of distinction between sympathetic arousal and physical perspiration is that sympathetic arousal is transient and highly localized. In [12] T. Kamei *et al.* present sweating as an indicator which is capable of identifying, surprise, or emotional stress. In [26] Shastri *et al.* did a comparative analysis of the EDA and the corresponding measurement at the perinasal region.

2.2 Thermal Imaging

Our research mainly focuses on facial physiology, monitored in the mid-wave thermal infrared spectrum. The thermal infrared spectrum is composed of electromagnetic energy with wavelengths between three and eight micrometers (see Figure 2.3). In contrast to energy in the visible band, which is reflected off surfaces, energy in the thermal infrared spectrum of the electromagnetic spectrum represents energy radiated by objects.

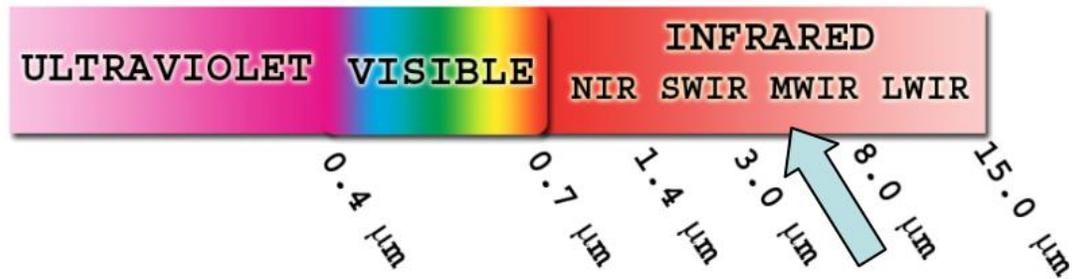


Figure 2.2: The electromagnetic spectrum. The mid-wave infrared spectrum is highlighted by the arrow.

All objects at finite temperature emit non-trivial amounts of electromagnetic radiation in the thermal infrared (3-14 m). According to Planck's law, the power of emission $M(\lambda, T)$ at a specific wavelength depends on the object's temperature as follows:

$$M(\lambda, T) = \frac{c_1}{\lambda^5} \left(\frac{1}{e^{(c_2/\lambda T)} - 1} \right) \frac{W}{m^2 - \mu m}, \quad (2.1)$$

where the first radiation constant, $c_1 = 3.7411 \times 10^8 W - \mu m^4 / m^2$, the second constant $c_2 = 1.4388 \times 10^4 \mu m - K$, and λ is the wavelength expressed in μm . As temperature increases, radiation ($M(\lambda, T)$) increases (see figure 2.3).

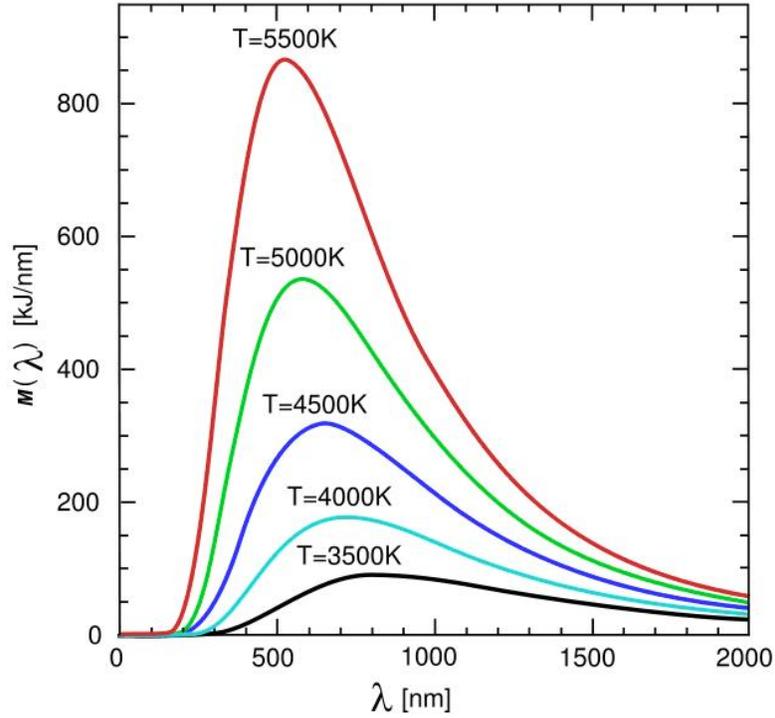


Figure 2.3: Black body spectrum. The graph shows a significant amount of energy radiation change in the thermal infrared band (300-1400 nm) due to change in temperature.

According to the Stefan-Boltzmann law, the power of emission over several wavelengths can be obtained by integrating equation 2.2:

$$M(\Delta\lambda, T) = \int_{\lambda_1}^{\lambda_2} M(\lambda, T)d\lambda, \quad (2.2)$$

where $\Delta\lambda = \lambda_1 - \lambda_2$. Since our sensing device operates in the Mid-Wave Infrared (MWIR) spectrum, $\lambda_1 = 3\mu m$ and $\lambda_2 = 5\mu m$ in our case.

Thermal data collection in our research has been accomplished by ATHEMOS (Automatic THERmal MOnitoring System), which we have developed in-house (see figure 2.4).

The 2D grid of the thermal camera captures the radiation energy of objects. Based on the excitation level of grid points, the processing unit converts the energy value at each grid point into a temperature value. Once the data have been converted to temperatures, they are transferred to a data storage unit, such as a computer hard disk. The raw image must be mapped to useful color values in order to visualize it easily.

We used two different thermal-imaging cameras in our experiments 2.6. In the deception detection experiment described in sections 3.1 respectively we used FLIR SC6000 thermal imaging camera from FLIR, and for the driving study described in section 3.3 we used the FLIR TAU 640 thermal imaging camera.



Figure 2.4: Custom thermal imaging system, ATHEMOS. The custom thermal imaging system developed by our group integrates a computer, a thermal camera, and several peripheral hardware components.

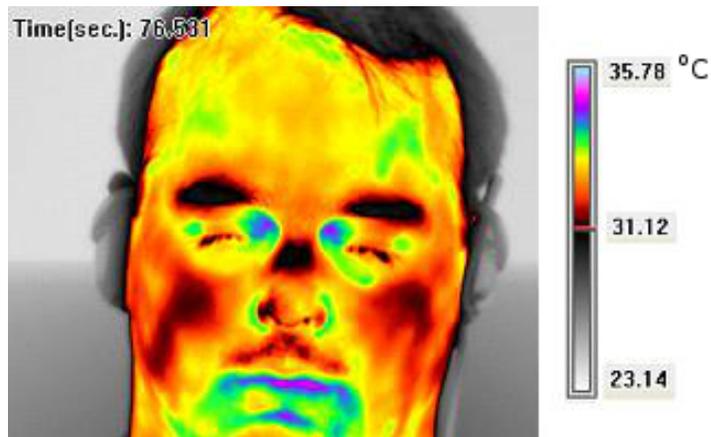


Figure 2.5: Thermal image. A sample thermal image of a subject. The bar on the right side shows the mapping between color and temperature in degrees centigrade.



(a) FLIR SC6000



(b) FLIR TAU 640

Figure 2.6: The thermal imaging cameras used in the experiments

2.3 Q Sensor

The Q Sensor is a wireless device that allows the user to conveniently record skin conductance as a function of activity of the sympathetic nervous system. It is designed for assessment of both cognitive as well as emotional stressors and can be worn on the palm or the distal forearm sites. Just like the traditional GSR instruments, the Q Sensor too has two Ag/AgCl electrode probes that come in contact with the region of measurement.



Figure 2.7: The Q Sensor worn on the palm, with the palm strap.

The Q Sensor is also embedded with a three-axis accelerometer. The data from these channels can be used for various forms of analysis such as posture analysis, and to measure the intensity and duration of fidgeting which is usually a sign of stress,

or disengagement, when looked at in proper context.

The Q Sensor can be setup with other recording devices and synchronized with a common clock to record data from different devices and compared on a synchronized time line. Its recording rate can be adjusted from 2 to 32 Hz. It is also capable of being used in longitudinal studies ranging from few hours to several days. There are a number of attachment straps which allow the Q Sensor to be worn on the palm, wrist and the ankle.

2.4 Zephyr BioHarness

The Zephyr BioHarness [1] is a compact physiological-monitoring module. It is attached to a lightweight Smart Fabric strap which incorporates ECG and breathing detection sensors. It is worn under the subjects clothing just above the sternum. The device can transmit physiological data wirelessly through Bluetooth or record it to internal memory and later on be downloaded to a computer for analysis.

The BioHarness device comes with an accelerometer, which can record the angle of the subjects' posture, and make determination about the subjects' activity. It can record heart rates from 0 to 240 BPM, and breathing rates from 1 to 120 BPM both with a sensitivity of ± 1 BPM. In addition to the calculated rates, the ECG and breathing waveforms are also stored along with the data which could be used for post-analysis to recompute rates based on needs. The Zephyr BioHarness data



Figure 2.8: The Zephyr BioHarness worn on the chest.

logger can also be synchronized with other devices like the thermal camera and Q Sensors being used simultaneously.

Chapter 3

Experiments

This chapter describes three experiments conducted to induce sympathetic loading and to cause stress. The main motivation for conducting these experiments is that sympathetic arousals affects human performance in critical tasks. We are able to quantify sympathetic arousals by measuring physiological changes in breathing rate, heart rate and skin conductance. Human performance also can be quantified depending on the domain. We choose three critical tasks of societal importance in our research, deception detection, educational testing, and driving. Since we can quantify both sympathetic arousals and human performance, we can study the interaction between them.

The motivation for the deception detection research is that there is a need for improved methods to detect criminal intent and detect lies, especially while interrogating suspect terrorists, detecting corporate and political lies and interrogating

suspect criminals to extract information from them.

The motivation for the exam study is that there is a need to reexamine exam assessment. Exams are an important part of a student's life, and typically students are judged by their grades only. Grades are singular data points and tell us just the end result. It would be interesting to measure a student's internal state. This can be done by monitoring physiology throughout the exam. Hence in this way, the student's progress could be measured more holistically, by considering not only the outward signs, which are their grades, but also looking at their inward state, which is how did they arrive at that outcome.

Driving is an important part of most adults' lives, and typically people spend about an hour driving themselves from home to work daily. It is usually during the rush hour commutes that there are accidents, sometimes even deadly. However, a traffic backup always incurs large economic losses (taking into account the loss of productivity of the large number of commuters sitting in traffic). A major known cause of accidents is distraction. Texting is a major distraction [31], for which there are now laws and rules to regulate such behavior while driving. This however is an apparent distraction. Distractions can be mental too. Our motivation in this study is to use sympathetic arousals, which is the most basic physiological response to any sort of stress, to study all sorts of potential stressors to a driver, which could be either cognitive, emotional or motoric.

Several factors were considered in designing each of the experiments. Some of these are 1) The experiments needed to be minimally obtrusive, 2) Realistic, 3)

Participants should be motivated to perform well, and 4) Subjects needed to be carefully and appropriately recruited.

The human body is highly sensitive and reactive. Every additional probe introduced makes the subject aware of the fact that they are being monitored and diverts their attention towards this fact. This introduces more stress on the subjects. As experimenters, we would like to have the subjects focused on the experimental task at hand and create an environment in which the only factors causing the subjects to experience stress are the experimental stressors. Hence, by using contact-free sensing and/or wearable devices, we achieve this requirement. Realism enables the subjects to engage completely into the experiment and effectively to treat it more seriously. In case of the deception-detection experiment, this is achieved through making them physically commit the simulated crime and having their rewards dependent on their performance at the tasks. In the case of the experiment involving the exam difficulty assessment, the experiment piggy-backed reality as students were monitored while they took real exams for a class they were enrolled in. While in the driving distraction study, we used a high-fidelity driving simulator, and subjects were told that their compensation depended on their driving performance and the extent to which they followed the road signs and speed-limits.

Though the experiments encompass few different areas of research, they are linked by several commonalities. The first aspect that unites these experiments is conceptual unity. Sympathetic arousals can affect human performance in critical tasks. We base our experiments on certain critical tasks of societal importance

and study the effects that these arousals have on these tasks. Our experiments are also united by their measurement approach. We measure sympathetic arousal by measuring either perinasal EDA or palmar EDA, done in minimally invasive ways. There is also unity in theme among these experiments. In each of the tasks that subjects are put through they are cognitively engaged and what we measure is in response to this engagement. In some cases, there is also emotional and motoric involvement, but cognitive challenge is common across the three testbeds. Also, each of these experimental protocols present subjects with some significant stakes to warrant their engagement. Finally, in each of these experiments are hypothesis-driven and we test our hypothesis using mostly statistical tests.

3.1 Deception Detection Study

3.1.1 Overview

The experimental setup for the experiment was designed and implemented by DACA (Defense Academy for Credibility Assessment) at the University of Arizona. DACA is an organization that trains in the methods of deception analysis such as the traditional polygraph approach in addition to more recent credibility assessment techniques and technologies. It was carried out in collaboration with the Computational Physiology Lab (CPL) from the University of Houston. This experiment was designed for the purpose of accumulating experimental data in various modalities to

develop new and effective techniques to detect deception. The methods discussed in this thesis use thermal imagery data and audio recording of the interviews captured using ATHEMOS, the thermal monitoring system of the CPL.

The experiment was to setup a mock crime scene in which some participants were randomly instructed to steal a ring from a designated office space in a building at the University of Arizona [8]. This experiment was conducted on a large scale over the period of a year and involving over 160 participants. When subjects arrived for the experiment, they were first asked to enter a room where they listened to a set of recorded instructions. These instructions briefed them about the experiment and about their role during the experiment. Following the briefing, they were instructed to leave that room and reach a room in another building nearby within 15 minutes. They were also advised against speaking to anybody while getting to the other room. They were asked to go up to the 4th floor of the building to room 429 and ask for a particular individual. As part of the experimental protocol, the attendant at room 429 would respond by saying that he did not know the individual in question and would go look for him, while he offered to have the subject wait within room 429. The ring that was to be taken by the selected subjects was placed in a cash box, within a desk drawer in this room. Subjects who were assigned a ‘Deceptive’ role in this experiment were asked to extract the ring from the cash box and place the ring on their person while they proceeded to the next room on the first floor of that building. Here, they were to be interrogated about their involvement in the crime. The role of all subjects whether they were asked to take the ring or not, was to try

and convince the interrogator that they were innocent and that they did not take any ring. All participants were awarded monetary compensation of \$15 for their time and involvement in the study. To serve as a motivation for participants to perform to the best of their ability at convincing the interrogator of their innocence, there was an additional bonus of \$50 for participants who were successfully able to convince the interrogator of their innocence.

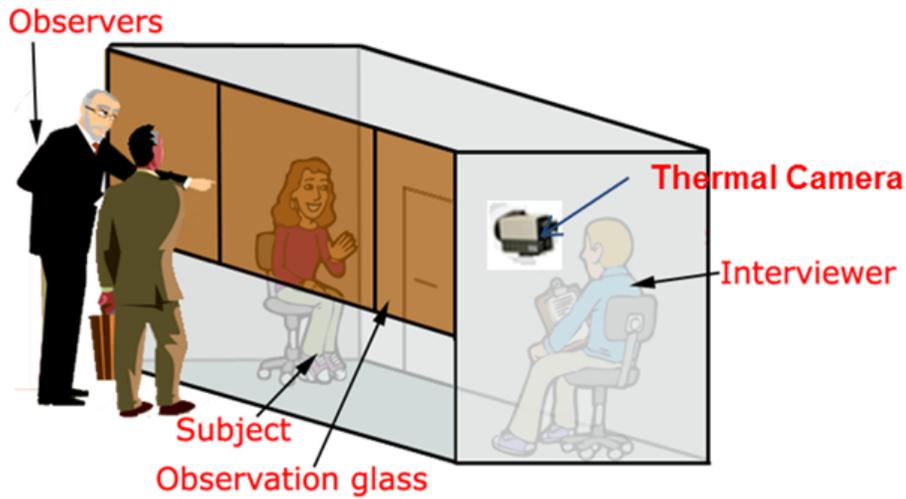


Figure 3.1: Diagram of the experimental setup.

Table 3.1: Reward Table

	Successful at convincing	Unsuccessful at convincing
Deceptive	\$15 + \$50	\$15
Truthful	\$15 + \$50	\$15

3.1.2 The Interrogation

The interview is designed to pose questions to the interviewees and to induce stress on them particularly while being strongly accused of being involved with the taking of the ring. The questions in the interview are structured purposefully to administer different levels of stress at each stage of the interview. The role of the initial questions is to extract the baseline response of the subjects. Some of the questions focus on the whereabouts of the participants since the time they report for the experiment while others directly accuses them of having some involvement in the crime which the interviewer could prove with the various surveillance equipment put in place.

During the interrogation, participants face three different types of interviews, namely the Behavioral Analysis Interview (BAI), Concealed Information Test (CIT) and the Stress Test (ST). The behavioral analysis interview poses questions which generally need to be answered in as much detail as possible. During this part of the interview, the interviewer gradually comes to a point where he tries to establish a link between the interviewee and the commission of the crime or action under discussion, therefore, causing the interviewee to admit his or her guilt. In the concealed information test, the interviewer says a word which is then repeated by the interviewee. This word may be a place or a description of an object or the crime scene, which when heard by a guilty interviewee may cause instantaneous physiological responses. Finally in the stress test interview, questions that can be answered in a word or two are asked, and the responses to which is expected quickly, allowing very little time to think about the response. In this thesis we only analyzed the first set of interview

questions, the BAI.

The questions asked as part of the behavioral analysis interview are as follows.

[Interviewer may follow up on any of the questions]

1. What color are the walls in this room?
2. Are there any lights on in this room?
3. Where were you born?
4. What is the name of the building we are in?
5. Did you ever take anything valuable from a place where you worked?
6. Have you ever lied to a person in position of authority?
7. How do you feel about taking the credibility assessment examination?
8. You know you are going to be tested about a crime committed in this building today. If you were involved in the crime in any way, you should tell me now.
9. Would you please describe everything you did for the first two hours after you awoke today? Think about what you were doing, how you felt, and what happened.
10. Now I'd like you to describe everything you did and saw from the moment you left the Esquire Building until you arrived here.
11. At any time were you on the fourth floor of this building?

12. We sometimes verify the information that people in this study give us. If I call the receptionist in Room 429, is there any reason that he will say you might have been near his desk? Im not saying that you are being dishonest, but we put you on the spot here and maybe you misremembered.
13. Is there any reason why your fingerprints should be on a desk on the 4th floor of this building? Maybe you just opened a receptionists desk to look for a pen?
14. Is there any reason why we may see you entering room 429 on any surveillance camera tapes?
15. Id like you now to describe in reverse order everything you did and saw from the moment you arrived here back to when you left the Esquire Building.
16. Is there anyone who can vouch for you coming directly to this room from the Esquire Building?
17. What is the item that was taken?
18. A ring was taken. Do you know where it is now?
19. What do you think should happen to the person who stole the ring?
20. Under any circumstances would you give that person a break?
21. What kind of person would steal jewelry from someones desk?
22. Is there anything that you said in this interview so far that you would like to change?

23. At any time during this study, were you instructed to lie?

24. How do you think this credibility assessment of you will turn out today?

3.2 Exam Study

3.2.1 Overview

This study has some exploratory goals for which we measured students physiological signals throughout their exams to try to uncover how sympathetic signals differ from good-to poor-performing students. Furthermore, we test to see if, through sympathetic signals, we can come up with some kind of assessment measure to assess their performance or to give them an alternate grade.

In this experiment, students enrolled in a Kinesiology course (Kin 3304) Anatomy and Physiology, at the University of Houston, volunteered to take part in an experiment aimed at studying if affective means could be used in determining a students performance in a test and the exam difficulty as experienced by the student. The affective sympathetic responses measured were provoked by the exam's perceived challenge and the level of student engagement.

Traditionally, students are evaluated by their grades in course exams. Provided that the course is well delivered and the exams are well designed, the weighted mean of exam scores is considered a reliable skill indicator. We challenge this conventional wisdom and we investigate if there is a hidden sympathetic cost that qualifies student

performance. In other words, we argue that in addition to apparent performance, the ‘internal’ cost at which this performance is attained matters. We are also interested to find if the mean sympathetic response of a class during an exam can track the difficulty of this exam. If this is true, then exam difficulty can be affectively sensed.

This study was conducted over the course of Fall of 2012, Spring of 2013 and Summer of 2013 and enrolled eight, eight and seven students respectively; a total of 23 subjects. Affective measures were acquired through an array of wearable sensors and self-report questionnaires. The course was designed by its instructor to have 5 exams in total of varying difficulty, and each contributing towards the students aggregate course performance. The most difficult of the exams being the final exam and the rest of the four exams in decreasing difficulty from exam 1 to exam 4. Another design factor which makes the selected course a desirable test bed for this study is the balance between free-response questions and multiple-choice questions and between critical-thinking questions and declarative-knowledge questions.

3.2.2 Measurements

We captured a number of physiological measurements from the students while they engaged in their exams. These measurements were taken through wearable sensors, so as to not constraint the students to any limitations. Students wore a Q Sensor on the palm of their non dominant hand, and one on the ankle. On their chest, we strapped a belt which recorded their heart rate and breathing rate throughout the exam. In addition both the Q Sensors and the Zephyr belt measured subject kinetics

from the hand, leg and torso. Self assessment forms about subjects trait anxiety levels was administered at the time subjects signed up for the study. State anxiety [2] and PANAS [3] questionnaires were also administered before and after each exam to evaluate subjects subjective stress levels. Further more, in addition to the previously described ‘inside’ measurement, we also took ‘outwards’ measurements in the form of observational recordings. Each students face and exam booklet was recorded by individual video cameras. These allow for contextual and causational links between their inwards stress levels, through physiological measurements, and outwards stress manifestation through facial observations.

3.2.3 Experiment

Subjects were recruited during the first week of class. In total students made 6 visits. During the first visit, subjects filled out the consent forms and took the baseline which involved filling out the Trait anxiety and the PANAS questionnaire which was relatively a non-stressful time as compared to during an exam. During each visit, four of the eight subjects took the exam starting an hour before the regular exam session, and the other four started their exam an hour after the regular exam start time. This was done because the experiment facility and resources allowed for monitoring of four students simultaneously. The students arrived 45 minutes prior to their exam start time and they filled the state anxiety and PANAS forms. These were to determine the pre-exam stress levels. Following these, the Q Sensors and Zephyr

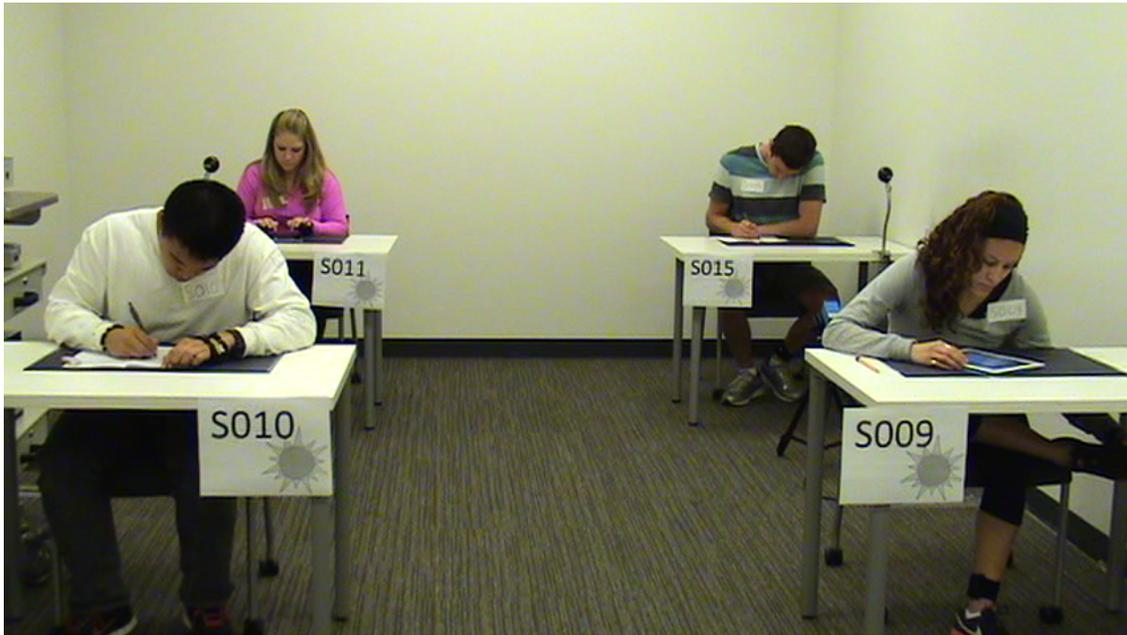


Figure 3.2: A view of the study in progress. 4 students take their examination while being monitored with wearable sensors, on the palm, ankle and a chest strap. Their facial expressions and exam booklets are recorded with video cameras.

belts were attached on each of the subjects, and they took their positions in the examination room. Once all subjects were fitted with their devices, the examination would start. When subjects finished their exams, they would leave the examination room and the devices would be unstrapped. They would also fill out the state anxiety form and PANAS after the exam to measure their post-exam stress levels.

