

EUSTRESSED OR DISTRESSED?
COMBINING PHYSIOLOGY WITH OBSERVATION
IN STRESS STUDIES

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
Avinash Wesley
December 2015

EUSTRESSED OR DISTRESSED?
COMBINING PHYSIOLOGY WITH OBSERVATION
IN STRESS STUDIES

Avinash Wesley

APPROVED:

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

Dr. Christoph Eick
Dept. of Computer Science, UH

Dr. Ricardo Vilata
Dept. of Computer Science, UH

Dr. Nikolaos V. Tsekos
Dept. of Computer Science, UH

Dr. Carla Sharp
Dept. of Psychology, UH

Dean, College of Natural Sciences and Mathematics

Acknowledgements

This research was supported by grants from the National Science Foundation (NSF) under Grant No. ISS-0812526 entitled ‘Do Nintendo Surgeons Defy Stress’, National Institute of Standards (NIST) entitled ‘Multi-Spectral Imaging for the Simultaneous Detection of Stress and Concealed Objects and Department of Defense (DOD) entitled ‘Spectral Imaging Sensor for Improved Biometric and Human Intent Analysis II. The views expressed in this work do not necessarily reflect those of the funding agencies.

Foremost, I must express my appreciation to Dr. Ioannis Pavlidis, whose orderly mind, consistency, and systematic approach helped me present this material in the clearest and most organized fashion. I would also like to thank Dr. Dvijesh Shastri for always being available with his time, his thoughts, and his help. My sincere gratitude also goes to Dr. Panagiotis Tsiamyrtzis for his valuable review of the dissertation. I am especially thankful to our previous lab manager Dr. Peggy Linder, also a skilled IT specialist, who contributed on numerous levels to my research.

I am very grateful to the Church in the Center (CITC) community and International Christian Fellowship (ICF). Their love and support has been invaluable.

And finally, but most of all, I want to acknowledge my beautiful wife — Phoebe Joy. Her constant encouragement, prayers, and support has enabled me to achieve

my doctorate. My dissertation is dedicated to her and my two sons Jedidiah and Emmanuel. They bring immense joy to my life.

**EUSTRESSED OR DISTRESSED?
COMBINING PHYSIOLOGY WITH OBSERVATION
IN STRESS STUDIES**

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
Avinash Wesley
December 2015

Abstract

In this dissertation, I describe a novel method to conduct stress studies via the combination of a physiological and an observational information channel. The method enables the quantification of aroused emotional states and their disambiguation into positive or negative instances. The physiological channel targets sympathetic responses and is materialized as a perspiratory signal extracted from thermal imagery of the perinasal area. The observational channel is materialized via decoding of facial expressions. Decoding is usually performed in the visible spectrum, however I have developed an algorithm to carry this out using thermal imagery instead. Thus, thermal imaging is used for both physiological and observational analysis. The potential of this dual-unobtrusive methodology was demonstrated with two stress studies. The first study was about surgeons' interaction with laparoscopic training boxes — representative of the dexterous genre. The second study was about operator overloading where the participants played a car driving game while being interrupted by phone calls and text messages.

Table of Contents

Table of Contents	vii
List of Figures	ix
List of Tables	xiii
1 Introduction	1
2 Background	5
2.1 Thermal Imaging	5
2.2 Emotion Classification	9
2.2.1 Basic Emotions	10
2.2.2 Circumplex Model of Emotions	11
2.2.3 Challenges and pitfalls of the existing emotion theories	11
2.2.4 Novel Stress Model Proposed	12
3 Facial Expression Recognition in Thermal Imagery	14
3.1 Benchmarking Thermal Imagery for Facial Expression Recognition	17
3.2 Thermal Action Coding System — TACS	31
3.3 System Architecture	34
4 Conducting Stress Studies	37
4.1 Stress Study I — Surgical Training	42

4.2	Stress Study II — Operator Overloading	46
5	Results	53
5.1	TACS Validation Results	53
5.2	Stress Analysis for the Two Stress Studies	54
5.2.1	Study Variables	54
5.2.2	Stress Analysis on Stress Study I — Surgical Training	57
5.2.3	Stress Analysis on Stress Study II — Operator Overloading	65
6	Conclusion	76
	Bibliography	78