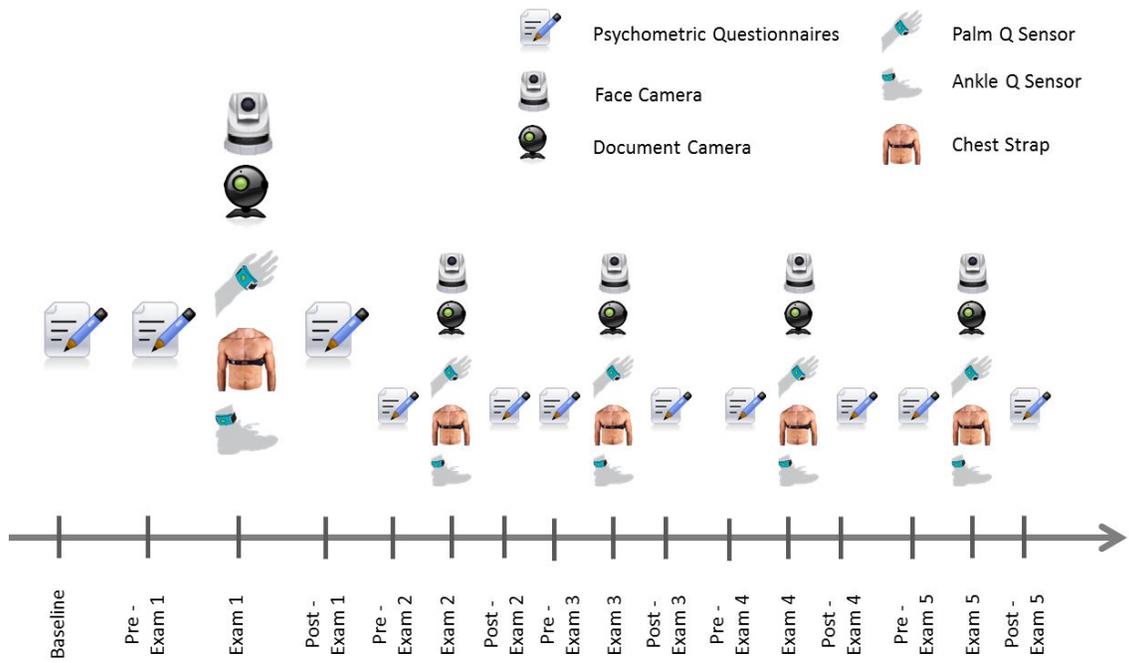


Figure 3.3: The timeline of administration of the various measurements throughout the course of the semester.

3.3 Driving Study

3.3.1 Overview

A central goal of this study is to understand better how driver-based and vehicle-based data might be used to detect higher-risk driving scenarios that might lead to motor-vehicle crashes. One of the focus is to test/validate driving conditions that induce high cognitive load, acute stress, and high emotional arousal and to determine the types and degree of driver performance decrements associated with these conditions. The knowledge and results derived from this study can be used for development of an online driver monitoring system and effective countermeasures for adverse cognitive and emotional states. More specifically we would like to find out if we can reliably detect various cognitive states of interest, and how these cognitive states affect drivers performance and safety.

3.3.2 Experiment

This experiment was a collaborative study between the Human Factors Group at the Texas A&M University and the Texas Transportation Institute (TTI). We recruited subjects from the Texas A&M campus (population about 60,000) through email solicitations and flyer postings. Subjects had to carry a valid driving license and have normal or corrected-to-normal vision. We restricted admission to individuals with at least one-and-a-half years of driving experience who were between 18 and 27 years of

age (Young cohort) or above 60 years of age (Old cohort). We excluded subjects on medications affecting their ability to drive safely. A total of 88 subjects participated in the experiment in a high fidelity driving simulator by Realtime Technologies, Inc (Fig. 3.4).



Figure 3.4: The driving simulator experimental setup. A thermal imaging camera mounted above the center TV screen records the subjects face. The virtual car is controlled by an accelerator pedal and a brake pedal, and steered with a steering wheel.

Upon signing the consent form, the subjects completed three questionnaires:

Biographic Questionnaire: It inquired key facts about the subject, such as gender, age, and driving record.

Trend Anxiety Inventory: Long-standing stress might have an effect on sympathetic responses and thus, scoring trend anxiety was of potential interest to this study.

Personality Type A/B: This was a modified version of the Jenkins Activity Survey. Some studies have shown association between type A personalities and specific driving behaviors; thus, scoring of type A/B personalities was also of potential interest to this study.

Next, the subjects went through $T_{session} = 7$ experimental sessions.

1: Baseline Session (BS) The subjects sat quietly in a dimly lit room, listening to spa music for 5 min. Following this baseline session, the subjects went through seven driving sessions on the simulator.

2: Practice Drive (PD) The subjects familiarized themselves with the simulator by driving on a 5 mi section of a four-lane highway at posted speeds; two lanes were dedicated to traffic in each direction, with the subject's car traveling in the right lane (R); the speed limits changed every couple of miles (50 mp \rightarrow 30 mph \rightarrow 60 mph). '3.5

Practice Drive

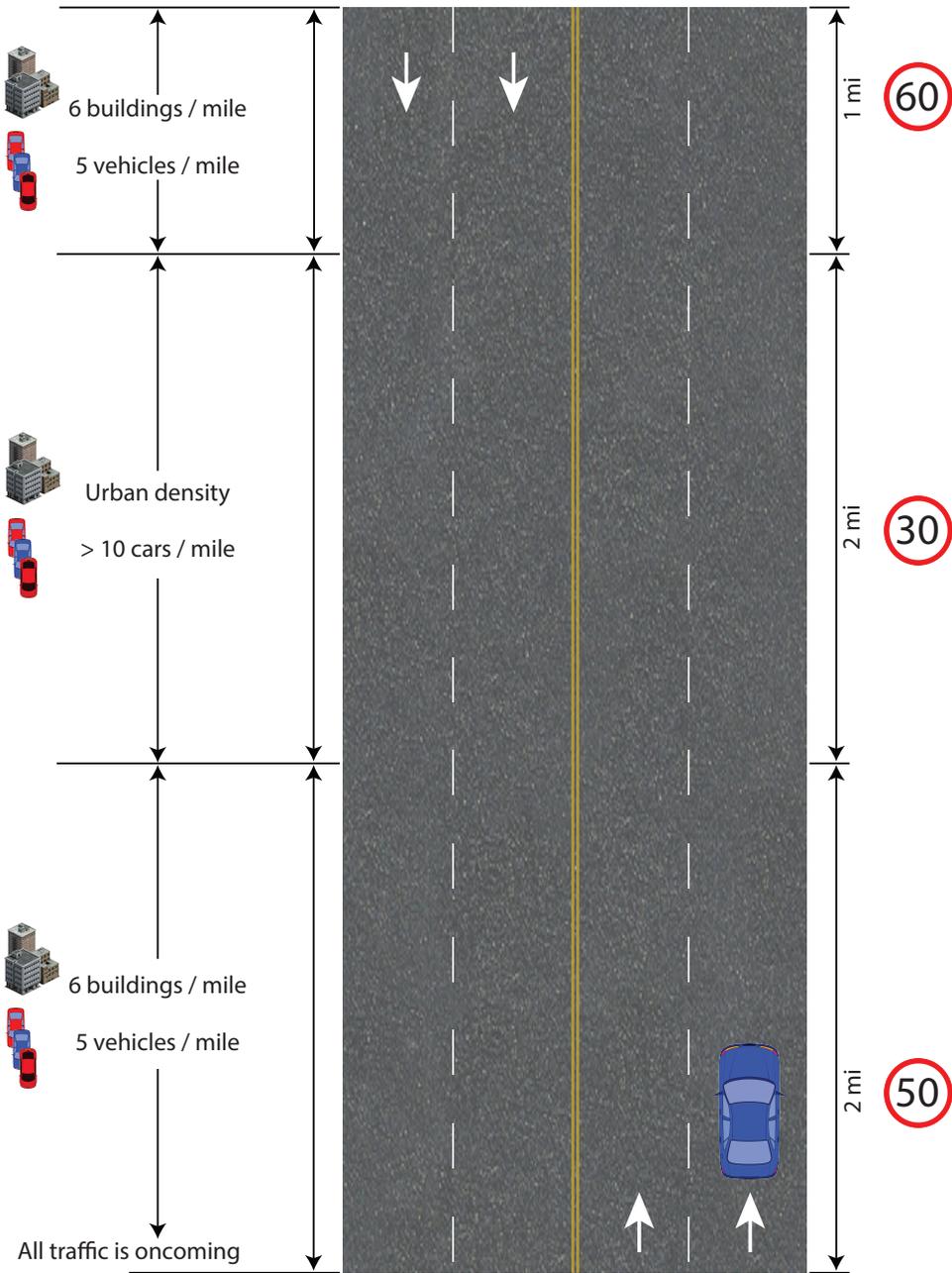


Figure 3.5: The practice drive scenario.

3: Relaxation Drive (RD) The subjects had to drive on a 6.75 mi section of a four-lane highway with posted speed limit of 45 mph; two lanes were dedicated to traffic in each direction, with the subject's car traveling in the right lane (R); there was light traffic on the oncoming lanes (5 vehicles per mile). The subjects were forced to change lane (R to L) after 3.25 mi into the drive. They stayed in the left lane (L) for 0.75 mi, before they were directed back to the right lane (R). The rationale for this lane change was to break the monotony of the drive.

Relaxing Drive

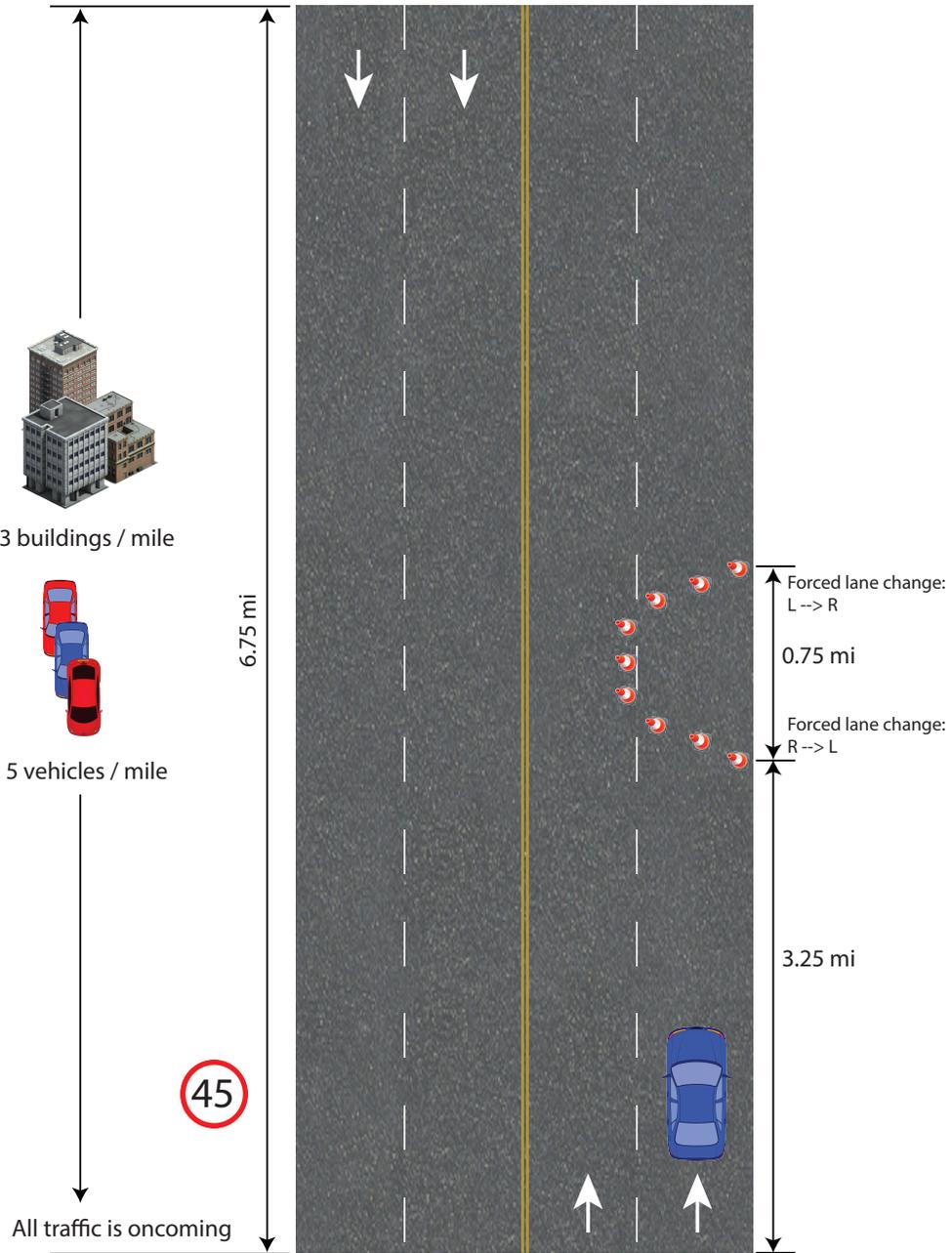


Figure 3.6: The relaxing drive scenarios.

4-7: Loaded Drives We randomized the order of four special driving sessions, called ‘loaded’ drives, featuring the same challenging driving conditions (construction zone). Each loaded drive was uniquely characterized by an additional stressor or absence thereof. This stressor assumed the form of a secondary activity that was forced in two phases during the course of the drive. All loaded drives were on the same 6.75 mi section of a two-lane highway with posted speed limit of 45 mph; one lane was dedicated to traffic in each direction, with the subject’s car traveling in the right lane (R). The drives featured heavy traffic on the oncoming lane (> 20 vehicles per mile) and construction pylons on both ends of the right lane (R). The subjects were forced to change lane (R to L) after 2.75 mi into the drive. They stayed in the left lane (L) for 0.75 mi, before they were directed back to the right lane (R). During the detour, construction barrels appeared on the right side of the lane. In more detail, the loaded drives were as follows:

Loaded Drive

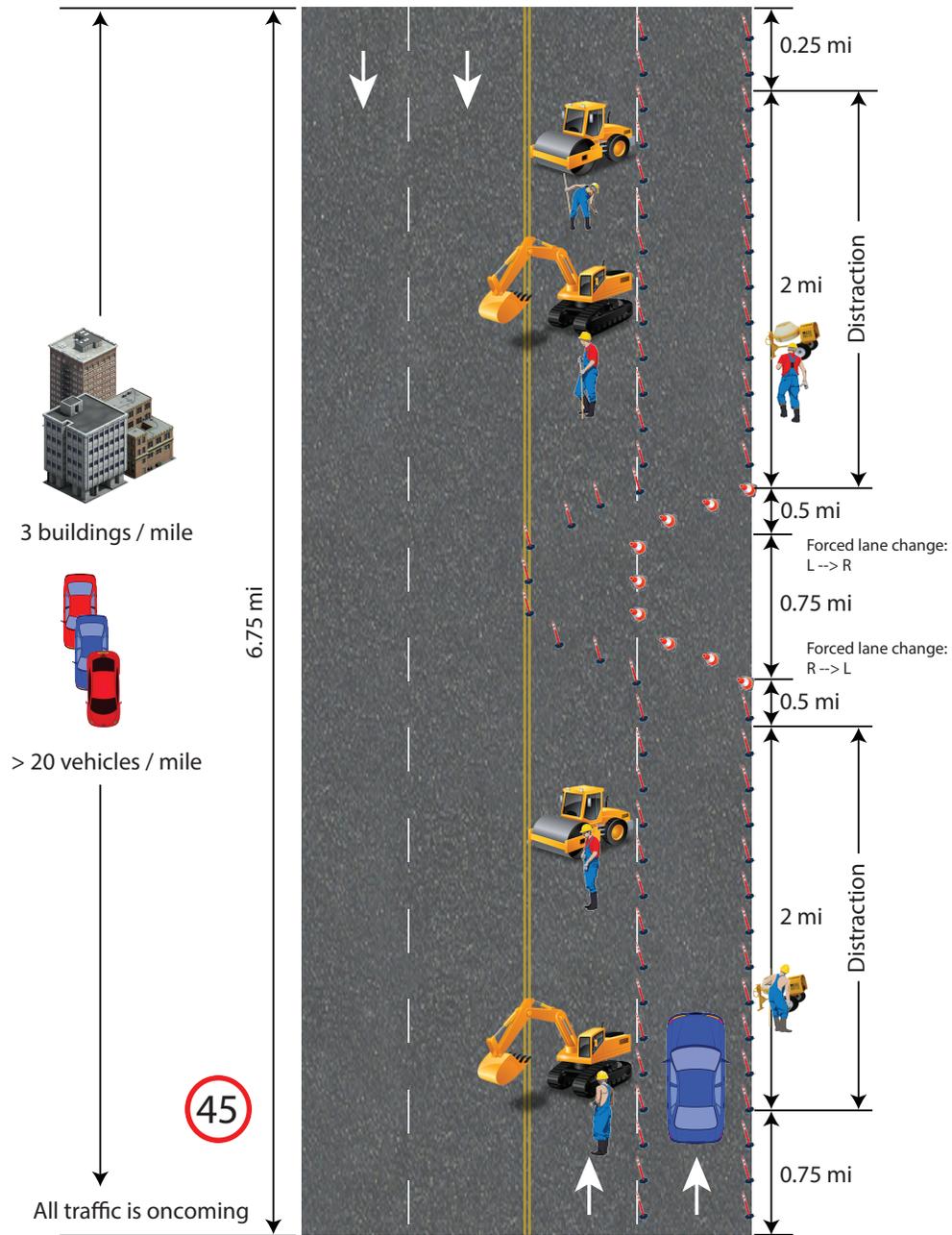


Figure 3.7: The four loaded drive scenarios.

- **Loaded Drive:** (LD) Driving with no secondary activity (no additional stressor).
- **Cognitive Drive:** (LD_C) Driving under a cognitive stressor. The cognitive stressor was mathematical questions in one phase of the drive and analytical questions in another phase of the drive, posed orally by the experimenter. The subjects had to answer these 5 questions to the best of their abilities. The mathematical vs. analytical phase order was randomized.
- **Emotional Drive:** (LD_E) Driving under an emotional stressor. The emotional stressor was emotionally stirring questions posed orally by the experimenter in two phases. The subjects had to answer these questions to the best of their abilities.
- **Motoric Drive:** (LD_M) Driving under a motoric stressor. The motoric stressor was texting back words, sent one by one to the subject's smartphone; this texting exchange took place in two phases.

The phase layout within each LD_X drive ($x \in [C, E, M]$) was as follows:

- **Phase $P1_{LD_X}$:** Driving without distractions for 80 s.
- **Phase $P2_{LD_X}$:** Driving while engaging in a secondary activity x for 160 s.
- **Phase $P3_{LD_X}$:** Driving without distractions for 240 s (coincided with the detour).

- **Phase P_{4LD_x}** : Driving while engaging in a secondary activity x for 160 s.
- **Phase P_{5LD_x}** : Driving without distractions for 120 s.

Chapter 4

Methods

4.1 Deception Detection Study

4.1.1 Interview Segmentation

To study the physiological effect caused in response to each question, we segmented each Question - Answer pair in the interview, including any follow-up questions asked by the interviewer. The start frame number and end frame number for the question and respective response was recorded, therefore we would be able to analyze question answer segments individually. However, because we apply a wavelet transform [30] on the physiological signals, the approach would not be suitable for performing analysis on each individual segment due to the short length of the question and its response.

The length of an individual segment would be too short to extract meaningful information as the physiological signals are weak having a very low signal-to-noise ratio and a very short length signal would only provide details about local or momentary physiological variations. Because we are more interested in analyzing the changes in physiology in succession to the increasingly incriminating questions, we grouped similar questions and their interlaced follow-up questions into segments.

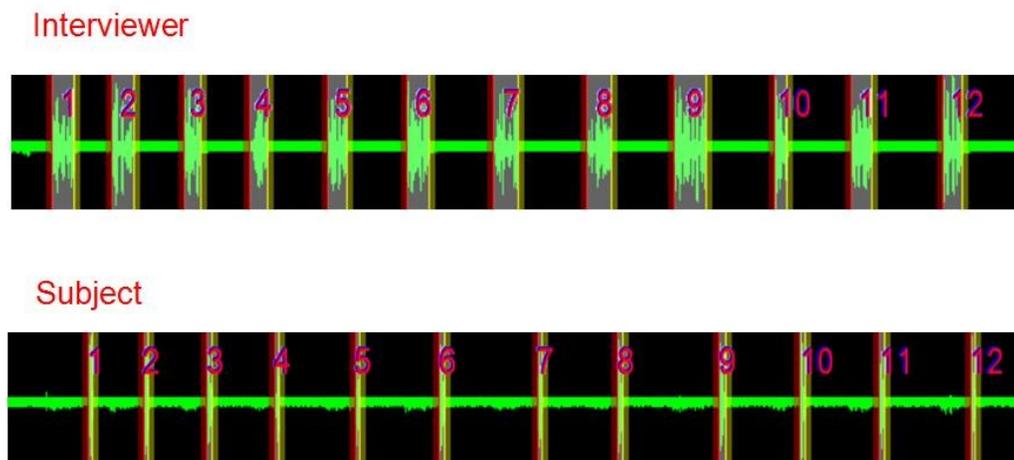


Figure 4.1: Audio segmentation of the interview questions.

For the Behavioral Analysis Interview (BAI), we grouped the 24 questions forming six segments. Each of these segments is of similar nature and is strategically included to serve a very specific purpose. The first group of questions is the baseline questions. The questions constituting this segment are answerable in one word or line. While these questions are being asked, a baseline response from the participant is extracted. It is against the baseline response that is collected, that we will compare the physiological response to a stimulus. The next group of questions inquires

about the subjects general deceptive behavior, e.g., if they ever lied to authority or stole from a work place in the past. The third segment is an irrelevant question, having no direct implication with the crime in question. It is a question requiring a descriptive response and its purpose is to bring the participants physiology at a state of rest so as to collect an additional baseline. The responses of the first and third segment are considered while determining the participants physiology at rest. The fourth segment contains questions which directly accuse the subject of having direct involvement in the crime, which can be proved by the interviewer with the help of the affixed surveillance and forensic equipment and through extraction of finger prints. The next groups of questions ask the participant about what they think would be the nature of the person who could have taken the ring and what would be an appropriate punishment for such an act. A deceptive participant may propose a more severe punishment for the person responsible in attempts to seem innocent; however the surprise of being faced by such a question causes physiological changes in the participants which they cannot control. The last group of questions concludes the behavioral analysis interview.

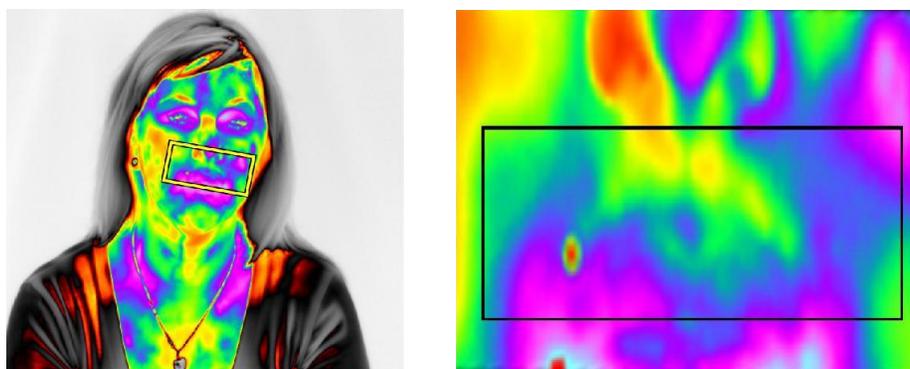
Table 4.1: Constitution of segments

Behavioral Analysis Interview Segments	
Questions	Segment Name
1 - 4	IR1
5 - 8	R1
9	IR2
10 - 16	R2
17 - 21	R3
22 - 24	R4

(IR - Irrelevant) (R - Relevant)

4.1.2 Tracking Region of Interest

The tracker used is a facial-tissue tracker based on the particle-filtering approach [33]. In contrast to a single particle filter-based tracker, the tracker used is a network-collaborative tracker, using multiple individual particle-filter trackers resulting in superior tracking abilities even in more challenging conditions. Since the resultant tracker is composed of multiple trackers, it is able to track dynamic scenes well, especially facial tissue in our case, hence compensating for deformation of the facial tissue. This kind of tracker is sufficiently robust to track abrupt changes in position and physiology. It handles head poses and partial occlusions well.



(a) Tracking region of interest

(b) Measurement region of interest

Figure 4.2: The tracking region of interest is depicted in 4.2(a). Within the tracking region of interest the measurement region is selected as shown in 4.2(b) from which the signal is extracted.

The method of tracking we used consists of the following steps:

1. First select the initial region for the tracker to track. This region is shown in

4.2(a), surrounded by the yellow box. This is the Tracking Region of Interest (TROI).

2. On the tracked region of interest, we outline the Measurement Region of Interest (MROI) 4.2(b). The MROI is a constant region from which physiological measurements will be extracted.
3. Once the two initial selections are made, the tracker will estimate the best matching TROI in the next thermal frame.

There are a few reasons for making two selections instead of one. Because the desired region of interest is very small, choosing a larger region would encompass lots more thermal features (steep gradients in temperature values) which makes tracking more robust. Also because physiology is constantly changing, a larger region is bound to have a more gradual change due to physiology than a smaller region. Finally, because we are extracting the perinasal signal (from the upper lip), this region is constantly being deformed due to the fact that the participants are speaking during the interview. Tracking only the perinasal region as the TROI would introduce huge amount of noise in the signal, mainly due to jittery tracking and excessive reselection of TROI.

4.1.3 Perspiration Signal Extraction

From the MROI in each thermal frame, we extract perspiration energy, which is obtained through a morphological based image processing method [26]. The rationale behind this method of extracting the perspiration signal is that the centers of the perspiration spots have low thermal emission due to emission of latent heat. Its signature is its physical characteristic of having ‘cold’ inner spots and ‘hot’ surrounding neighborhood. Also the boundaries of the perspiration spots are fuzzy due to thermal diffusion.

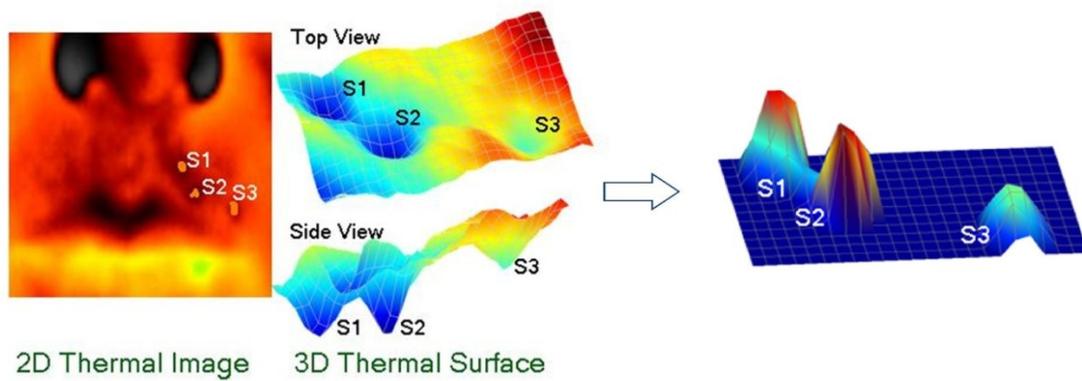


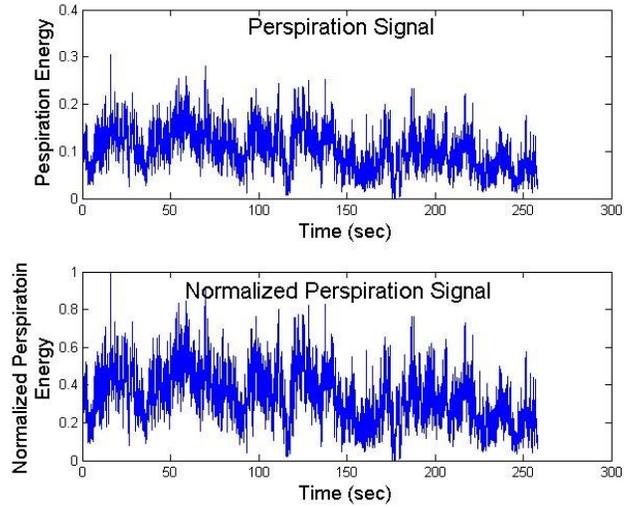
Figure 4.3: Extraction of the perspiration signal.

4.1.4 Signal Processing

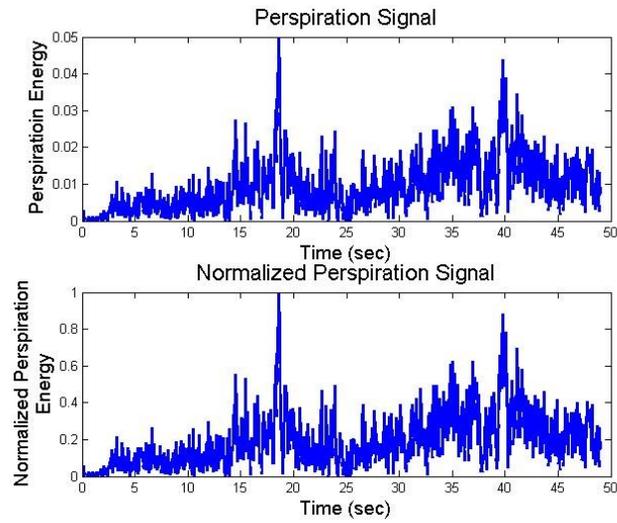
4.1.4.1 Normalization

Each signal segment which is created by segmenting the perspiration signal corresponding to the entire interview is considered individually by the wavelet transform. The main motivation behind applying the normalization step to these signals is that the perspiration signals extracted are very weak, and therefore have low signal-to-noise ratios. The physiological responses to stimuli vary from one segment to another, as well as physiological intensity varies from one subject to another. Because we are not interested in the magnitude of the energy but the rate at which it changes within the segments; we normalize each segment in the range of 0 to 1 before proceeding with applying the wavelet transform.

$$S_{norm} = \frac{S - Min(S)}{Max(S) - Min(S)} \quad (4.1)$$



(a)



(b)

Figure 4.4: Figure 4.4(a) (top) and 4.4(b) (top) represents two perspiration signals whose initial intensities are different by an order of magnitude. 4.4(a) (bottom) and 4.4(b) (bottom) show the normalized perspiration signals within the same intensity range.

4.1.4.2 Symmetric Extension

The wavelet transform is a convolution operation of the wavelet with the signal. Our signals are of finite length, while convolving at the ends of the signal, the operation would need additional data points beyond each end. This introduces an error called borderline-discontinuity error. The borderline-discontinuity error would introduce errors in wavelet analysis since it would lead incorrect local and global maximums in wavelet energy curves. There are a few ways to do signal extension, such as zero-padding, wraparound and symmetric extension. We used the symmetric approach as our signals are non-stationary in nature. Our choice of length for the extension was $N/4$ where N is the length of the signal.

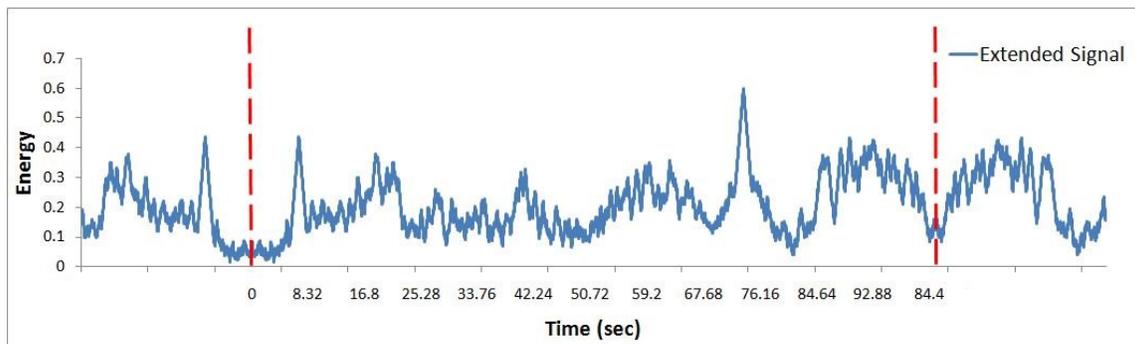


Figure 4.5: Symmetrically extended signal.

The extension provides signal points beyond the ends of the original signal, hence alleviating the result from the border line discontinuity problem. The wavelet energy curve no longer has wrong global maximums in its wavelet energy curves, therefore once again avoiding selection of improper scales for the frequency calculations.

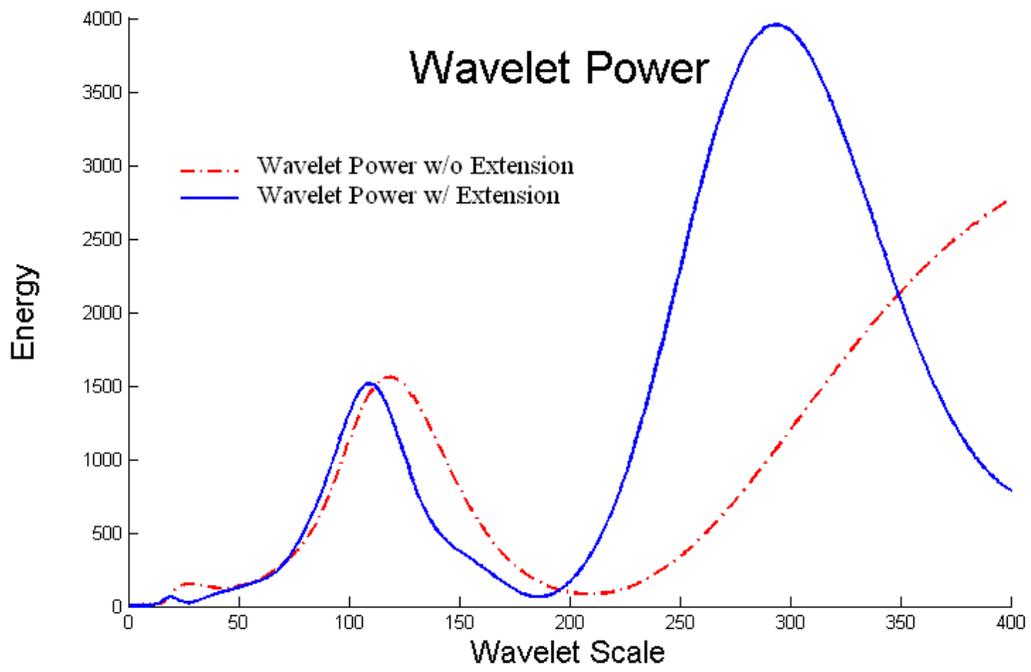


Figure 4.6: Effect of symmetric extension of signal on wavelet energy.

4.1.4.3 Impulse Cleaning

Due to imperfections in tracking, caused by the reasons mentioned previously, the perspiration energy signal is noisy. A large amount of this noise can be characterized as spiky or an impulse, caused due to the momentary displacement of the tracker for a frame or two. Due to the artificially introduced spikes of high intensity in the signals, it affects the normalization of signals and does not allow the signal to be amplified to entire potential. The presence of impulse noise is a category of high frequency noise, which is characterized by stochastic impulses (i.e., random in nature), occurring very frequently. The occurrence of such noise in the signals lead to the domination of high frequency peaks in the wavelet energy curve, which exists at the lower wavelet scales. Therefore it is necessary to perform impulse cleaning.

$$E_{diff} = \left[\sum_{f=1}^{n-1} abs(E_{(f)} - E_{(f+1)}) \right] / n \quad (4.2)$$

If the difference in energy between two consecutive frames differs by an amount greater than the mean difference (E_{diff}) between all consecutive frames in the signal, we reduce the difference in energy by an amount equal to the mean difference (E_{diff}).

$$\begin{aligned} & if(abs(E_f - E_{f-1}) > E_{diff}) \\ & E_f = E_{f-1} - E_{diff} \end{aligned}$$

After applying the impulse-cleaning step, a wavelet transform applied on the

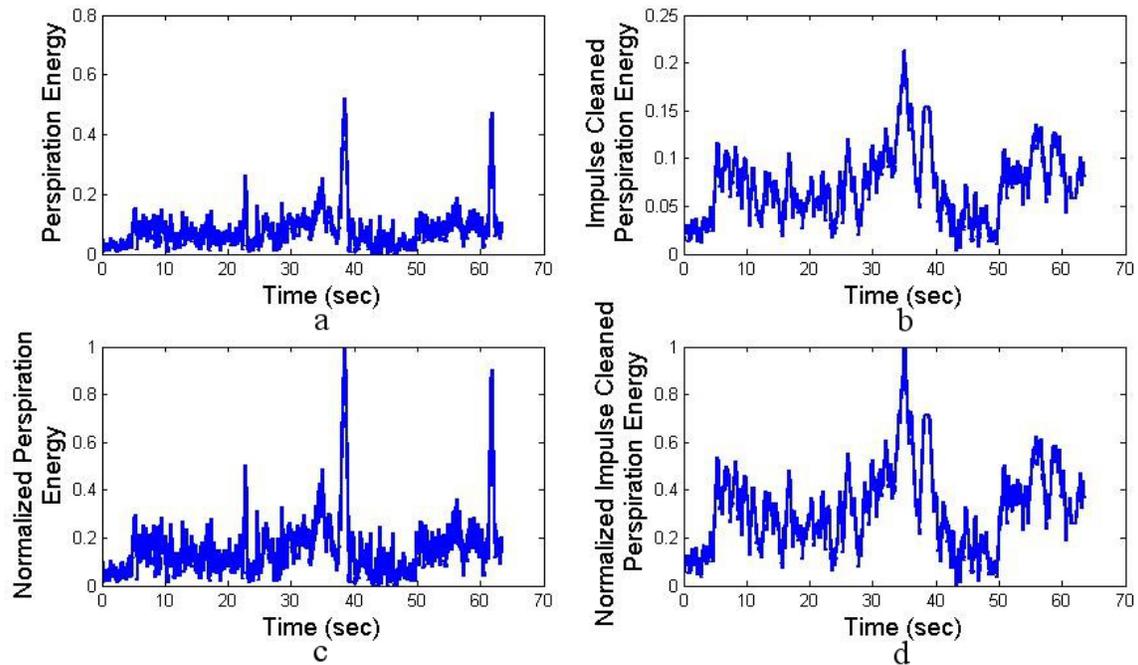


Figure 4.7: Figure (a) represents the original perspiration signal, figure (b) represents impulse cleaning applied on (a). Figure (c) and (d) represent normalization applied on (a) and (b) respectively.

signal would now identify appropriately the mid frequency components as the dominating harmonics in the signal. The same is observed in the wavelet energy graphs of the wavelet transforms applied to the same signal, with and without impulse cleaning.

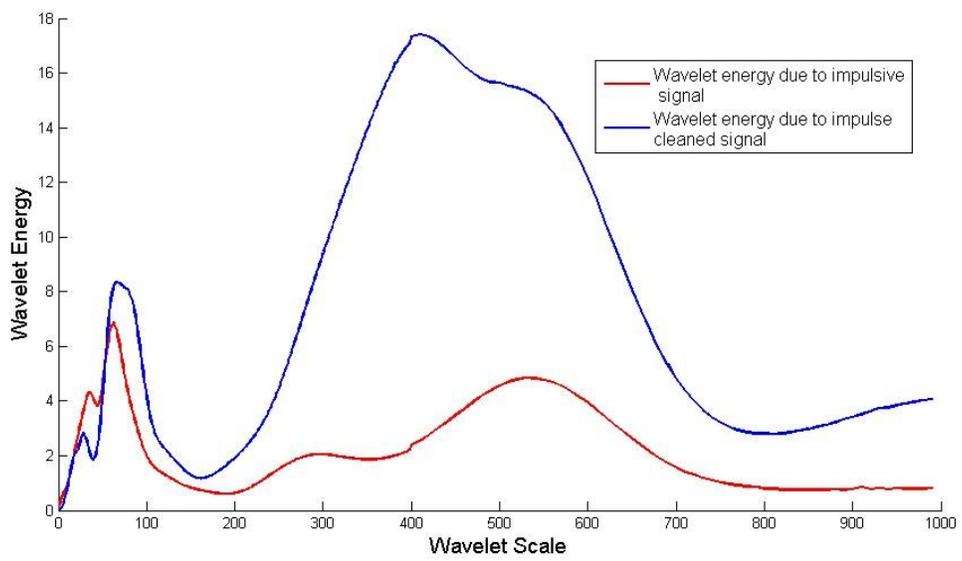


Figure 4.8: Effect of impulse on computed wavelet energy.

4.1.4.4 Noise Cleaning

Even though impulse noise is cleaned, the perspiration signals are still embedded with high-frequency noise. This is mainly due to systemic noise and tracker oscillations. Our intension is to analyze the lower-frequency variations in the perspiration signals which are indicative of the changing rate of perspiration. We cancel out the effect of the high-frequency fluctuations by passing the signal through a Fast-Fourier-Transform-based low-pass filter setting with decay (σ) of 0.005. The decay (σ) is experimentally chosen to be small so the influence of the signal points on the output noise cleaned signal will not be confined just to its near vicinity in the input signal, but it will be strongly influenced by signal points at a larger distance in both directions. Finally, in figure 4.9, the noise reduced signal is free from sharp changes in amplitude; however it still preserves the overall shape of the signal indicating true physiological phenomenon.

Noise cleaning is the final step required for preprocessing the signal before the wavelet transform can be applied on it. The wavelet transform applied on the normalized, impulse cleaned and symmetrically extended signal contains only one globally prevalent peak among the mid to low frequency scales, therefore, avoiding the dilemma of selection of improper peaks and can also be used to automatically calculate frequency.

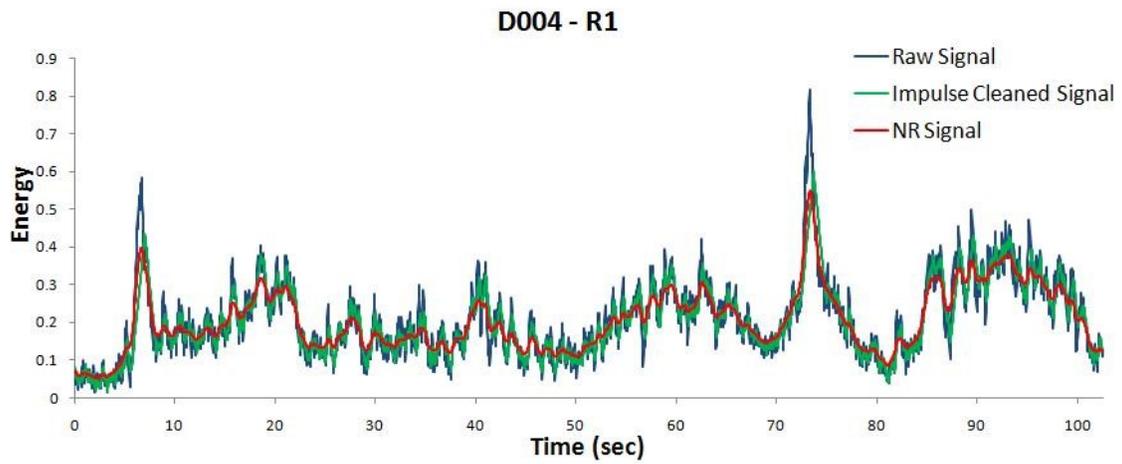


Figure 4.9: Raw, Impulse cleaned and Noise reduced signal.

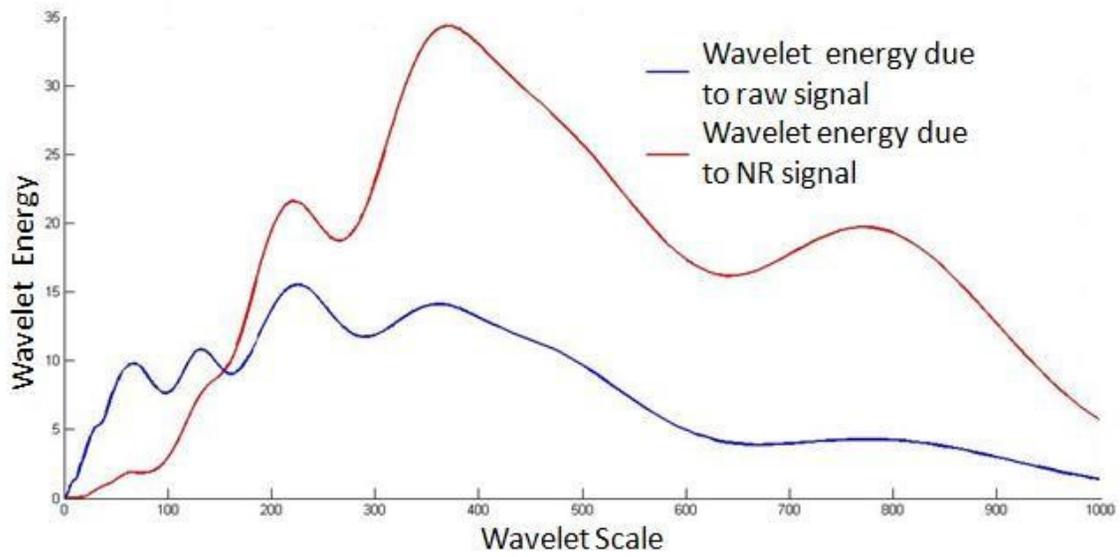


Figure 4.10: Effect of noise cleaning on wavelet energy.

4.1.4.5 Wavelet Transformation

The perinasal-perspiration signal, as its name suggests, measures the amount of perspiration at any given point in time. This region of interest is constantly being modulated by two factors. Firstly, by the large concentration on sweat glands in this region and secondly, from the thermal effect of breathing, because it is situated just under the nostrils [25]. Because we are interested in measuring the rate of modulation of perspiration or the rate of activation of the sweat glands, we apply a wavelet transformation to the perspiration signal. The Wavelet Transform with its time-frequency analysis is a very useful tool in performing harmonic analysis. Thus we can determine the strongest harmonic, which could be the representative the phasic response of an entire interview segment.

Because our signal is extracted from a region on the face that is being affected by a number of factors, we use multi-resolution wavelet in order to separate the effects due to these factors. For example, the breathing is clearly evident at lower scales while the perspiration is more pronounced at higher wavelet scales.