List of Figures

1.1	The actual version of Yerkes-Dodson law based on the findings and theorizing of Yerkes and Dodson	2
1.2	Perspiration signals for Eustress (a) and Distress (b)	3
2.1	The electromagnetic spectrum. The mid-wave infrared spectrum is highlighted by the arrow.	6
2.2	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
2.3	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.	8
2.4	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.	9
2.5	Six basic emotions	10
2.6	Circumplex model of emotions	12
2.7	Thermal Affective Stress Model – TASM	13
3.1	The experimental setup used to simultaneously collect thermal and visual facial data from subjects.	19
3.2	The 13 regions of interest used to capture facial movement and deformation.	20

3.3	PCA values from ROIs 3 (depicted by dotted lines) and 9 (depicted by solid lines) while the subject is making (a) angry expression, AU 4 (Brow lowerer) and (b) happy expression, AU12 (Lip corner puller). It can be clearly seen that PCA values are larger for ROI 3 during angry expression, while it has large values for ROI 9 during happy expression.	24
3.4	(a) A sample from Phase I (illumination variance) dataset. The top row shows the thermal and visual images acquired while lighting was from subject's right, second row shows corresponding images from another subject when the lighting was from the top. (b) A sample from Phase II (temperature variance) dataset. The top row shows the thermal and visual images acquired while air is blown from subject's right, second row shows corresponding images from another subject while air is blown from the front. The bottom row shows the color map used for thermal images.	26
3.5	The seven ROIs used to capture facial muscle movement and deformation	32
3.6	Feature vector formation for expression AU 1+2 (Inner+ Outer Eyebrow Raise)	33
3.7	System architecture for the stress model.	36
4.1	Sensing instruments	39
4.2	(a) First Image Row: Virtual tissue tracker (black rectangle) at work as a subject exhibits head motion during execution of Task 3 in Stress Study I. Second Image Row: Motion-corrected perinasal area snapshots throughout the performance period. The end effect could be considered virtual tethering of a virtual probe, so that the measurement area remains as consistent as in conventional measurements with tethered physical probes. (b1) Thermal image of the face. Spots S1, S2, and S3 are 'cold' spots indicative of perspiration. (b2) 3D thermal plot of the area surrounding perspiration spots S1, S2, and S3. The conic shape of the spot profiles denotes the gradual transition from a 'cold' core to a 'hot' surrounding background. (b3) Outcome of the morphological extraction algorithm as 3D energy plot.	41

4.3	Surgeon imaged thermally and visually during trial execution in the inanimate-laparoscopic skills lab of the Methodist Institute for Technology, Innovation, and Education (MITTIE SM). The panels to the right of the figure show details inside the surgical training box during the execution of the three different tasks.	44
4.4	Experimental setup for the driving simulation game.	49
4.5	Experimental protocol for operator overloading study.	52
5.1	The gray 1D signal is the perspiration signal. The horizontal red and green bars at the bottom are the annotated time intervals for negative and positive facial expressions. With the aid of observations, the perspiration signal can be disambiguated into distress or eustress signal.	58
5.2	A bout of eustress for an experienced surgeon (D005), as indicated by the locally elevated perinasal signal and the positive facial expressions.	59
5.3	Bouts of distress for a novice surgeon (D025), as indicated by the fluctuating perinasal signal and the negative facial expressions.	60
5.4	Distribution per skill level of mean perspiration intensity $\mu_E(x)$ ($\ln(\cdot)$ transformation to comply with the ANOVA assumptions). $\ln(\mu_E(x))$ is -6.0626 ± 0.6926 (s.d.) for novices and -6.8614 ± 0.6319 (s.d.) for expert surgeons.	61
5.5	(a) Distribution per skill level of mean stress responses during expressed negative feelings $\mu_{EN}(x)$ ($\sqrt{\cdot}$ transformation to comply with the ANOVA assumptions). $\sqrt{\mu_{EN}(x)}$ is 0.0447 ± 0.0204 (s.d.) for novices and 0.0205 ± 0.0169 (s.d.) for experienced surgeons. (b) Distribution per skill level of mean distress extent $\mu_N(x)$ ($\sqrt{\cdot}$ transformation to comply with the ANOVA assumptions). $\sqrt{\mu_N(x)}$ is 0.6457 ± 0.2279 (s.d.) for novices and 0.3963 ± 0.2928 (s.d.) for experienced surgeons.	63
5.6	The stress model representing distress for all the trials of all the surgeons. The red points are distress levels for the novice surgeons and the red points are for the expert surgeons. The bigger solid points represents an average distress level for the each surgeon level.	64