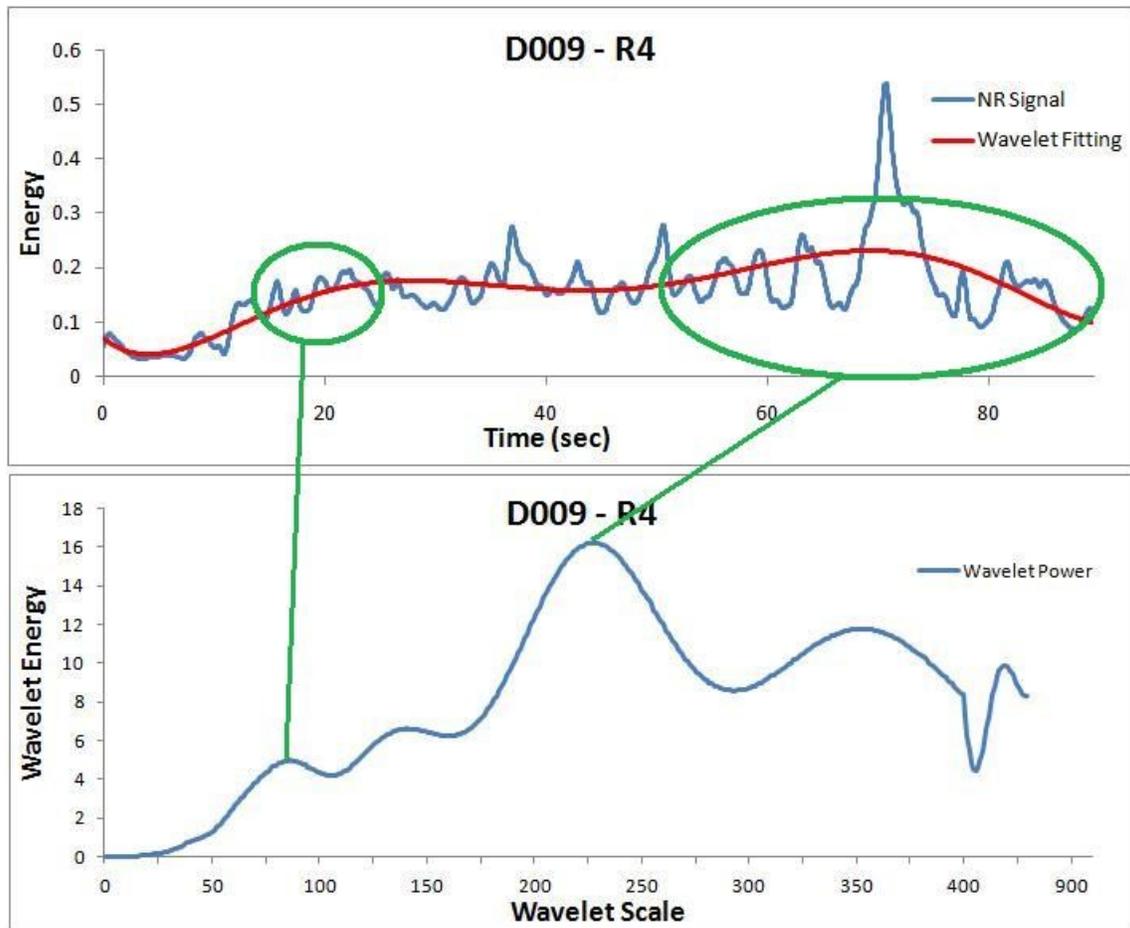
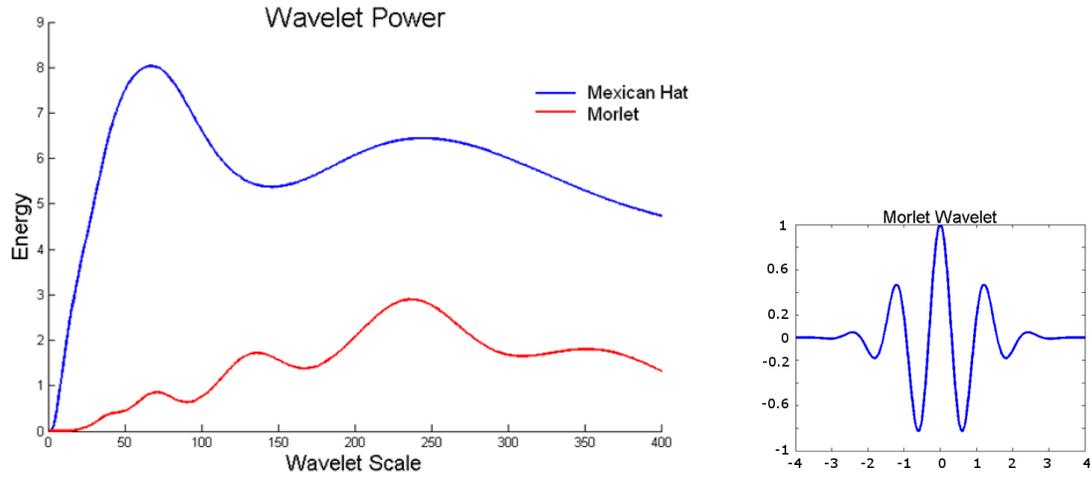


Figure 4.11: Noise-reduced signal superimposed with the wavelet fitting (top). Wavelet energy graph (bottom). The peak at the lower wavelet scale corresponds to high-frequency noise while the peak at the higher wavelet scale corresponds to true perspiration phenomenon.

4.1.4.6 Continuous Wavelet Transformation

We applied Continuous Wavelet Transform (CWT) on the signals with the “Morlet” wavelet as the choice of Mother Wavelet. We tried other mother wavelets like Daubechies-2 and Mexican Hat wavelet. Among the different mother wavelets we used, the power diagrams generated using the “Morlet” wavelet was able to produce clearly distinguishable peaks. The choice of the wavelet was also fortified due to the close resemblance between the perspiration signal and the “Morlet” wavelet.



(a) Wavelet energies using Mexican Hat and Morlet mother wavelet

(b) Morlet mother wavelet

Figure 4.12: The figure compares the wavelet energies of the two mother wavelets and it is observable that while using Morlet mother wavelet, the global maxima is clearly distinguishable.

We apply the continuous-wavelet transformation on the noise-reduced perspiration signal at all scales from 1 to 1000. From the wavelet power curve we choose

the wavelet scale corresponding to the maximum energy in the range of 100 to 650. We choose this range as such because in case of the Morlet wavelet, scales below 100 correspond to very high frequency. The presence of high energy at the lower scales is due to the modulation effect of breathing on the perinasal region. In the figure a, the high energy at scale 60 is caused to the strong cyclic pattern of period 4-6 seconds. In figure b, we see that there is a region of high energy around scale 650. The reconstructed signal at that scale fits a global trend in the signal spanning well over a 100 seconds.

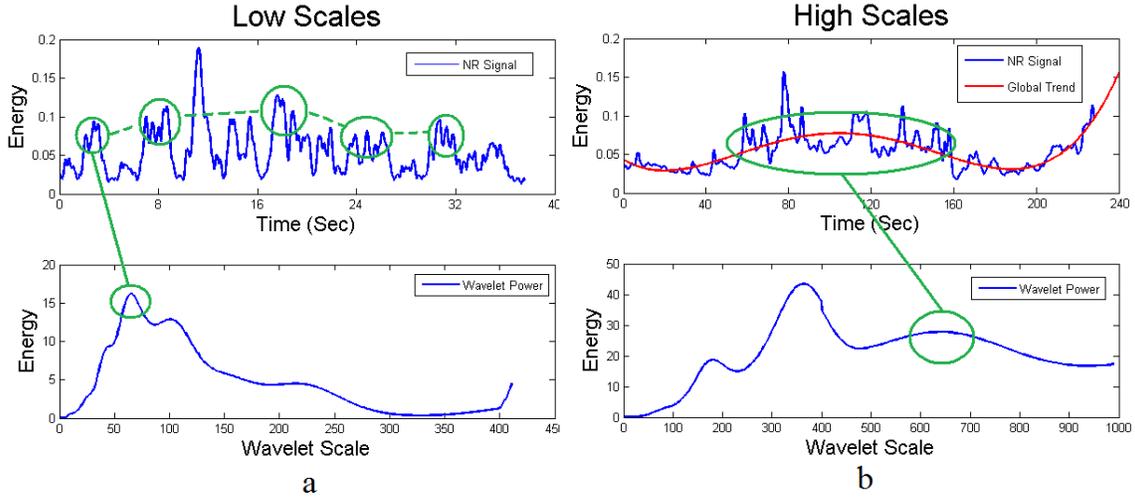
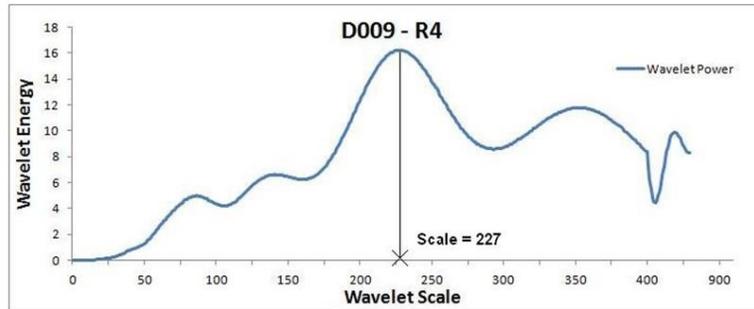


Figure 4.13: The figure points out correspondence of low and high wavelet scales, with high-frequency noise and global trends in the perspiration signal. (a) top: Perspiration signal with dominant high-frequency noise. (a) bottom: Wavelet energy corresponding to the (a) top. (b) top: Perspiration signal containing a very long cycle. (b) bottom: Wavelet energy corresponding to the (a) bottom.

4.1.4.7 Feature Extraction

From each segment of the perspiration signal, we wish to find the most prevalent harmonic in the signal using which we can compute the mean rate of perspiration during that segment of the interview. To do so we perform a scale to frequency computation considering the scale with highest energy. This way we obtain a single feature value for each segment of the interview. These values are frequency of perspiration. From these features we further create other features Relevant, Irrelevant and Difference.

Irrelevant is the average of the two irrelevant segments IR1 and IR2, while Relevant is the average of the relevant segments R1, R2 and R3. The final feature Difference is the difference of the Relevant and Irrelevant.



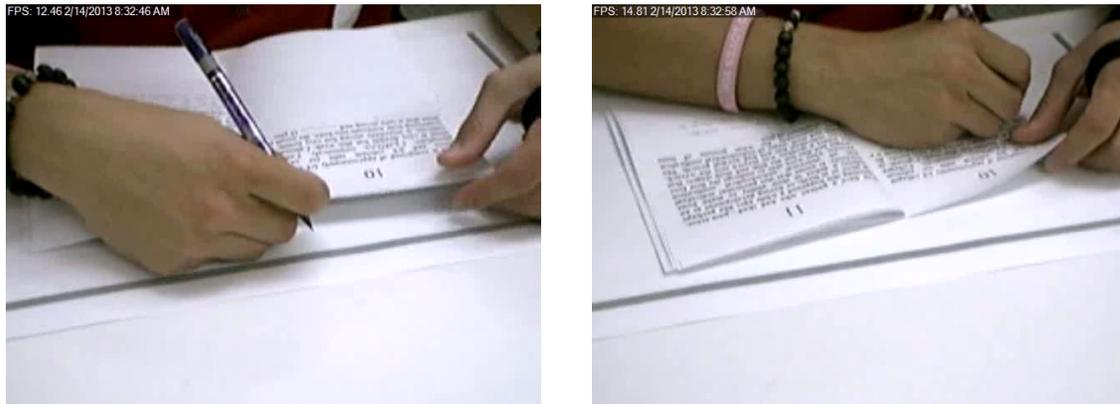
scale2freq(227,"Morlet",1/25)
0.089089

Subject	IR1	R1	IR2	R2	R3	R4	Relevant	Irrelevant	Difference
D001	0.033854167	0.091497748	0.058369253	0.048363095	0.036272321	0.072286477	0.058711055	0.04611171	0.012599345
D004	0.032242063	0.053174084	0.189836449	0.03329918	0.056111878	0.049542683	0.047528381	0.111039256	-0.063510875
D009	0.199142157	0.037615741	0.070529514	0.0559573	0.0390625	0.089089912	0.044211847	0.134835835	-0.090623988
D012	0.150462963	0.123856707	0.083590535	0.056738827	0.042317708	0.05387931	0.074304414	0.117026749	-0.042722335
D014	0.05721831	0.056897759	0.053174084	0.058369253	0.035635965	0.152725564	0.050300992	0.055196197	-0.004895205
.
.

Figure 4.14: Translation of wavelet scale at maximum wavelet energy to perspiration frequency values. (top) Wavelet energy graph with maximum energy at scale 227. (bottom) Table containing perspiration frequency values for each segment for all subjects.

4.2 Exam Study

In this experiment, we do not have a definite number of stimuli or events that we analyze. These events are each attempt of each question during an exam session which are different for each subject. Hence we segment each attempt sequentially using the feeds from the document cameras (see figure 4.15).

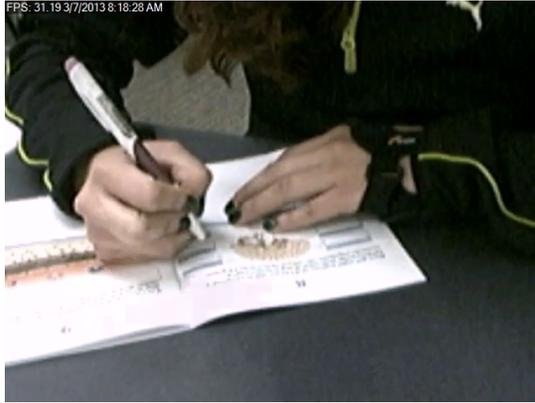


(a) Video annotation of subject reading a question.

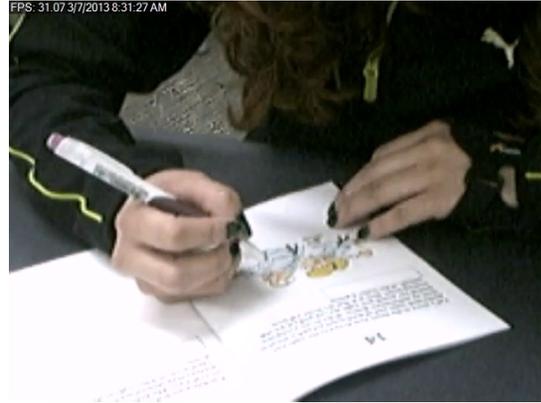
(b) Video annotation of subject writing an answer.

Figure 4.15: Video annotation of subjects' actions at every instance when they start and end an action.

Each of the questions belongs to one of two categories. 1) Declarative Knowledge Questions and 2) Memory Response Questions. Declarative knowledge questions require the students to recollect answers to questions which she/he has previously learned from the study material, while critical thinking questions test the extent to which the student has internalized the subject matter, and present an analytic answer to the question.



(a) Labeling parts of the brain.



(b) Analytical thinking.

Figure 4.16: (a) - Declarative Thinking Question. (b) - Critical Thinking Question

Based on the video annotation, we split the physiological signals from the Q Sensors and Zephyr chest straps. Each of the measurement devices and video feeds were synchronized with a common time server. Figure 4.17 illustrates the segmentation of the palm EDA signal, palm and ankle motion, heart rate and breathing rates signals with indexing from the video annotation. The green-highlighted portion indicates that the subject is reading or thinking during those durations, while the pink-highlighted portion indicates that the subjects are writing down the answers during those durations. Each segmented question-answer segment is also labeled indicating whether it is declarative-knowledge or critical-thinking type. We also obtained a copy of the students graded exams and recorded the scores for each question, allowing us to analyze their performance on the different types of questions.

We condense the physiological data per question and answer segments to a single value by representing it by its mean intensity.

K – Declarative Knowledge
 T – Critical Thinking



Figure 4.17: Physiological signal segmentation and annotation.

4.3 Driving Study

During the baseline session and all the subsequent drives, we continuously imaged the subject's face with a thermal camera 2.6(b). At the same time, the simulator was programmed to save a record of the evolving driving parameters. These parameters included speed, acceleration, braking, steering angle, and lane position. We use the perinasal perspiration as the indicator of sympathetic arousal. Out of the 88 subjects that participated, we performed analysis on 59 (26 male / 33 female) subjects. We could not perform analysis on nine male subjects because they had facial hair, rendering extraction of perinasal perspiration signal problematic. For 20 other subjects, either the thermal imaging was out of focus, or certain session files were missing. The ratio of males to females was 12/18 and 14/15 in the young and old cohorts, respectively. Similar methods for tracking the region of interest and extraction of the perinasal perspiration signal were used as described in 4.1.2 and 4.1.3 respectively.

The experiment had three parts:

Introductory Sessions One sitting session with soothing music to establish the subjects' resting baselines and two drives. These two initial drives were meant to familiarize subjects with the simulator (PD drive) and relax them (RD drive), respectively, in preparation for the main experimental drives.

Loaded Drives Four drives (order randomized) repeated on the same highway

segment under similar ambient conditions; they featured a modicum of driving difficulty (LD_j drives). One of these ‘loaded’ drives had no additional stressor ($LD, j = \text{NULL}$). The remaining three loaded drives were characterized by forced distraction through a j stressor - cognitive ($LD_C, j = C$), emotional ($LD_E, j = E$), or motoric ($LD_M, j = M$), respectively. The distraction applied twice during the corresponding drive. The presence or absence of distraction divided the LD_j drives into five phases: $P1_{LD_j}$ (no distraction for all); $P2_{LD_j}$ (distraction in LD_C, LD_E, LD_M vs. no distraction in LD); $P3_{LD_j}$ (no distraction for all); $P4_{LD_j}$ (distraction in LD_C, LD_E, LD_M vs. no distraction in LD); $P5_{LD_j}$ (no distraction for all).

We segmented and annotated these various phases of the drives using Subject-Book(Fig. 4.18) [4], and represented each of these phases’ physiological and simulator output signals by its mean value.

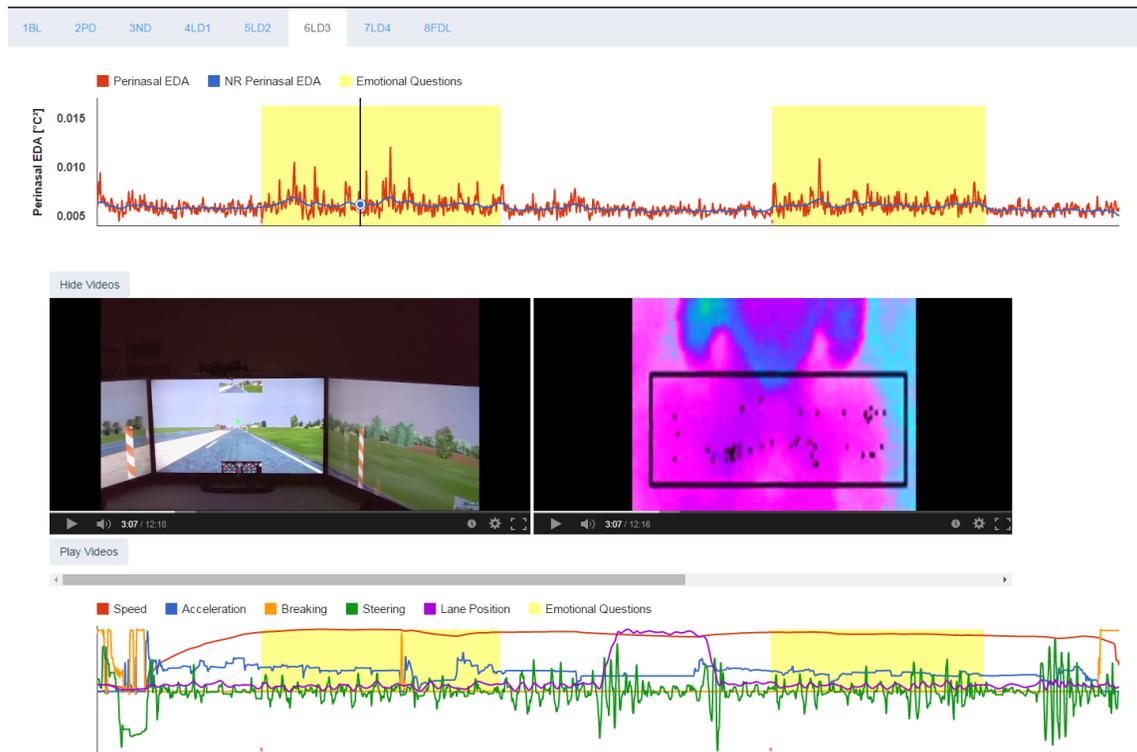


Figure 4.18: A screenshot of SubjectBook showing Subject 29, LD_E . The signal on the top is the perinasal perspiration and the signals on the bottom are the output of the various simulator parameters. Between the signals, we see a synchronized playback of videos corresponding to the black vertical marker. The highlighted section annotates the loaded part of the drive.

Chapter 5

Results

5.1 Deception Detection Study

A total of 67 participants were considered as part of the analysis, of which ground truth was available for only 25 subjects. Using the ground truth data set, we built models of ‘Truthful’ and ‘Deceptive’ subjects. Using the model generated by these 25 subjects, we tested the classification on the remaining 42 subjects as ‘Truthful’ or ‘Deceptive’. We performed two independent tests of batch sizes 27 and 15 subjects respectively. On the first test batch of 27 subjects, our model was able to classify 78 % accurately, and on our second attempt at testing with 15 subjects, the model classified 80% correctly. For the total of 42 subjects tested, the overall classification success rate is 78.5%. Table 5.1 and 5.3 contain the results of classification for the validation set and test set 2 respectively.

Table 5.1: Prediction on the validation set

Subject	Ground Truth	TH	MLP	AB(DS)	DT	AB(NB)	Probability
D001	Deceptive	1	1	1	1	1	0.999953
D004	Truthful	1	1	1	1	1	1
D009	Truthful	1	1	1	1	1	1
D012	Truthful	1	1	1	1	1	1
D014	Truthful	1	0	1	0	1	0.80239
D016	Truthful	1	0	1	1	1	0.946574
D017	Truthful	1	1	0	1	1	0.794048
D018	Truthful	1	1	1	1	1	1
D020	Deceptive	1	1	1	1	1	0.998949
D024	Truthful	1	1	1	1	1	1
D025	Deceptive	1	0	1	1	0	0.999992
D026	Truthful	1	1	1	1	1	0.996255
D027	Truthful	1	1	0	1	1	0.69268
D029	Deceptive	1	1	1	1	1	1
D033	Truthful	1	1	1	1	1	0.999969
D036	Truthful	1	1	1	1	1	0.941929
D038	Deceptive	1	1	1	1	1	1
D039	Truthful	1	1	1	1	1	0.998106

Continued on next page

Table 5.1 – continued from previous page

Subject	Ground Truth	TH	MLP	AB(DS)	DT	AB(NB)	Probability
D043	Truthful	1	1	1	1	1	1
D044	Truthful	1	1	1	1	1	0.999953
D046	Truthful	1	1	1	1	1	1
D052	Deceptive	1	1	1	1	1	0.998949
D053	Deceptive	0	0	1	1	1	0.757142
D055	Truthful	1	1	1	1	1	1
D057	Deceptive	0	0	0	0	1	0.871645
Success Rate		92%	80%	88%	92%	96%	

1 Correct Classification

0 Incorrect Classification

TH Thresholding

MLP Multilayered Perceptron

AB(DS) Boosting (Decision Stump)

DT Decision Tree

AB(NB) Boosting(Naïve Bayes)

The threshold classification approach classified 100% of the truthful subjects correctly and 6 out of 8 (75%) deceptive subjects correctly. The other machine-learning

algorithms misclassify different subjects however other than the Multilayered Perceptron, none of them misclassify more than 2 deceptive subjects. Also if we consider a majority voting to decide classification, doing so will result in a classification of 96%. We also see this trend in the prediction results for test set 2. There seems to be a high degree of agreement among the classifiers. The only information about the classification results pertaining to the undisclosed test set is that it achieved an overall success rate of 78% in blind predictions. The ground truth for these subjects has not been revealed, probably with the intention to use the data to promote further research in this field.

Table 5.2: Prediction of Test Set 1

Subject	Ground Truth	TH	MLP	AB(DS)	DT	AB(NB)	Probability
D064	?	T	T	T	T	T	0.957
D065	?	D	D	D	D	D	1
D069	?	D	D	D	D	D	0.988
D072	?	D	D	D	D	D	0.738
D076	?	D	D	D	D	D	1
D077	?	T	T	T	T	T	1
D078	?	T	T	T	T	T	1
D080	?	T	T	T	T	T	1
D082	?	T	T	T	T	T	1
D083	?	T	T	T	T	D	0.519
D084	?	D	T	D	D	D	1
D089	?	T	T	T	T	T	1
D091	?	T	T	T	T	T	1
D094	?	D	D	D	D	D	0.98
D096	?	T	D	D	D	T	0.765
D107	?	D	D	D	D	D	1
D110	?	T	T	T	T	T	1
D113	?	T	T	T	T	T	0.999

Continued on next page

Table 5.2 – continued from previous page

Subject	Ground Truth	TH	MLP	AB(DS)	DT	AB(NB)	Probability
D114	?	T	T	T	T	T	1
D115	?	T	T	T	T	T	1
D118	?	T	T	T	T	T	1
D121	?	T	T	T	T	T	0.999
D122	?	D	D	D	D	D	0.978
D123	?	T	T	T	T	T	1
D124	?	D	D	D	D	D	0.613
D127	?	T	T	T	T	T	1
Success Rate		?%	?%	?%	?%	?%	

T Predicted Truthful
D Predicted Deceptive
TH Thresholding
MLP Multilayered Perceptron
AB(DS) Boosting (Decision Stump)
DT Decision Tree
AB(NB) Boosting(Naïve Bayes)

Table 5.3: Prediction of Test Set 2

Subject	Ground Truth	TH	MLP	AB(DS)	DT	AB(NB)	Probability
D130	Deceptive	1	1	1	1	1	0.984
D131	Deceptive	0	0	0	0	0	1
D132	Truthful	1	1	1	1	1	0.913
D136	Deceptive	0	0	0	0	0	1
D140	Deceptive	1	1	1	1	1	0.964
D144	Truthful	1	1	1	1	1	1
D145	Deceptive	1	1	1	1	0*	0.501
D147	Truthful	1	1	1	1	1	1
D149	Truthful	1	1	1	1	1	1
D151	Truthful	1	1	1	1	1	0.95
D153	Truthful	1	1	1	1	1	1
D154	Truthful	1	1	1	1	1	1
D161	Truthful	1	1	1	1	1	1
D162	Truthful	0	0	0	0	0	0.976
D163	Deceptive	1	1	1	1	1	0.894
Success Rate		92%	80%	88%	92%	96%	

*	Refuse to classify
1	Correct Classification
0	Incorrect Classification
TH	Thresholding
MLP	Multilayered Perceptron
AB(DS)	Boosting (Decision Stump)
DT	Decision Tree
AB(NB)	Boosting(Naïve Bayes)

5.2 Exam Study

We obtained State Anxiety Inventory (SAI) scores just before and after the end of each exam. We found significant paired differences between the pre-exam and post-exam SAI scores ($p < 0.05$), with the post-exam SAI scores being significantly lower than the pre-exam SAI scores Figure 5.1.

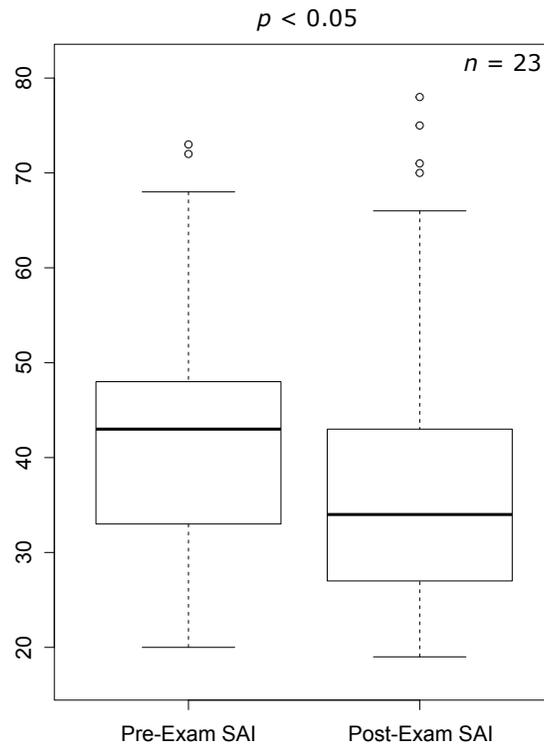


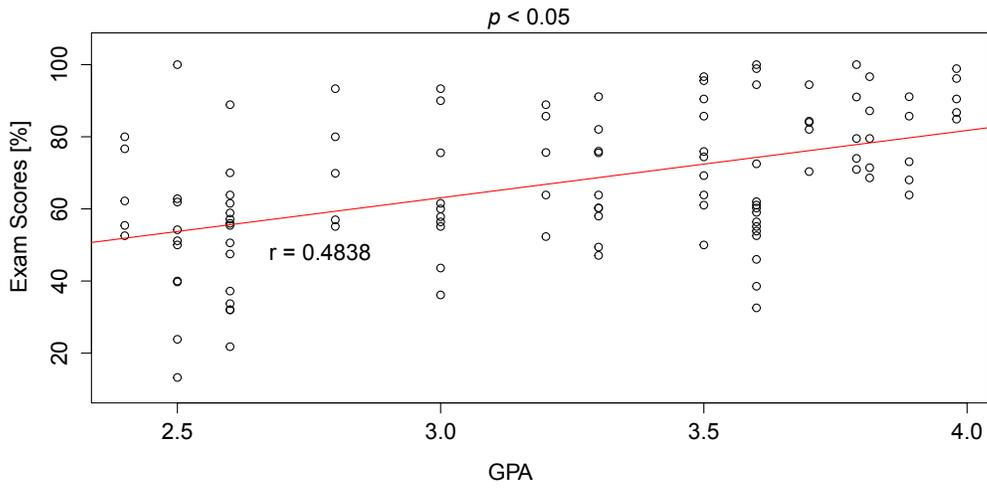
Figure 5.1: Distribution of SAI scores before (Pre) and after (Post) the exams for all subjects.

Students arrived at the classroom with elevated (anticipatory) stress that subsided once the exam was over. This is an unmistakable sign of a stressful experience,

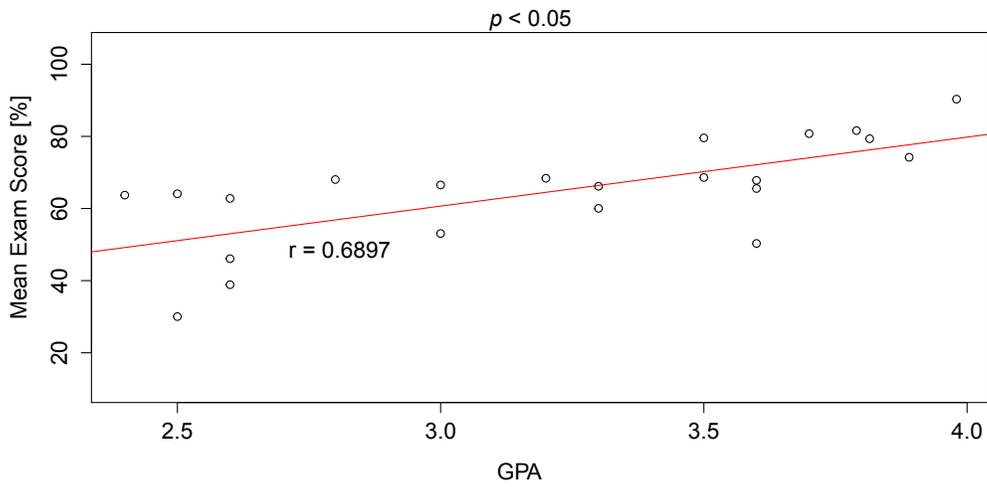
which supports the validity of this study's design.

The main analysis depends on two variables: student performance in the exams and student stress caused during the exams. Because each of the 5 exams were out of a different maximum possible score, we used percent obtained ($\text{Score Obtained} / (\text{Maximum Possible Score}) * 100$) to normalize the exam score. Per the syllabus, the five exams were weighted with (15%, 15%, 15%, 15%, 40%) to compute the the students final score. By regressing Exam Scores over GPA we find that there is a significant correlation between the two (Figure 5.2). In other words, the student's quality predicts her/his course performance, as expected. This is one more indicator of normalcy in the study data.

Next, we put to the test the exam difficulty ranking suggested by the instructor. The standard way to do this is by comparing the subjective ranking with a ranking based on an exam performance measure. For each exam, we calculated the percent exam score ($\text{Score Obtained} / (\text{Maximum Possible Score}) * 100$). Figure 5.3(a) shows the distribution of percent obtained scores of all the students enrolled in the three semesters per exam. Figure 5.3(b) shows the distribution of percent obtained scores of the 23 students participating in the study across the three semesters per exam. From the exam scores obtained by the students, we see that the objective exam score ranking and subjective exam difficulty are in agreement. Further more, the subjects participating in the study seem to be a representative sample of the students in the class.

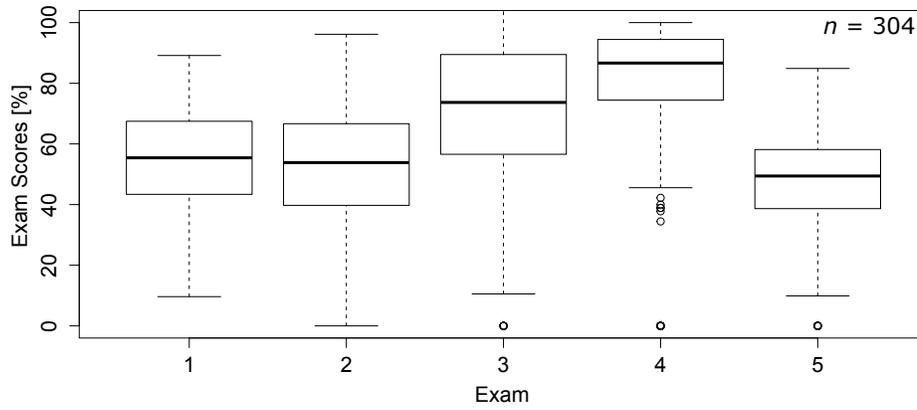


(a)

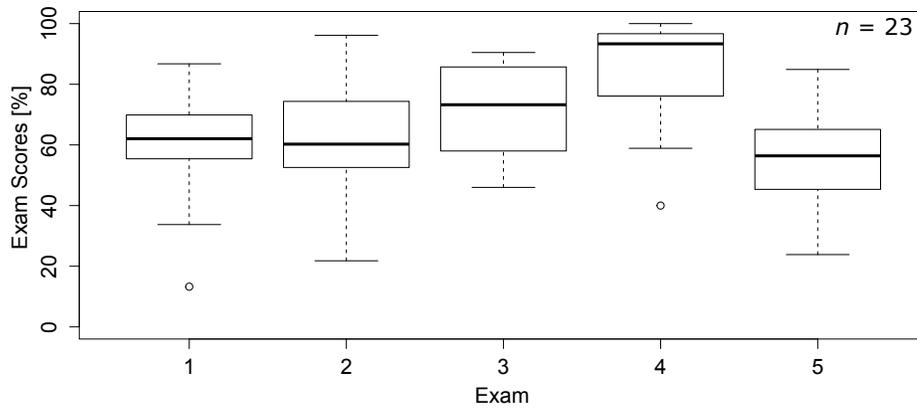


(b)

Figure 5.2: Figure 5.2(a) represent scatterplots of exam scores vs. student GPA for all 5 exams attempted by the 23 subjects. Figure 5.2(b) represent the scatterplots of the mean exam scores vs. student GPA.



(a)



(b)

Figure 5.3: 5.3(a) represent the percent score distribution of all students in the class. 5.3(b) represent the percent score of the 23 students who participated in the study.

Next, we test if any of the physiological or kinetic indicators of sympathetic arousal can track exam performance, much like the GPA does. Figure 5.4 depicts the exam boxplots for the 8 sympathetic indicators we recorded. We observe that

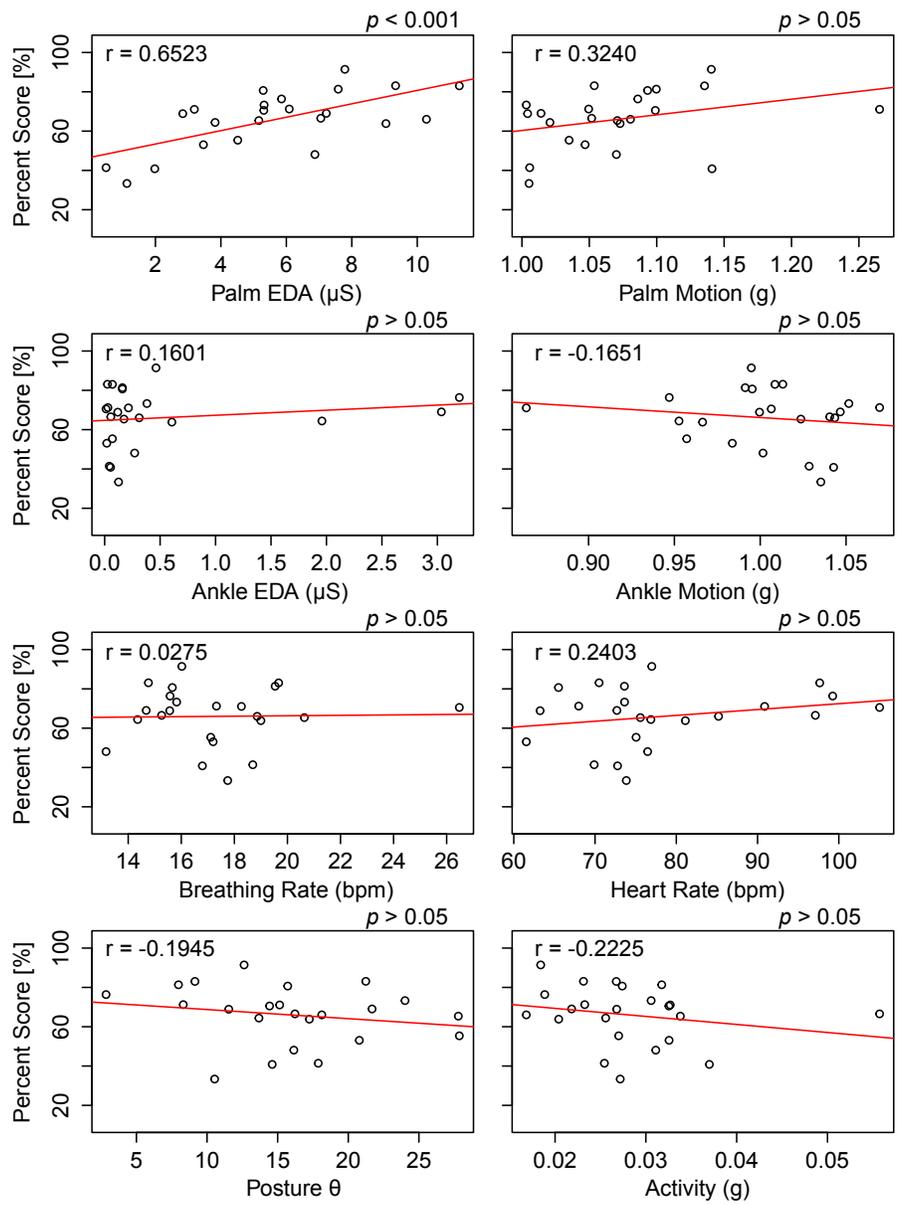


Figure 5.4: Mean physiological and kinetic indicators of sympathetic arousal per student.

only palm EDA appears to track exam performance. There appears to be a high correlation ($r=0.6523$) between the mean palm EDA per student and the mean percent score per exam.

Because palm EDA seems to be able to predict exam performance, we further explore the existence of any correlation between palm EDA and student's GPA. Figure 5.5 displays the distribution of mean palm EDA per exam for each of the GPA's of the participating students. Figure 5.6 displays the mean palm EDA per student against GPA. There exists a positive correlation ($r = 0.4032$) between palm EDA and GPA. Figure 5.7 displays the mean palm EDA per student against mean percent score achieved. The correlation between these variables is even higher ($r = 0.6523$). The GPA of the student is a measure of their academic achievement and is indicative of their ability. The positive correlations indicate that students scoring higher, certainly do so due to the fact that they have prepared themselves for the exams, however these students perform as such during their exam at a much higher sympathetic cost. That is, they care more about the outcome of the exam and hence stress more. On the other hand, students performing poorer seem to care less or are satisfied with achieving lower, hence they exhibit lower sympathetic arousal.

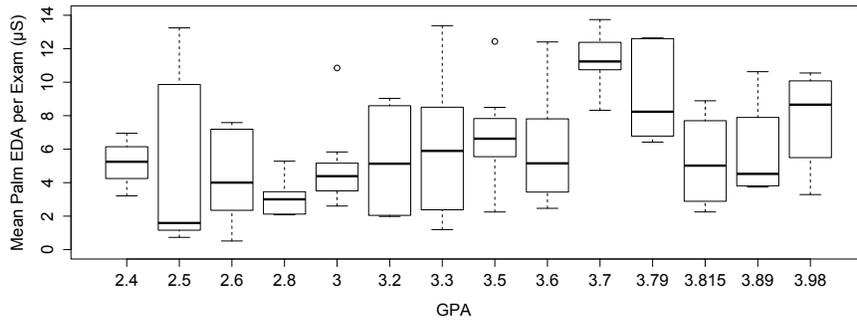


Figure 5.5: Distribution of mean palm EDA per exam per GPA.

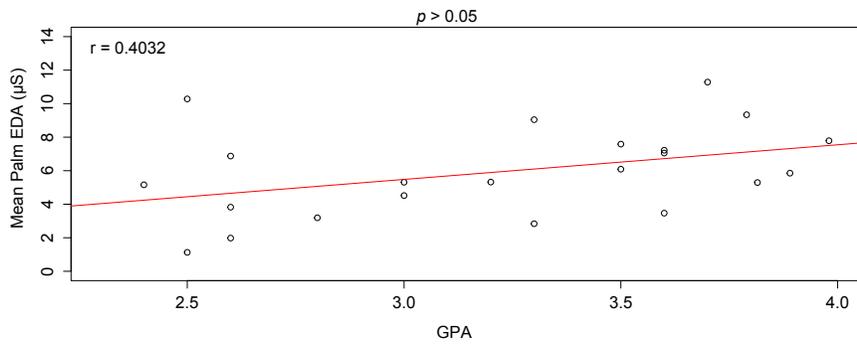


Figure 5.6: Mean palm EDA per student per GPA.

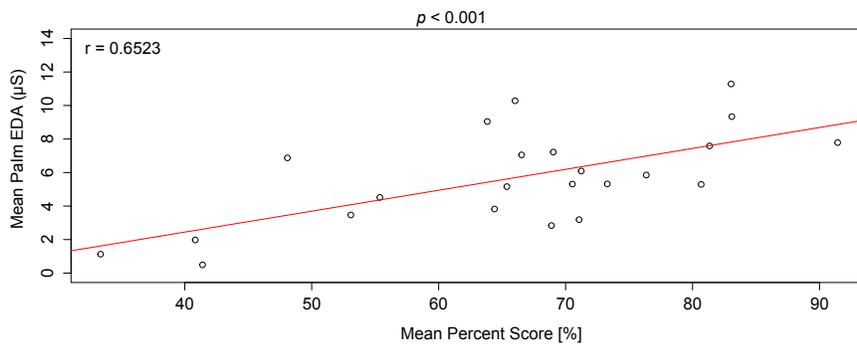


Figure 5.7: Mean palm EDA per student against mean percent score.

5.3 Driving Study

We measured the strength of the linear relationships on the scatterplots with the Pearson’s product moment correlation coefficient and performed the respective test of significance. We did hypothesis testing against a two-tail alternative, setting levels of significance at $= 0.05$ designated by *, or $= 0.01$ designated by **, or $= 0.001$ designated by ***.

Introductory Sessions Analysis

Comparison among the perinasal perspiration signals in the baseline session, the practice drive, and the relaxing drive indicated the absence of any significant differences ($p < 0.05$, analysis of variance). This suggests that in all these cases, subjects were hovering close to their tonic levels, in the absence of any serious challenge.

Loaded Drives Analysis

We analyzed perinasal perspiration (explanatory variable) to determine the sympathetic effect of distractions in loaded drives. Next, we analyzed steering angle (response variable I) to determine how sympathetic effects associate with attempted actions. Finally, we analyzed lane departure (response variable II) to ascertain how attempted actions are modulated, shaping error prone driving behaviors.

Specifically, for the explanatory variable we computed the mean perinasal perspiration signal intensity $\overline{E(k, LD_j, P_i)}$ for each driving phase P_i , of each loaded drive LD_j , for each subject k . These values represented the mean sympathetic arousals exhibited by the subjects in response to the presence or absence of stress stimuli.

For response variable I, we computed the mean angular steering deviation (in absolute terms) $|\overline{\text{ST}(k, \text{LD}_j, \text{P}_i)}|$ for each driving phase P_i , of each loaded drive LD_j , for each subject k . These steering values served as indicators of attempted actions. Given that the subjects were traveling on a straight highway for 6 out of the 6.75 mi of the drive, the mean absolute steering value should have been close to zero; the further away from zero, the stronger the sympathetic effect on instantaneous motor responses.

For response variable II, we computed the range of lane departures $X(k, \text{LD}_j, \text{P}_i)$ for each driving phase P_i , of each loaded drive LD_j , for each subject k . Here we define lane departure as the distance of the car's center from the right edge of the road. Ideally, the driver should maintain a nearly constant distance from this reference edge, driving in the middle of her/his lane ($X \approx 0$). If her/his lateral position deviates significantly, then the range of lane departure values X would change substantively (erratic driving).

For each subject, we normalized the explanatory and response variables with respect to the corresponding variables in the LD_0 drive that featured no stressor. These LD_0 baselines represented the subject's sympathetic state, steering performance, and driving performance under normal conditions. Since the itinerary and environment remained the same in all loaded drives, any mean deviations from the subject's LD_0 baselines should be attributed to the forced distractions.

Effect of Cognitive Load on Sympathetic State, Steering and Driving Performance

For our sample of 59 subjects, we computed for each driving phase P_i the distributions of paired differences between:

- Mean perinasal perspiration in LD_C and LD_\emptyset (Eq. 5.1) - proxy for sympathetic changes
- Mean absolute steering angle in LD_C and LD_\emptyset (Eq. 5.2) - proxy for steering changes
- Range of lane departures in LD_C and LD_\emptyset (Eq. 5.3) - proxy for driving changes

$$\Delta \ln(\bar{E}(\cdot, C, P_i)) = \ln(\bar{E}(k, LD_C, P_i)[^\circ C^2]) - \ln(\bar{E}(k, LD_\emptyset, P_i)[^\circ C^2]) \quad (5.1)$$

$$\Delta \ln(|\overline{ST}(\cdot, C, P_i)|) = \ln(|\overline{ST}(k, LD_C, P_i)|[\text{rad}]) - \ln(|\overline{ST}(k, LD_\emptyset, P_i)|[\text{rad}]) \quad (5.2)$$

$$\Delta X(\cdot, C, P_i) = X(k, LD_C, P_i)[\text{m}] - X(k, LD_\emptyset, P_i)[\text{m}] \quad (5.3)$$

From Equation 5.1 we obtained the first row of boxplots in Fig. 5.8, which suggests that cognitive distraction of subjects in phases $P2_{LD_C}$ and $P4_{LD_C}$ had as a result significant elevation of their mean sympathetic arousal, with respect to phases $P2_{LD_\emptyset}$ and $P4_{LD_\emptyset}$ in the no-stressor drive ($p < 0.001$, paired t-tests in both cases).

From Equation 5.2 we obtained the second row of boxplots in Fig. 5.8, which suggests that cognitive distraction of subjects in phases $P2_{LD_C}$ and $P4_{LD_C}$ had as a result significant deterioration in mean steering performance, always with respect to phases $P2_{LD_\emptyset}$ and $P4_{LD_\emptyset}$ in the no-stressor drive ($p < 0.001$ in $P2_{LD_C}$ and $p < 0.05$ in $P4_{LD_C}$, paired t-tests in both cases). It is interesting that deterioration in mean steering performance remained significant in phases $P3_{LD_C}$ and $P5_{LD_C}$ with respect to phases $P3_{LD_\emptyset}$ and $P5_{LD_\emptyset}$ in the no-stressor drive ($p < 0.001$, paired t-tests in both cases), indicating that there was a lingering behavioral effect on subjects wrt response variable I, which outlived each application of the cognitive stressor.

From Equation 5.3 we obtained the first row of boxplots in Fig. 5.9, which suggests that cognitive distraction of subjects in phase $P4_{LD_C}$ had as a result significant improvement in the range of lane departures, with respect to phase $P4_{LD_\emptyset}$ in the no-stressor drive ($p < 0.001$, paired t-test). This behavioral effect on subjects wrt response variable II tended to linger in $P5_{LD_C}$, outliving the application of the cognitive stressor ($p < 0.01$, paired t-tests).

Effect of Emotional Load on Sympathetic State, Steering and Driving Performance

For our sample of 59 subjects, we computed for each driving phase P_i the distributions of paired differences between:

- Mean perinasal perspiration in LD_E and LD_\emptyset (Eq. 5.4) - proxy for sympathetic changes
- Mean absolute steering angle in LD_E and LD_\emptyset (Eq. 5.5) - proxy for steering changes
- Range of lane departures in LD_E and LD_\emptyset (Eq. 5.6) - proxy for driving changes

$$\Delta \ln(\bar{E}(\cdot, E, P_i)) = \ln(\bar{E}(k, LD_E, P_i)[^\circ C^2]) - \ln(\bar{E}(k, LD_\emptyset, P_i)[^\circ C^2]) \quad (5.4)$$

$$\Delta \ln(|\overline{ST}(\cdot, E, P_i)|) = \ln(|\overline{ST}(k, LD_E, P_i)|[\text{rad}]) - \ln(|\overline{ST}(k, LD_\emptyset, P_i)|[\text{rad}]) \quad (5.5)$$

$$\Delta X(\cdot, E, P_i) = X(k, LD_E, P_i)[\text{m}] - X(k, LD_\emptyset, P_i)[\text{m}] \quad (5.6)$$

Using Equation (5.4) we produced the third row of boxplots in Fig. 5.8, which suggests that emotional distraction of subjects in phases $P2_{LD_E}$ and $P4_{LD_E}$ had as a result significant elevation of their mean sympathetic arousal, with respect to phases $P2_{LD_\emptyset}$ and $P4_{LD_\emptyset}$ in the no-stressor drive ($p < 0.001$, paired t-tests in both cases).

Using Equation (5.5) we produced the fourth row of boxplots in Fig. 5.8, which suggests that emotional distraction of subjects in phases $P2_{LD_E}$ and $P4_{LD_E}$ had as a result significant deterioration in mean steering performance, always with respect to phases $P2_{LD_0}$ and $P4_{LD_0}$ in the no-stressor drive ($p < 0.01$ in $P2_{LD_E}$ and $p < 0.05$ in $P4_{LD_E}$, paired t-tests in both cases). It is interesting that deterioration in mean steering performance remained significant in phase $P3_{LD_E}$ with respect to phase $P3_{LD_0}$ in the no-stressor drive ($p < 0.05$, paired t-test - Fig. 5.8), indicating that there was a lingering behavioral effect on subjects wrt response variable I, which outlived the first application of the emotional stressor. This lingering behavioral effect did not appear in phase $P5_{LD_E}$, after the second application of the emotional stressor.

Using Equation (5.6) we produced the second row of boxplots in Fig. 5.9, which suggests that emotional distraction of subjects in phases $P2_{LD_E}$ and $P4_{LD_E}$ had as a result significant improvement in the range of lane departures, with respect to phases $P2_{LD_0}$ and $P4_{LD_0}$ in the no-stressor drive ($p < 0.001$, paired t-tests in both cases).

Effect of Sensorimotor Load on Sympathetic State, Steering and Driving Performance

For our sample of 59 subjects, we computed for each driving phase P_i the distributions of paired differences between:

- Mean perinasal perspiration in LD_M and LD_\emptyset (Eq. 5.7) - proxy for sympathetic changes
- Mean absolute steering angle in LD_M and LD_\emptyset (Eq. 5.8) - proxy for steering changes
- Range of lane departures in LD_M and LD_\emptyset (Eq. 5.9) - proxy for driving changes

$$\Delta \ln(\overline{E}(\cdot, M, P_i)) = \ln(\overline{E}(k, LD_M, P_i)[^\circ C^2]) - \ln(\overline{E}(k, LD_\emptyset, P_i)[^\circ C^2]) \quad (5.7)$$

$$\Delta \ln(|\overline{ST}(\cdot, M, P_i)|) = \ln(|\overline{ST}(k, LD_M, P_i)|[\text{rad}]) - \ln(|\overline{ST}(k, LD_\emptyset, P_i)|[\text{rad}]) \quad (5.8)$$

$$\Delta X(\cdot, M, P_i) = X(k, LD_M, P_i)[\text{m}] - X(k, LD_\emptyset, P_i)[\text{m}] \quad (5.9)$$

Using Equation (5.7) we produced the fifth row of boxplots in Fig. 5.8, which suggests that sensorimotor distraction of subjects in phases $P2_{LD_M}$ and $P4_{LD_M}$ had as a result significant elevation of their mean sympathetic arousal, with respect to

phases $P2_{LD_0}$ and $P4_{LD_0}$ in the no-stressor drive ($p < 0.001$ in $P2_{LD_M}$ and $p < 0.01$ in $P4_{LD_M}$, paired t-tests in both cases).

Using Equation (5.8) we produced the sixth row of boxplots in Fig. 5.8, which suggests that sensorimotor distraction of subjects in phases $P2_{LD_M}$ and $P4_{LD_M}$ had as a result significant deterioration in mean steering performance, always with respect to phases $P2_{LD_0}$ and $P4_{LD_0}$ in the no-stressor drive ($p < 0.001$, paired t-tests in both cases). It is interesting that deterioration in mean steering performance remained significant in phases $P3_{LD_M}$ and $P5_{LD_M}$ with respect to phases $P3_{LD_0}$ and $P5_{LD_0}$ in the no-stressor drive ($p < 0.001$, paired t-tests in both cases), indicating that there was a lingering behavioral effect on subjects wrt response variable I, which outlived each application of the sensorimotor stressor.

Using Equation (5.9) we produced the third row of boxplots in Fig. 5.9, which suggests that sensorimotor distraction of subjects in phases $P2_{LD_M}$ and $P4_{LD_M}$ had as a result significant deterioration in the range of lane departures, with respect to phases $P2_{LD_0}$ and $P4_{LD_0}$ in the no-stressor drive ($p < 0.001$ in $P2_{LD_M}$ and $p < 0.05$ in $P4_{LD_M}$, paired t-tests in both cases). This behavioral effect on subjects wrt response variable II tended to linger in $P3_{LD_M}$, outliving the first application of the sensorimotor stressor ($p < 0.05$, paired t-test).

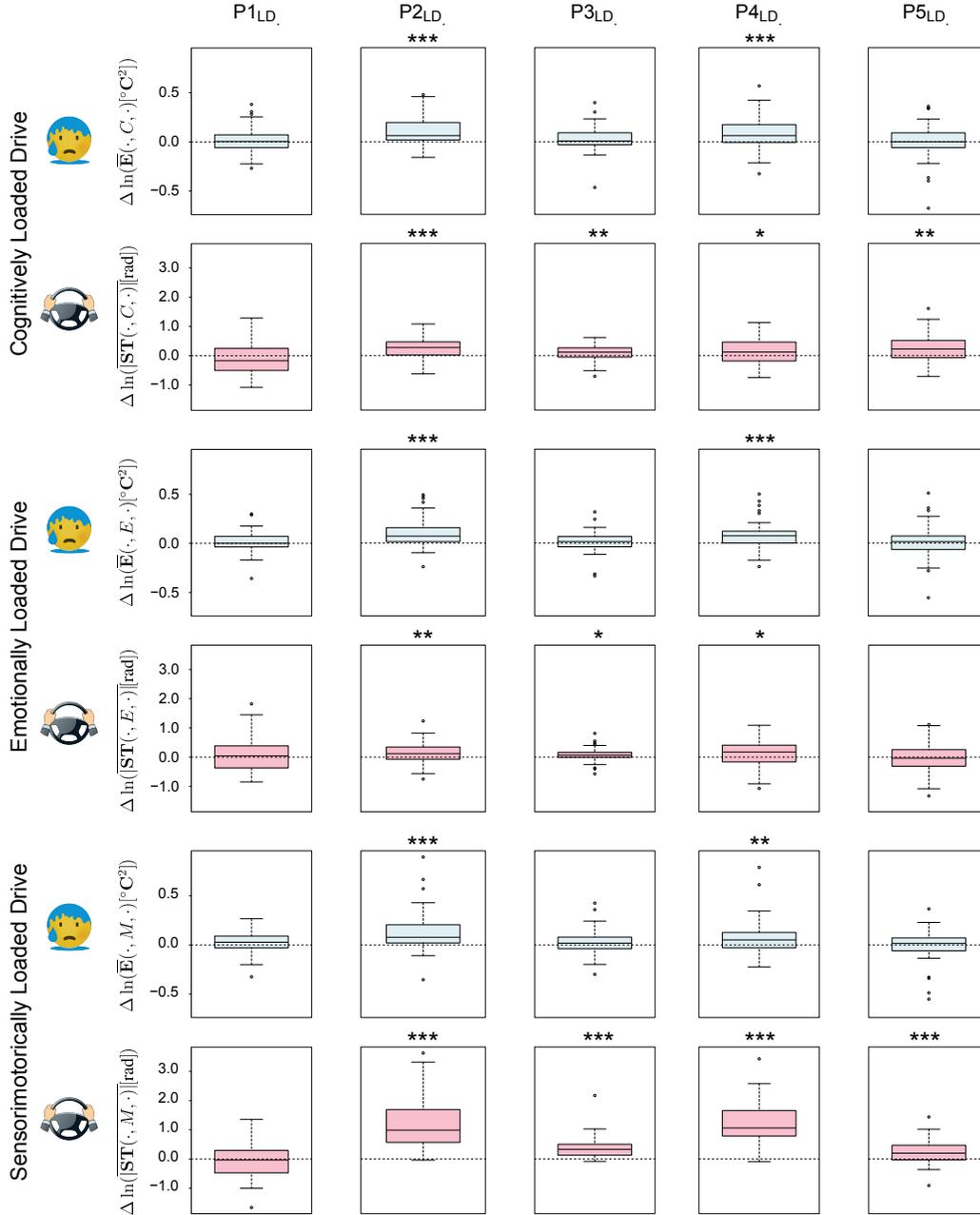


Figure 5.8: Paired tests for the explanatory (perinasal perspiration) and response I (steering) variables in each phase of the cognitively, emotionally, and sensorimotorically loaded drives.

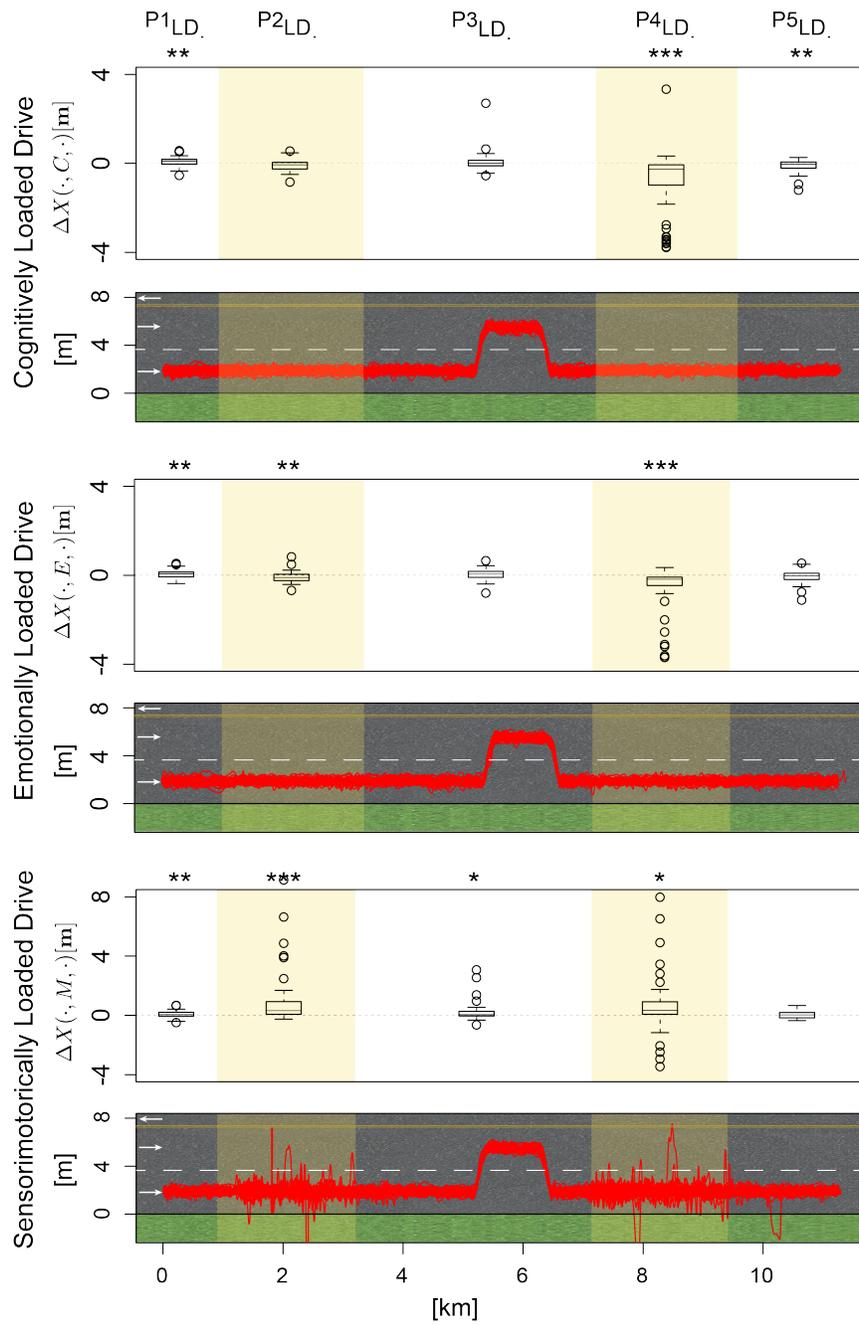


Figure 5.9: Paired tests for the response II (lane departure) variable in each phase of the cognitively, emotionally, and sensorimotorically loaded drives.

Chapter 6

Conclusion

6.1 Deception Detection Study

To the best of our knowledge, this is the first research effort that investigated the value of facial perspiration in high-stakes deceptive behavior. We validated the proposed framework for a mock crime experiment where the subjects faced intense investigation. A pool of 67 subjects were used in our analysis (25 for training and 42 for testing). We devised a threshold-based classifier and modeled four machine learning classifiers. The classifiers scored above 80% successful prediction rate on the training set and close to 80% successful prediction rate on the testing set, indicating that the proposed method scales up.

This research makes two significant contributions in the field of deception analysis: one at the feature level and one at the system level. At the feature level, we

demonstrated that high-stakes lying causes detectable changes in facial perspiration patterns. Specifically, most deceptive subjects in our experiment exhibited significant increase in perspiration frequency when faced with questions related to the mock crime. Most truthful subjects, on the other hand, did not show any substantial changes in perspiration frequency between the relevant and irrelevant question sets. These findings indicate that the perspiration frequency can be used as a discriminating feature for classifying deceptive from truthful behavior. Such a discrete system for interrogating individuals can also be used to determine any mal-intent as demonstrated in [7].

Our contribution at the system level is the proposed thermal imaging-based deception detection framework. The framework features unobtrusive measurement, rapid analysis, and generalizable classifiers. Specifically, the framework quantifies facial perspiration responses in a contact-free manner. Thus, it eliminates the need for contact probing that may compromise the validity of a sympathetic measurement. Given a thermal video with a synced interview audio, the framework offers a semi-automated process for deception detection. In particular, its image processing-based perspiration extraction module is near real-time. Only its signal processing-based feature extraction module needs some human intervention for audio demarcations. This process, however, requires only a few mouse clicks and thus, it is fast. Most importantly, the proposed framework features a generalizable classification method. Unlike the previous deception detection approaches that are limited to a specific

interview question or a set of interview questions [14] [17] [29], the proposed classifiers operate on the difference in perspiratory responses between the relevant and irrelevant question sets. Thus, this approach makes the framework a natural fit to behavior analysis interviews (BAI) with any number of irrelevant and relevant questions.

6.2 Exam Study

It is interesting that Palm EDA (a cholinergic response), is the only sympathetic indicator of consequence in this cognitive performance vs. stress study. The adrenergic indicators (breathing and pulsation) do not hold any differentiating power. This is obvious even by causal observation of the adrenergic signals, which are much less variable than the cholinergic signals. The exams induce moderate sympathetic arousal to the students, enough to cause instantaneous perspiration bouts, but not sufficiently strong to accelerate their cardiopulmonary rhythms at times.

Cholinergic responses also proved to be highly informative in a study of dexterous performance vs. stress, tracking level of experience, and hence perceived task difficulty [18]. Provided that the subjects are pre-occupied with an intensive task, and thus the context is well defined, the cholinergic channel emerges as the affective channel of choice across moderately stressful domains. One might wonder why the cholinergic signal on the ankle did not hold any differentiating power. This has to do with the sensor placement. In the lower limbs, instantaneous perspiratory responses during sympathetic arousal are maximal on the underfoot and wear off with distance. Hence, at the ankle are minimal. Unfortunately, there is no Q sensor form for underfoot placement and this is the best we could do.

There has been some work in education about dynamic adjustment of exam difficulty [11] [21]. Thus far, this adjustment is based solely on actual or anticipated

grade performance, which is domain specific and is not always real-time. The research presented here complements and transforms the state of the art by introducing the notion of ‘attainment cost’, which qualifies grade performance. On similar lines, Taamneh et al. investigated what sympathetic arousals can tell about children’s performance in reading. [28].

6.3 Driving Study

Driver safety is improved when the driver operates sensibly and her/his environment does not change abruptly. An interesting question is what happens if either of these conditions is not met. Here we restricted ourselves to the study of distracted driving and unintended acceleration in busy thorough fares. Although distracted driving is not the only form of non-sensible driving (e.g., driving under intoxication is another infamous variety), it is certainly the most prevalent, especially during rush hours, when the individual effects on traffic flow are maximized.

Pivotal to our approach is the abstraction of distracted driving into three main categories, depending on the stressor involved: cognitive, emotional, and sensorimotor. This is a comprehensive but diverse stressor set; thus, sensing its physiological effects through a universal indicator can streamline the measurement process rendering future applications practical. We used perinasal perspiration as a measure of sympathetic arousal - a prime indicator of stressor effects, irrespectively of the stressor type. We extracted this perspiratory signal using a clinically validated method based on thermal imaging [26]. The sensing modality rendered the physiological measurement process totally unobtrusive.

We measured the direct sympathetic effect on driving using the evolving absolute value of the steering angle. We measured the filtered effect on driving using the range of involuntary lane departures - an indicator that tracks propensity for error and thus, accident.

In a simulator experiment, designed to isolate each stressor type, we found that all three stressors resulted in significant increases of the drivers' sympathetic arousal levels, all other things being equal (i.e., itinerary and traffic conditions). Furthermore, we found that these elevated arousal levels were associated with significant increases in the mean value of the absolute steering signal - an indication of attempted erratic action directly driven by sympathetic arousal. Interestingly, we also found that these attempted erratic actions were overcorrected when the hand-eye feedback loop was not interrupted; this was true in the cognitive and emotional stressor cases. A likely explanation for this paradox is that cognitive or emotional conflict activated the anterior cingulate cortex (ACC), which successfully counter-balanced erroneous motor reactions. However, in the case of pure or mixed sensorimotor conflict, where the hand-eye feedback loop was interrupted, ACC filtering was slipping, failing at times to counterbalance instinctive motor reactions and thus, resulting into occasional lane departures.

It appears that moderate levels of pure cognitive or emotional loading have beneficial effects on driving behaviors. This result is intriguing. One should not rush to generalize this conclusion, however, as it is almost certain that extreme cognitive and emotional loads will tilt the scale towards unsafe driving behaviors. The question is where is the threshold. Furthermore, the experiment's results shed light to the likely neurophysiological mechanism that renders texting while driving so disruptive and dangerous, even in moderate amounts; it knocks out human's last line of conflict resolution defense, that is, the anterior cingulate cortex.

A limitation of the perinasal imaging method is that it does not perform reliably when the subject has facial hair. For this reason about 13% of the original data set (nine male subjects) could not be processed. Other methods for peripheral sensing of sympathetic arousal, such as palmar electrodermal activity sensing, have their own set of problems, especially in the context of driving where the subjects' hands are engaged. Further research into measurement methods will solve these problems in due course. What the current study convincingly demonstrates, however, is that distractions, over-arouse the average driver and may result in significant deterioration of her/his driving performance. Furthermore, real-time unobtrusive measurement of driver's arousal and its behavioral effects are within reach, opening the way for engineering orthotic feedback loops. These loops will notify drivers (and perhaps others in the vicinity) of their predicament, which often goes unnoticed because it is subtle and subconscious, yet no less dangerous.

Chapter 7

Appendix

7.1 Overview

We needed to first perform a validation study to measure the effectiveness of the Q Sensor to detect sympathetic arousal from the palms of the students while they took their exams, in the *Exam Difficulty Assessment Study 3.2*. In 2010, the first mobile EDA sensor appeared in the market - Q Sensor (Affectiva, Waltham, Massachusetts). The Q Sensor measured perspiratory responses on the palm or on the wrist. For the former case the sensor was packaged in an open glove form (pod), while for the latter case the sensor was packaged in a wristband form (curve). The Q Sensor was discontinued in 2013, but several other mobile EDA sensors have appeared in the market since then, such as the E3 Wristband (Empatica, Cambridge, Massachusetts) and the Shimmer3 GSR + Optical Pulse Development Kit (Shimmer, Dublin, Ireland).

The brands may differ but the basic technology within each family of EDA instruments (conventional vs. mobile) is the same. Hence, one can pick any model from each family and perform a study that would be highly representative. Comparing features across the two families of instruments one notes both similarities and differences. Both instrument families use Ag/AgCl disc electrodes with contact areas of 1.0 cm^2 for their recordings, as recommended in the literature [9]. However, they differ in terms of power mode (AC vs. DC), packaging (large vs. small form factor), standard communication capabilities (wired vs. wireless), and wearable options. The power mode has some implications in measurement accuracy (AC is better [5], while the wearable options in the mobile family of sensors are both a blessing and a challenge. The challenge stems from the fact that sympathetically induced perspiratory responses differ in their strength among various body locations [5]. Unfortunately, the strongest responses do not necessarily correlate with the most ‘wearable’ body locations. The combination of lesser accuracy with lesser responses may bias results in studies of user affect. Such biases have not been sufficiently appreciated or studied in the literature.

In the foundational paper of mobile EDA sensing, Poh et al. [20] compared the Q Sensor measurements on the wrist against measurements taken with conventional EDA sensors on the fingers; the stimuli included physical activity, mental arithmetic, Stroop, and horror movies. Then, Poh and other researchers from the same lab used the Q Sensor to collect affective data in a number of daily activity studies including reading, walking, and sleeping [20] [24]. They even used the Q Sensor in an epilepsy

study [19]. Chaspari et al. [6] used mobile EDA sensing to model verbal response latencies in autistic children. McDuff et al. [15] used mobile EDA sensing to quantify user emotions in a reflection study of past events. Overwhelmingly, in all these studies researchers used the wristband form of mobile EDA sensing and reported interesting findings.

There has been no rigorous validation of a representative mobile EDA device for various body locations. Poh et al. [20] did validate the Q Sensor against a conventional EDA instrument, but they did this only for the wrist location, and using an array of stimuli that are not considered baseline in sympathetic studies. A universal arousal stimulus is auditory startle, invoking a threshold response on the fingers of healthy subjects [13]. An EDA device that successfully captures startle responses on the fingers, meets the gold standard, as it has the capability of measuring sympathetic arousal of low intensity and minimal duration on a prime neurophysiological site. Conventional EDA devices belong to this category. Even a gold standard device, however, may not capture startle responses on a different body location (e.g., wrist), where there may be lower concentration of perspiration glands and sparser innervation. Hence, an EDA device should be validated for every intended body location.

Using EDA sensors on body locations that have not been validated against the gold standard (i.e., conventional EDA on the fingers), is inherently prone to bias. In a field study with sustained stressors of high intensity, the sensing device will likely show responses, irrespectively of the sensor type and body location. In a field study

with short stressors of mild intensity, if the sensor/location combination does not correspond to the gold standard, the device may not show anything at all, giving the illusion that the sympathetic system is in tonic mode.

7.2 Experimental Design

We conducted this study per a protocol approved by the University of Houston (UH) Institutional Review Board (IRB). We recruited subjects through email solicitations and flyer posting in the UH campus community (population about 35; 000). We excluded children (< 18), subjects with hearing impairments, and subjects on medications. Age brings psychological and physiological changes that are especially prominent during developmental and late years [5]; for this reason, we did not include children and older adults (> 59) in the subject pool. The use of auditory startle stimuli in the experiment necessitated the exclusion of subjects with hearing problems. Certain medications affect sympathetic responses [10]; to simplify screening and minimize confounding factors, we excluded all medication cases. A total of $n = 25$ subjects fulfilling the inclusion/exclusion criteria volunteered for the experiment.

After each subject consented, s/he filled out a biographic questionnaire, the Trait Anxiety Inventory, and the State Anxiety Inventory. The last two meant to check if any pathological or extraordinary conditions were biasing responses. Next, we asked each subject to wash her/his hands and feet prior to sensor attachment. We

waited 10 min for normal moisture levels to reestablish on their skin, before attaching the sensors. This preparation ensured optimal and uniform skin conditions in the attachment areas for all subjects. We used Galvanic Skin Response probes connected to an ADInstruments PowerLab data acquisition unit (ADInstruments, Bella Vista, Australia) as *C* EDA sensors; we used Q Sensor sets (Affectiva, Waltham, Massachusetts) as *M* EDA sensors(16). Figure 7.1 depicts the node configuration.

We attached seven EDA sensors to each participant: three sets of a conventional EDA sensor (*C* EDA) on the participant's right-hand side and four sets of a mobile EDA sensor (*M* EDA) on the participant's left-hand side. We used the Galvanic Skin Response probes from ADInstruments (ADInstruments, Bella Vista, Australia) as *C* EDA sensors and Q Sensor sets as *M* EDA sensors. Specifically, we had the following sensor arrangement (Fig. 7.1):

- Classic EDA on the right hand fingers (*C* Fingers)
- Classic EDA on the right palm (*C* Palm)
- Classic EDA on the right wrist (*C* Wrist)
- Mobile EDA on the left hand fingers (*M* Fingers)
- Mobile EDA on the left palm (*M* Palm)
- Mobile EDA on the left wrist (*M* Wrist)
- Mobile EDA on the left sole (*M* Sole)

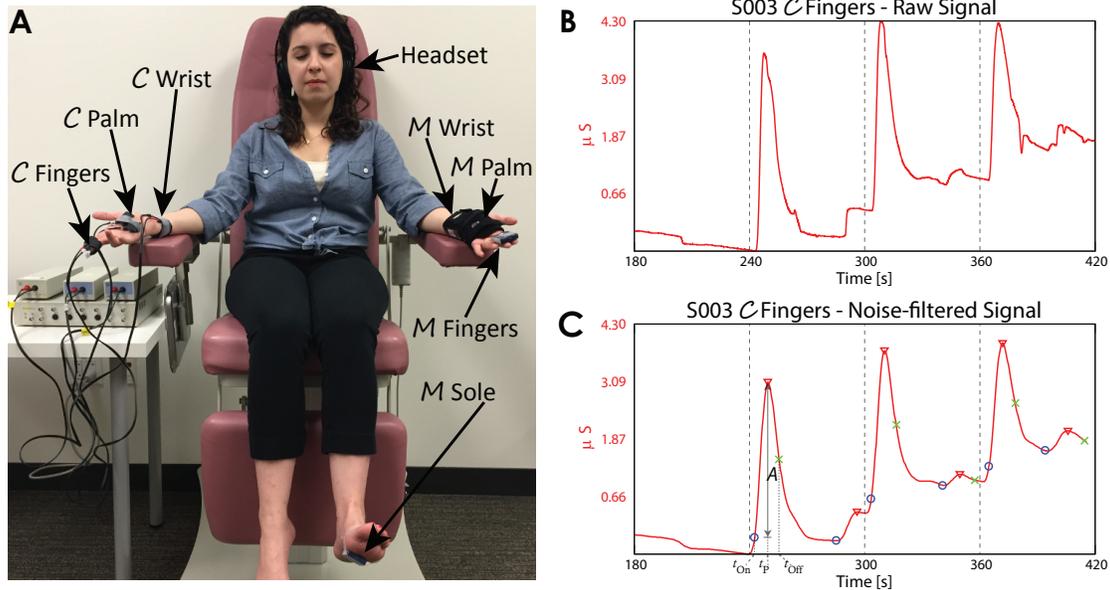


Figure 7.1: Experiment and sample outcome. A, Experimental setup, demonstrating all seven nodes. B, Raw EDA signal on the fingers of subject S003, captured via the conventional sensing device. The dotted lines mark the occurrence of the three stimuli. Only the last minute of the baseline period is depicted to economize space. C, The signal after noise filtering. There are multiple arousals after each stimulus; circles mark onsets, triangles mark peaks, and crosses mark offsets; t_{on} , denotes the time of Onset occurrence, t_p the time of Peak occurrence, and t_{off} the time of Offset occurrence; A stands for the arousals amplitude.

For two of the 25 subjects, the recorded EDA responses in all body locations had a low signal to noise ratio, rendering analysis impossible; we excluded these two subjects from any further consideration. Hence, the usable data set included 23 subjects (11 males / 12 females; age: 23.91 ± 8.12). **Figures 7.2 - 7.5 depict the**

entire data set per location, facilitating cross-sensor comparison where applicable. The main experiment lasted seven minutes for each subject. During these seven minutes, the subject listened to soothing music via a headset in a near dark room. This soothing music was interrupted three times with a glass breaking sound, which served as the auditory startle stimulus. The first stimulus was delivered in the fourth minute. After that, two more stimuli were delivered at one-minute interval each. All sensor measurements were synchronized and recorded throughout the experiment. Hence, our usable data bank accrued 161 EDA signals (23 subjects \times 7 EDA signals per subject - one for each node).

7.3 Methods

The C sensor samples at the rate of 25 points per second while the M sensor samples at the rate of 32 points per second. For this reason, we resampled the M signals down to 25 points per second to establish uniformity across the signal bank. Then, we applied a moving window filter twice to reduce noise. We set the window size at $W = 125$ points. We arrived at this selection by performing sensitivity analysis with window sizes $W = 25 \times k$; $k = 1, 2, 3, \dots$. We chose the size of the increment to be 25 points because it matches the minimum resolution (i.e., rate of sampling). At $k = 5$ we attained optimal performance by substantively reducing noise without destructing signal information; at $k > 5$ the signals exhibited over-smoothing, which affected the performance of the Peak and Onset detection algorithms.

On the noise-filtered signals, we focus on the non-baseline period from $t = 4$ to 7 min. We divide this period into three segments: $S_1 = [4; 5)$ min, $S_2 = [5; 6)$ min, and $S_3 = [6; 7]$ min; each segment includes the respective stimulus S_1 , S_2 , S_3 and the period up to the delivery of the next stimulus or the end of the experiment. Within each segment we seek arousals (firings), each characterized by a Peak and an Onset. An Offset may or may not exist, depending on the recovery rate and the timing of the next arousal. We execute the following algorithmic steps:

1. Peak Detection: We apply a Peak detection algorithm on the non-baseline portion of the signal $S_1 \cup S_2 \cup S_3$.
2. First Onset: Between each stimulus and the first Peak that follows it within the respective segment, we determine the corresponding Onset; this is the Onset of the first arousal in response to the specific stimulus.
3. Additional Onsets; If there are multiple firings in response to a stimulus, then more than one Peak point exist within the respective segment, S_i , $i = 1, 2, 3$. We locate the Onset point corresponding to such an additional Peak in the valley between the present and preceding Peak.
4. Offsets: For each Peak we determine the matching Offset as the $50 \times (\text{Peak} - \text{Onset})$ drop-off point. If this point occurs after the next Onset or stimulus, then it is rejected and the Offset is treated as a missing value; this is an indication that the subject has not recovered at the time a new arousal set in.

7.4 Results

Even under the best circumstances, EDA signals are noisy and variable (Fig. 7.1 B). For this reason, signal abstraction is quintessential to fair comparisons; standard low pass filtering is not sufficient (Fig. 7.1 C). What is of interest here is the ability of the sensor to measure the essence of the stimulus' response at the specific site. A normal neurophysiological response (arousal) can be reconstructed to a good approximation from three key points in the corresponding EDA signal [5]: Onset, Peak, and Offset. The Onset point represents the start of the EDA activation; the Peak point represents the culmination of the activation; and, the Offset point represents the ebbing of the activation.

An electrodermal response may feature more than one Peak. Such peaks correspond to multiple neurophysiological firings provoked by a single stimulus [5]. Hence, for each stimulus a subject can experience none, one, or several peaks (Fig. 7.1 C). Peaks are the most characteristic points of an electrodermal response. Hence, our analysis proceeds at two levels:

- **Detection-level** We pay attention to the occurrence of peaks, as proxies of neurophysiological responses; the absence of peaks signifies the absence of a response at a node.
- **Measurement-Level** We pay attention to the parameters of the neurophysiological responses recorded at the various nodes. These parameters include the times of occurrences for Onset (t_{on}), Peak (t_p), and Offset (t_{off}) as well as

the amplitude of the recorded response ($A = \text{Peak} - \text{Onset}$). The times (t_{on}), (t_p), (t_{off}) quantify the arousal's evolution, while the amplitude quantifies the arousal's intensity.

To ensure that the subjects' trait and state anxiety levels did not affect manifestation of sympathetic activation [16], we obtained their trait (TAI) and state (SAI) anxiety psychometrics [27]. Then, we examined in each node the correlations between the number of peaks detected per subject versus the corresponding TAI and SAI scores. None of these correlations was statistically significant ($p > 0.05$)

Detection-Level Analysis

The numbers of peaks quantify arousal levels, with zero signifying a non-responsive subject in the specific node. Different numbers of subjects exhibit different arousal levels at the different nodes. As we observe in Fig. 7.6, the wrist locations have the highest number of non-responsive subjects irrespectively of the sensor type. This indicates that wrists respond poorly to stimuli. The rest of the locations have minimal numbers of non-responsive subjects, indicating the regular presence of arousals in response to stimuli.

We are interested to examine if indeed the proportion of non-responsive subjects differs significantly among the nodes. To do so we use the Binomial distribution. In each node we sampled 23 subjects, where each subject presented a binary outcome: failure, if s/he was nonresponsive or success, if s/he had at least one arousal. Hence, for each node if we call Y_i the random variable that denotes the outcome of subject

i and θ the unknown probability of success, we have $Y_i|\theta \text{ Bernoulli}(\theta)$, giving:

$$Y_i = \begin{cases} 0 & \text{if no peaks in the entire experiment} \\ 1 & \text{if at least one peak in the entire experiment} \end{cases}$$

We count on the total number of successes in the 23 subjects, forming the random variable $X|\theta \text{ Binomial}(n, \theta), n = 23$:

$$X = \sum_{i=1}^n Y_i$$

As we observe in Fig 7.6, the maximum likelihood estimates (MLE) at the wrists are quiet smaller compared to the other locations. In fact, running a seven-sample test for equality of proportions shows that there are significant differences among the seven nodes ($p < 0.01$). Next, we exclude the two wrist nodes that appear to be the culprit and we compare the proportions in the remaining five nodes. This time the test returns a non-significant number ($p > 0.05$), which indicates that with the exception of the wrist nodes, all the other nodes are statistically equivalent in terms of peak presence.

Measurement-Level Analysis We start the measurement-level analysis by studying the relationship of arousal timing between each pair of nodes, taking into account all three stimuli. Figure 7.7 shows a matrix that is split along its diagonal. The matrix portion below the diagonal shows the scatterplots of Onset (t_{on}), Peak (t_p), and Offset (t_{off}) times for all node pairs. The matrix portion above the diagonal shows the correlation coefficients for the corresponding scatterplots. As indicated

both visually and numerically, the timing agreement between nodes is exceptionally high. Hence, the arousal's evolution is captured accurately at the locations of interest (fingers, palm, wrist, sole), irrespectively of the type of sensor used. One has to note, however, that in the case of wrist nodes, the number of points is small. This is consistent with the finding in Fig. 7.6. Arousal detection at the wrist is rare, but when it occurs, both the C and M sensors track equally well its evolution.

Next we study the relationship of arousal intensity between each pair of nodes, taking into account all three stimuli. Figure 7.8 also shows a matrix that is split along its diagonal. The matrix portion below the diagonal shows the amplitude (A) scatterplots for all node pairs. The matrix portion above the diagonal shows the correlation coefficients for the corresponding scatterplots. As indicated both visually and numerically, the agreement is poor for pairs involving one wrist node, irrespectively of the sensor type attached to this wrist node. Agreement is also poor for pairs involving the mobile sole node. Agreement gets at least moderately strong for pairs involving fingers, palms, or finger-palm, irrespectively of the sensing modality.

This brings to the fore a problem with EDA sensing: while the timing of the phenomenon is accurately captured across locations and sensor types, the magnitude of the phenomenon poses a challenge. This challenge is bigger for the sole and insurmountable for the wrist location, irrespectively of the sensor type.

7.5 Discussion

We recognize the importance of quantifiable and objective information the EDA responses can provide to studies of subject affect. Traditionally, EDA sensing is performed on the fingers with conventional EDA devices. Although, this has serious usability problems, it captures well minimal bursts of sympathetic activation. The new mobile EDA sensors, attached on various body locations, have obvious usability advantages. However, one has to be careful not to compromise measurement accuracy or if s/he does, s/he should at least be aware of it.

In this study we found that in response to minimal standardized sympathetic stimuli, conventional EDA devices are in moderate agreement with mobile EDA devices on the fingers and palm. At the same time, we found that both conventional and mobile EDA devices give significantly inferior measurements when attached to the wrist. Actually, not only measurement but also mere detection of sympathetic responses on the wrist is quite challenging. Given the proliferation of mobile EDA devices, often in the form of affective wristbands, and the accompanying marketing hype, these results are a waking call for a more careful examination of operational limitations. Should this call go unheeded, the introduction of measurement bias in affective studies appears likely.

Another body location - the sole - that is a candidate site for wearable EDA sensing provides strong detection capability, but relatively poor measurement capability.

This study tested sensors and locations in a stationary context. While this was

necessary to minimize confounding factors and establish a clear first-level comparison, it certainly does not account for additional effects that are present in practice. The main such effect is ambulatory motion that needs to be studied in a subsequent study.

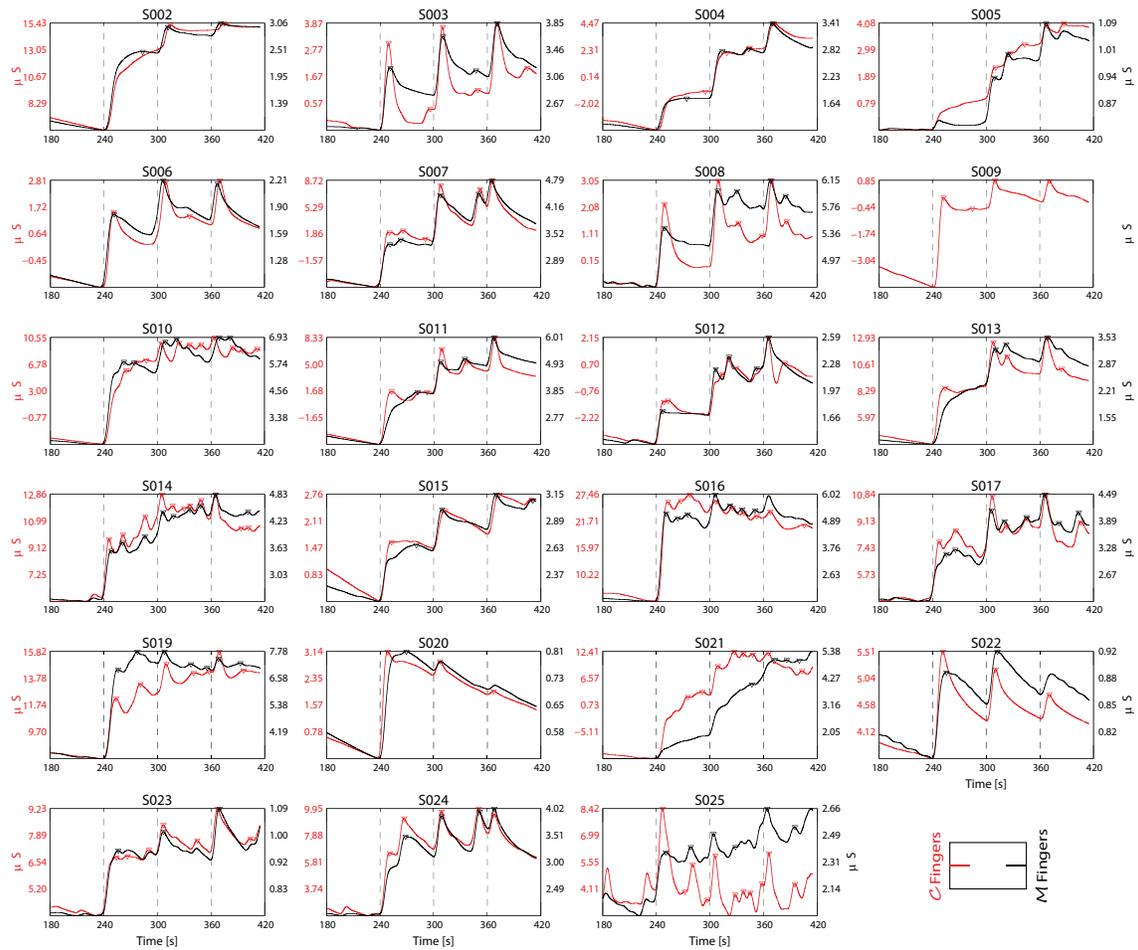


Figure 7.2: *C* and *M* EDA responses on the fingers per subject. From the four minutes of baseline only the last minute is depicted in the graphs to economize space. Vertical dotted lines identify stimuli times and inverted triangles denote peaks, facilitating arousal comparison. The *M* EDA signal for subject S009 was not collected due to technical reasons. The graphs confirm qualitatively the high responsiveness of the fingers location and the agreement between the two sensor modalities.

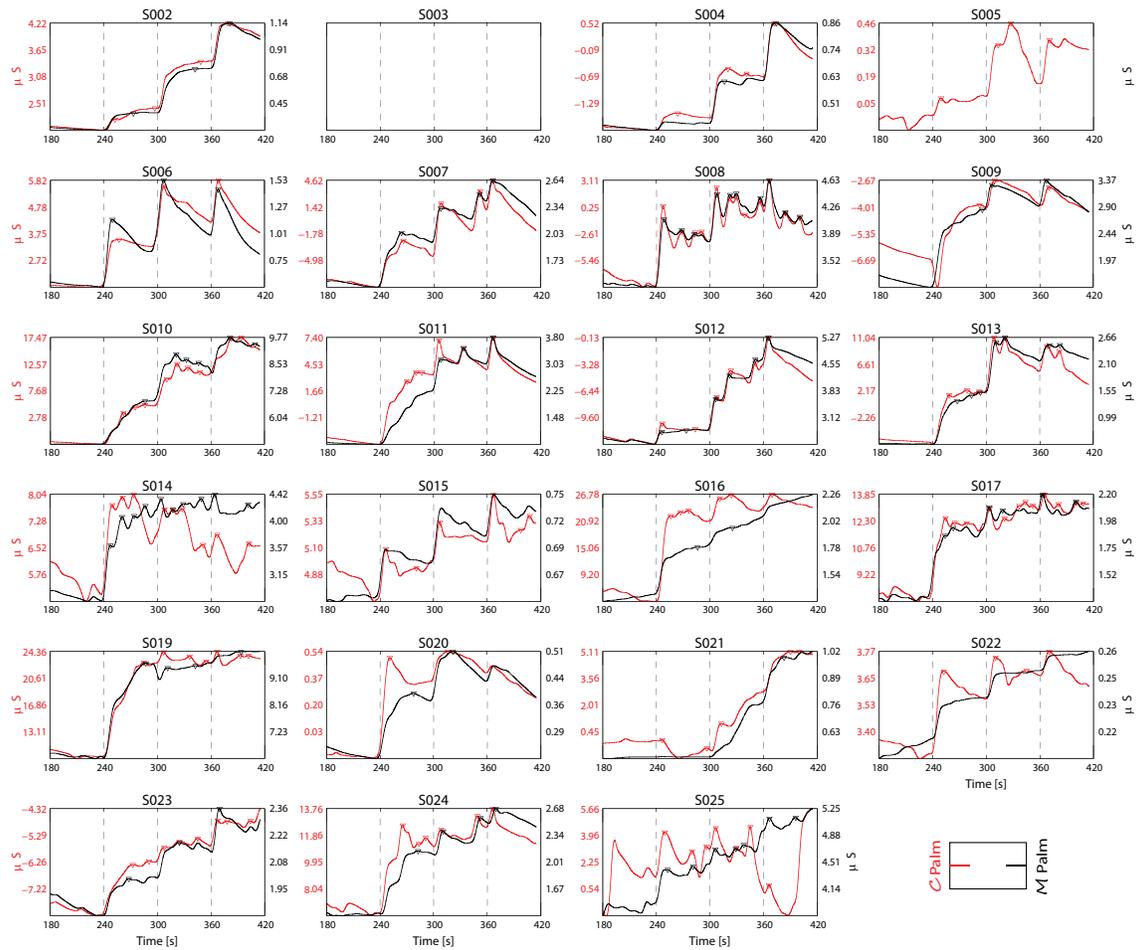


Figure 7.3: *C* and *M* EDA responses on the palm per subject. From the four minutes of baseline only the last minute is depicted in the graphs to economize space. Vertical dotted lines identify stimuli times and inverted triangles denote peaks, facilitating arousal comparison. The *M* EDA signal for subject S005 and both and EDA signals for subject S003 were not collected due to technical reasons. The graphs confirm qualitatively the high responsiveness of the palm location and the agreement between the two sensor modalities.

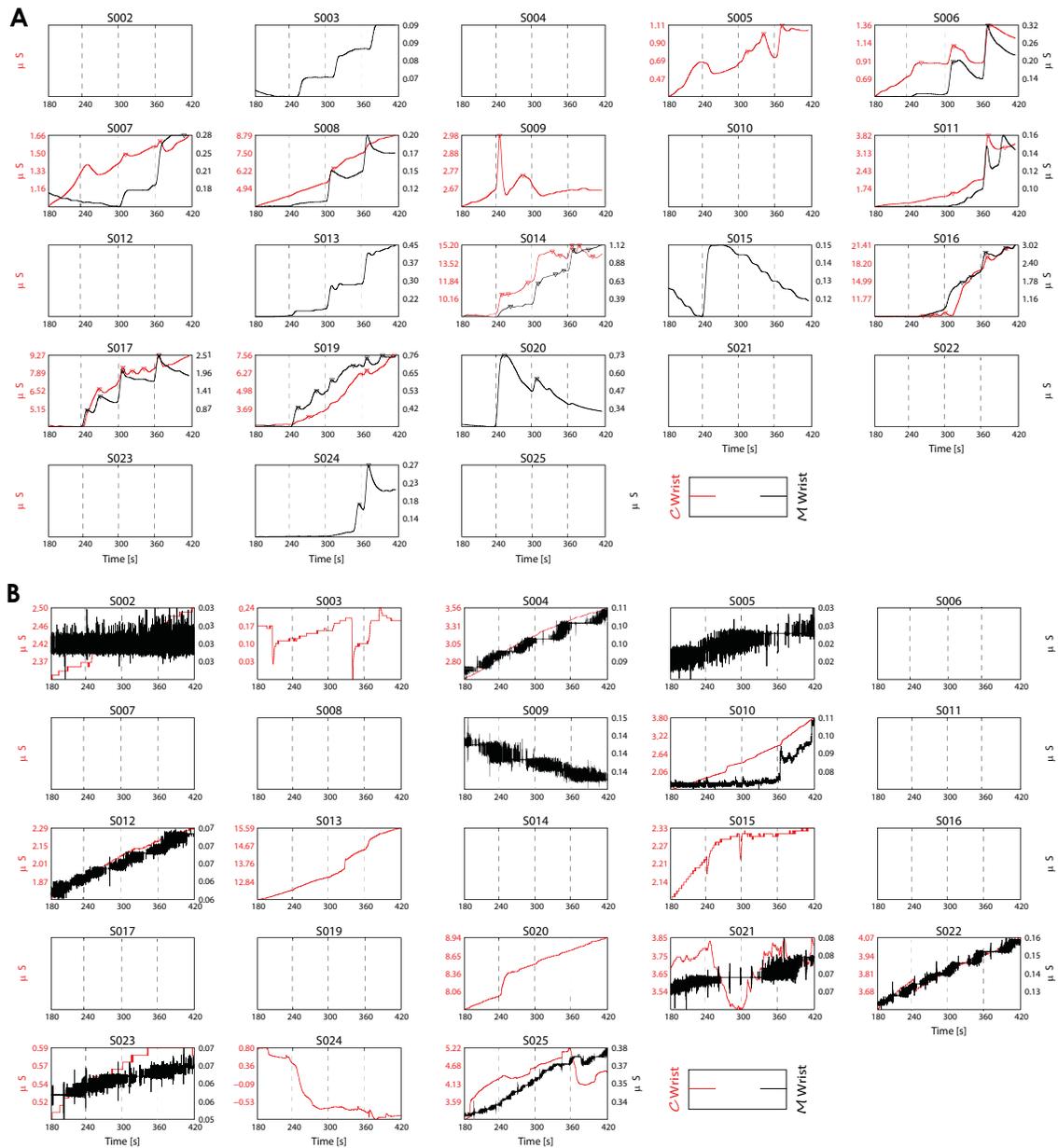


Figure 7.4: *C* and *M* EDA responses on the wrist per subject. From the four minutes of baseline only the last minute is depicted in the graphs to economize space. Vertical dotted lines identify stimuli times and inverted triangles denote peaks, facilitating arousal comparison. **A**, Signals manifesting sympathetic responses on the wrist. **B**, Signals manifesting the absence of sympathetic responses on the wrist, which were excluded from further processing and analysis. The graphs confirm qualitatively the low responsiveness of the wrist location, a result that was documented quantitatively in the text.

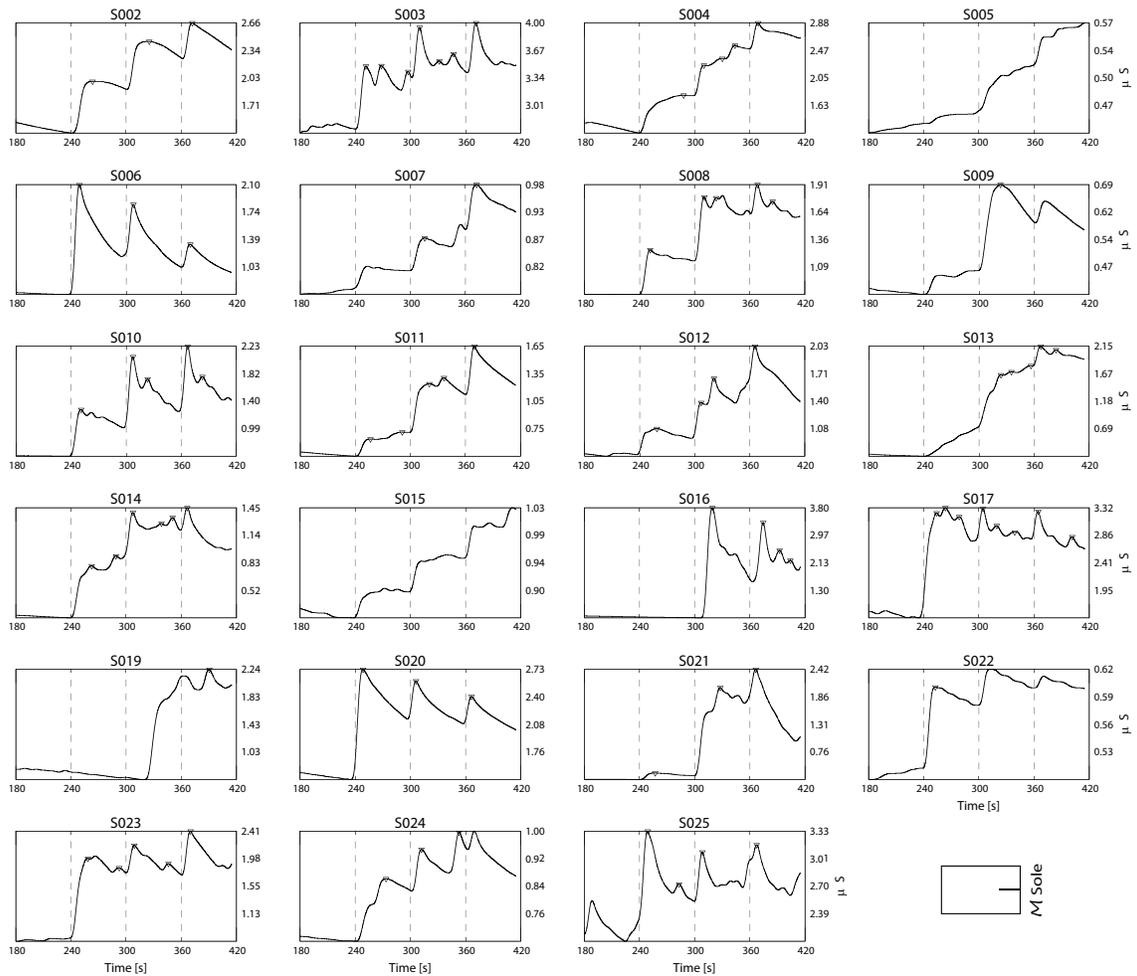


Figure 7.5: *M* EDA responses on the sole per subject. From the four minutes of baseline only the last minute is depicted in the graphs to economize space. Vertical dotted lines identify stimuli times and inverted triangles denote peaks, facilitating arousal comparison. The graphs confirm qualitatively the responsiveness of the sole location, a result that was documented quantitatively in the text. However, amplitude correlations with the finger and palm locations are not strong, suggesting that precise EDA measurements on the sole may be challenging.

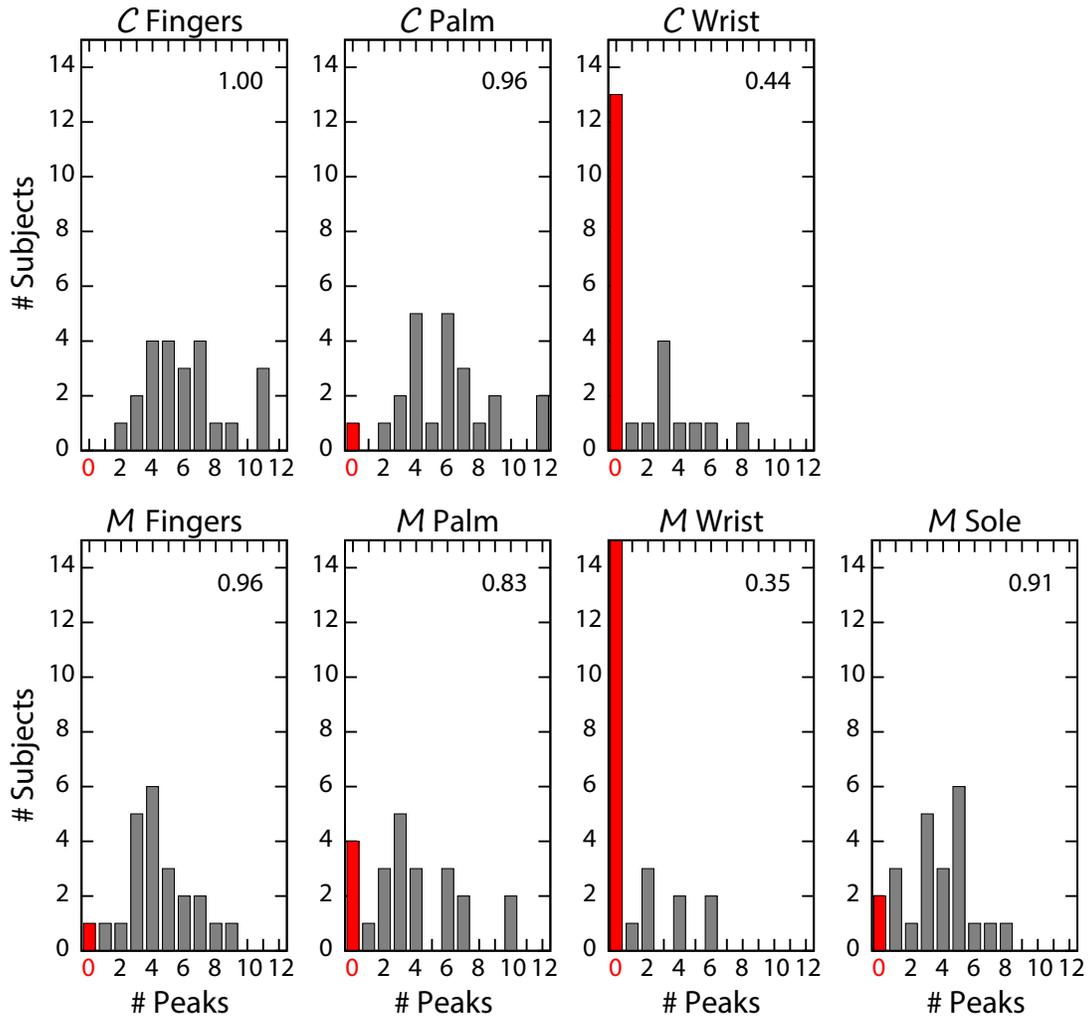


Figure 7.6: Responsiveness of subjects per node. Each bar chart depicts the distribution of the number of subjects over different levels of response (i.e., number of recorded peaks in the node). Red bars indicate the number of completely non-responsive subjects for the specific node. The maximum likelihood estimates (MLE, $\hat{\theta} = X/n$) for getting at least one peak appear on the upper right corner of the node's panel.

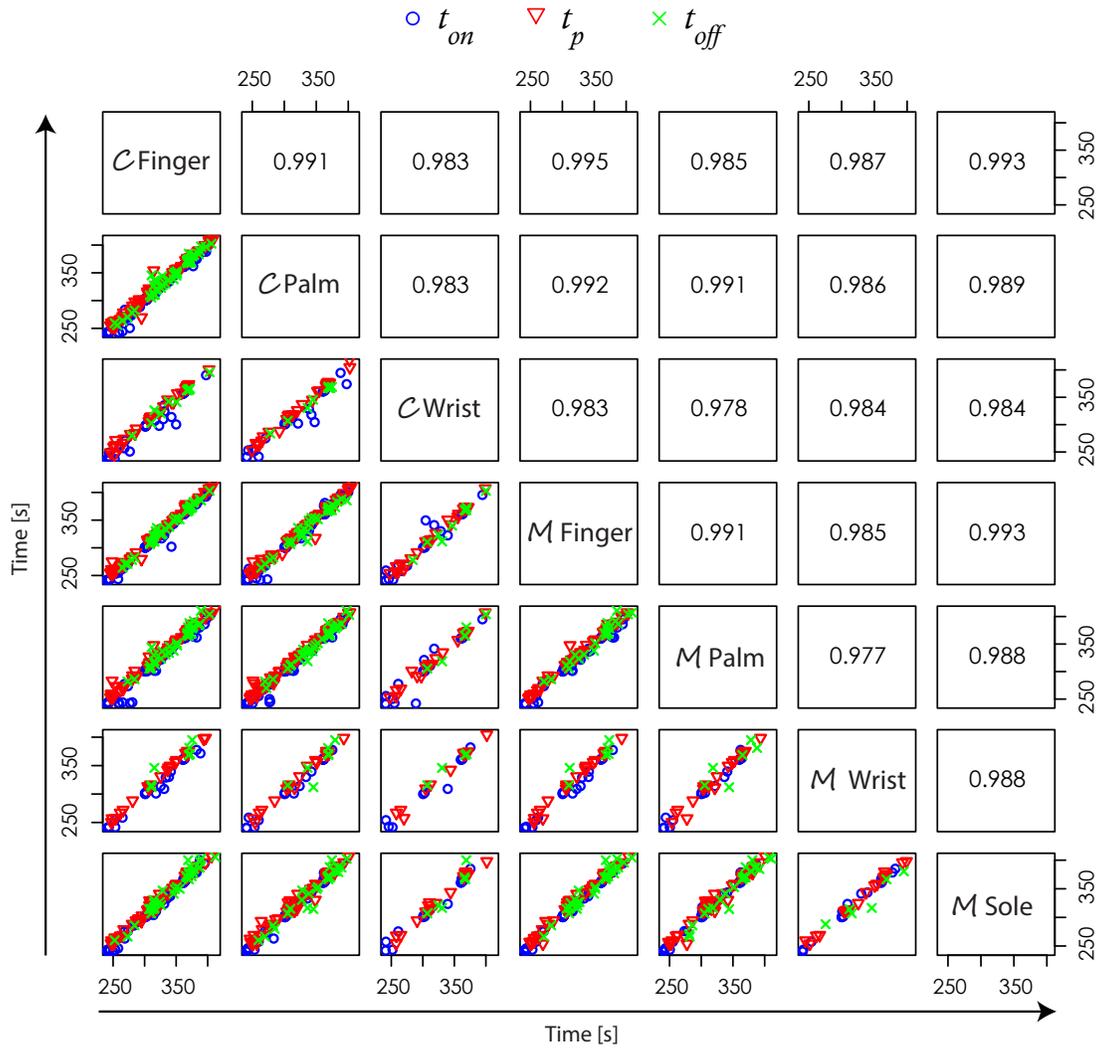


Figure 7.7: Arousal timing relationships between nodes. **BELOW THE DIAGONAL:** Scatterplots of Onset (t_{on}), Peak, (t_p), and Offset (t_{off}) times for the various node pairs. **ABOVE THE DIAGONAL:** The correlation coefficients that correspond to the strength of the linear relationships depicted in the scatter-plots below the diagonal (all are significant, $p < 0.01$).

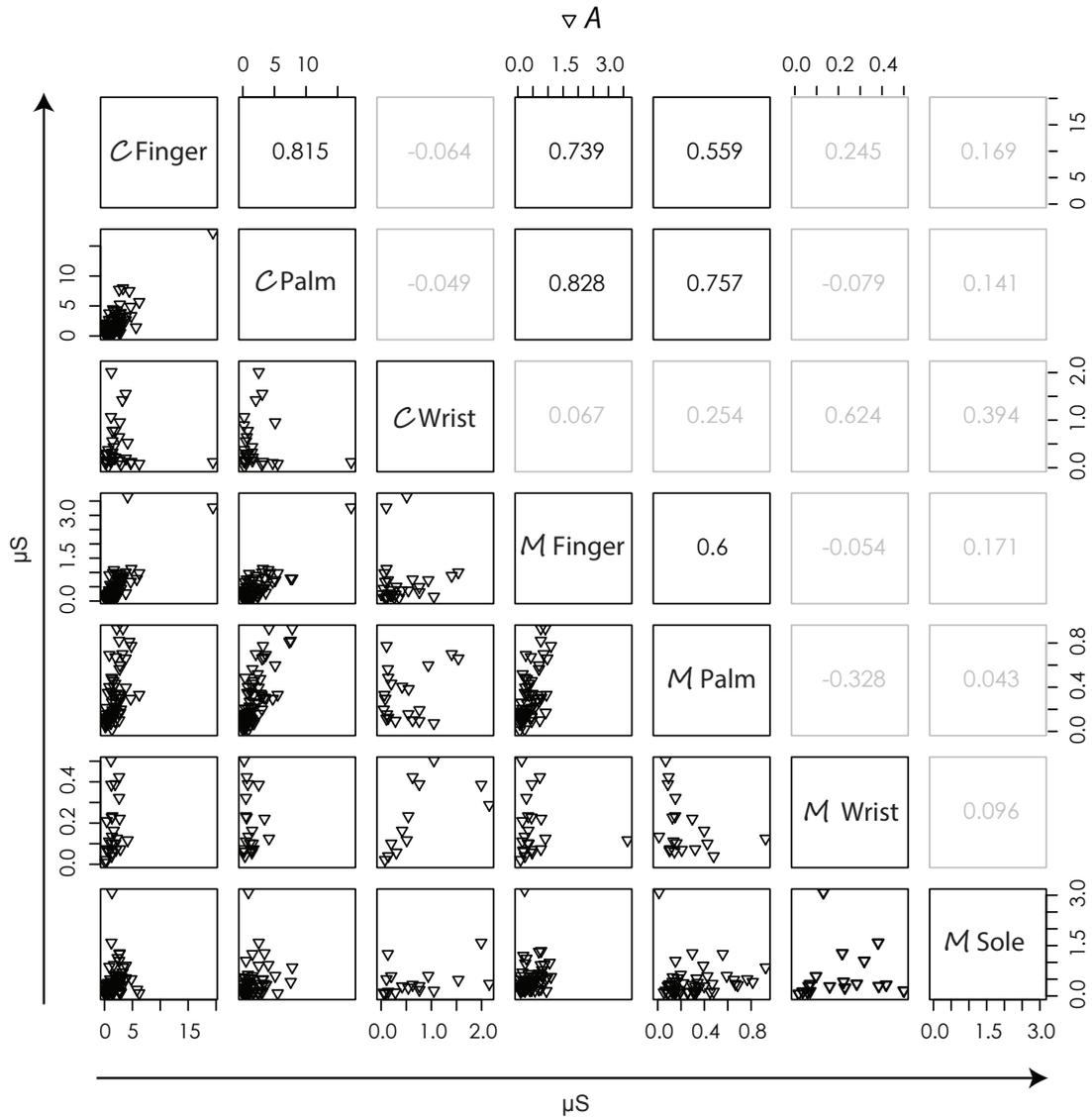


Figure 7.8: Arousal intensity relationships between nodes. **BELOW THE DIAGONAL**: Scatterplots of amplitude (A) for the various node pairs. **ABOVE THE DIAGONAL**: The correlation coefficients that correspond to the strength of the linear relationships depicted in the scatter-plots below the diagonal (coefficients in bold are significant, $p < 0.01$).

Bibliography

- [1] <http://www.zephyranywhere.com/products/bioharness-3/>.
- [2] <http://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/trait-state.aspx>.
- [3] <http://www2.psychology.uiowa.edu/faculty/Watson/PANAS-X.pdf>.
- [4] <http://subjectbook.times.uh.edu>.
- [5] W. Boucsein. *Electrodermal Activity*. Springer, New York, NY, USA, 2012.
- [6] T. Chaspari, D. Bone, J. Gibson, C. Lee, and S. Narayanan. Using physiology and language cues for modeling verbal response latencies of children with asd. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 37023706, Buenos Aires, Argentina, 2013.
- [7] M. Dcosta, D. Shastri, and I. Pavlidis. Perinasal indicators of malevolence. In *Proceedings of the 2015 IEEE International Conference on Automatic Face and Gesture Recognition (FG2015)*.
- [8] M. Dcosta, D. Shastri, R. Vilalta, J. Burgoon, and I. Pavlidis. Perinasal indicators of deceptive behavior. In *Proceedings of the 2015 IEEE International Conference on Automatic Face and Gesture Recognition (FG2015)*.
- [9] D. Fowles, M. Christie, R. Edleberg, W. Grings, D. Lykken, and P. Venables. Publication recommendations for electrodermal measurements. *Psychophysiology*, 18(3):232–239, 1981.
- [10] A. S. Horn, J. T. Coyle, and S. H. Snyder. Catecholamine uptake by synaptosomes from rat brain structure-activity relationships of drugs with differential effects on dopamine and norepinephrine neurons. *Molecular Pharmacology*, 7(3):66–80, 1971.

- [11] E. Jettmar and N. C. Adaptive testing: Effects on user performance. In *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
- [12] T. Kamei, T. Tsuda, S. Kitagawa, K. Naitoh, Koji, and T. Ohhashi. Physical stimuli and emotional stress-induced sweat secretions in the human palm and forehead. *Analytica Chimica Acta*, 365(1):319–326, 1998.
- [13] P. Lang, M. Bradley, and B. Cuthbert. Emotion, attention, and the startle reflex. *Psychological review*, 97(3):377–395, 1990.
- [14] D. T. Lykken. The gsr in the detection of guilt. *Journal of Applied Psychology*, 43(6):385–388, 1959.
- [15] D. Mcduff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski. Affect aura an intelligent system for emotional memory, 2012.
- [16] R. S. Neary and M. Zuckerman. Sensation seeking, trait and state anxiety, and the electrodermal orienting response. *Psychophysiology*, 13:205–211, 1976.
- [17] I. Pavlidis and J. Levine. Thermal image analysis for polygraph testing. *IEEE Engineering in Medicine and Biology Magazine*, 21(6):56–64, 2002.
- [18] I. Pavlidis, P. Tsiamyrtzis, D. Shastri, A. Wesley, Y. Zhou, P. Lindner, P. Buddhharaju, R. Joseph, A. Mandapati, B. Dunkin, and B. Bass. Fast by nature - how stress patterns define human experience and performance in dexterous tasks. *Scientific Reports*, 2, 2012.
- [19] M. Poh, T. Loddenkemper, N. Swenson, S. Goyal, J. Madsen, and R. Picard. Continuous monitoring of electrodermal activity during epileptic seizures using a wearable sensor. In *In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4415–4418, Buenos Aires, Argentina, 2010.
- [20] M.-Z. Poh, N. Swenson, and R. Picard. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, 57(5):1243 – 1252, 2010.
- [21] T. Rocklin, A. O’Donnell, and P. Holst. Effects and underlying mechanisms of self-adapted testing. *of Educational Psychology*, 87(1):103–116, 1995.
- [22] C. Sandi. Stress and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4:245–261, 2013.

- [23] T. Sandi, C. Pinelo-Nava. Behavioral effects and neurobiological mechanisms. *Neural Plast*, 2007.
- [24] A. Sano and R. Picard. Toward a taxonomy of autonomic sleep patterns with electrodermal activity. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 777–780, Boston, Massachusetts, 2011.
- [25] D. Shastri, A. Merla, P. Tsiamyrtzis, and I. Pavlidis. Imaging facial signs of neurophysiological responses. *IEEE Transactions on Biomedical Engineering*, 56(2):477–84, 2009.
- [26] I. Shastri, Papadakis, P. M. Tsiamyrtzis, B. Bass, and I. Pavlidis. Perinasal imaging of physiological stress and its affective potential. *IEEE Transactions on Affective Computing*, 3(3):366–378, 2012.
- [27] C. D. Spielberger. *State-Trait Anxiety Inventory*. Wiley Online Library, 2010.
- [28] S. Taamneh, D. Shastri, C. Currie, M. Dcosta, and I. Pavlidis. What sympathetic responses can tell about children’s performance in reading? In *Society for Affective Science Conference*, 2015.
- [29] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. T. Pavlidis, M. G. Frank, and P. Ekman. Imaging facial physiology for the detection of deceit. *International Journal of Computer Vision*, 71(2):197, 2006.
- [30] I. Uyanik, P. Lindner, P. Tsiamyrtzis, D. Shah, N. V. Tsekos, and I. T. Pavlidis. Applying a level set method for resolving physiologic motions in free-breathing and non-gated cardiac mri. In *Functional Imaging and Modeling of the Heart*, pages 466–473. Springer, 2013.
- [31] A. Wesley, D. Shastri, and I. Pavlidis. A novel method to monitor driver’s distractions. In *CHI ’10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA ’10, pages 4273–4278, New York, NY, USA, 2010. ACM.
- [32] R. Yerkes and J. Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18:459–482, 1908.
- [33] Y. Zhou, P. Tsiamyrtzis, P. Lindner, I. Timofeyev, and I. Pavlidis. Spatio-temporal smoothing as a basis for facial tissue tracking in thermal imaging. *IEEE Transactions on Biomedical Engineering*, 60(5):1280–1289, 2013.