5.8	Distribution of extent of expressed positive feelings $\mu_P(x)$ during tasking per skill level. For novices $\sqrt{\mu_P(x)}$ is 0.0682 ± 0.1938 (s.d.) while for experienced surgeons is 0.1316 ± 0.2946 (s.d.). We used $\sqrt{\cdot}$ transformation to comply with ANOVA assumptions.	65
5.7	Experienced surgeon's (subject ID: D001) thermo-physiological (perinasal) and observational (facial) images during execution of Task 3, Session 4, Trial 2. The corresponding perspiration (stress) signal is shown in the middle. There is an elevation in the signal due to some excitation, about 30 sec into the execution of the trial. This excitation is positive, as the FACS-decoding [12] of facial expressions indicates along the timeline (bottom). This response was due to an error committed by the surgeon (needle drop), from which he quickly recovered.	66
5.9	Protocol for the operator overloading. There are three sessions — Baseline, Cellphone, and Texting. Cellphone and Texting sessions are further divided into three segments — solo game play, interrupted game play, and post solo game play.	68
5.10	Box-plots for the overall perspiration response for all the subjects across the segments. We used the $\ln(\cdot)$ transformations to comply with the analysis of variance assumptions.	71
5.11	Box-plots denoting the distress intensity and distress extent for the multi-tasking segments. We used the $\sqrt{\cdot}$ transformations to comply with the analysis of variance assumptions.	72
5.12	The stress-model representing distress for all the subjects in the multi-tasking segments. The red points are distress levels for the phone segment and the red points are for the texting segment. The bigger solid points represent an average distress level for each segment. . . .	73
5.13	Box-plots for errors committed in the multi-tasking segments. We used the $\sqrt{\cdot}$ transformations to comply with the analysis of variance assumptions.	74
5.14	The plot depicts the percentage of the negative and positive facial expressions made by the participants in each study. Green region represents the positive facial expression percentage and the red region represents the negative facial expression percentage.	75

List of Tables

3.1	Confusion matrix for facial expressions from deception study dataset [8]. CERT software was used for facial expression recognition in the visual imagery.	16
3.2	General classification of the 8 selected action units.	18
3.3	Regions of interest based on the facial anatomy.	22
3.4	Confusion matrices of Phase I (illumination variance) and Phase II (temperature variance) experiments; for thermal and visual modalities, and their fusion	28
3.5	Accuracy of Phase I (illumination variance) and Phase II (temperature variance) experiments for thermal and visual modalities.	29
3.6	Dynamic trackers placement based on facial anatomy.	35
4.1	Stress Studies designed for different types of stressors.	38
4.2	Make-up of surgeon pool.	43
5.1	Confusion matrix for facial expressions from Stress Study I — Surgical Training.	55
5.2	Confusion matrix for facial expressions from Stress Study II — Operator Overloading.	55
5.3	Confusion matrix for facial expressions from an extra dataset from the deception study conducted by the CPL lab in the University of Houston.	56
5.4	Confusion matrix for facial expressions from the two studies and the extra dataset.	56

Chapter 1

Introduction

Stress (defined here as physiological arousal) is an ever-present mechanism that helps humans cope with perceived or real threats or challenges. Upon stimulus, the sympathetic nervous system signals the adrenergic and the cholinergic receptors. The former result in the elevation of cardiovascular operations, while the latter result in activation of sweat glands on the fingers, perinasal area, and other peripheral parts of the body. We focus on the cholinergic receptors as they are more dynamic.

Stress has been suspected to play a key role in human performance for quite some time [22]. In the early 20th century, Yerkes and Dodson postulated their famous hypothesis regarding the relation of stress to human performance [44]. Figure 1.1, taken from [9] indicates the two curves representing two kinds of tasks — simple and difficult. For simple or well-learned tasks, the relationship can be considered linear with

improvements in performance as arousal increases. For complex, unfamiliar, or difficult tasks, the relationship between arousal and performance becomes inverse, with declines in performance as arousal increases. For such tasks, performance increases with stress up to a point and decreases past that — a relationship that proved to be true in several experimental studies.

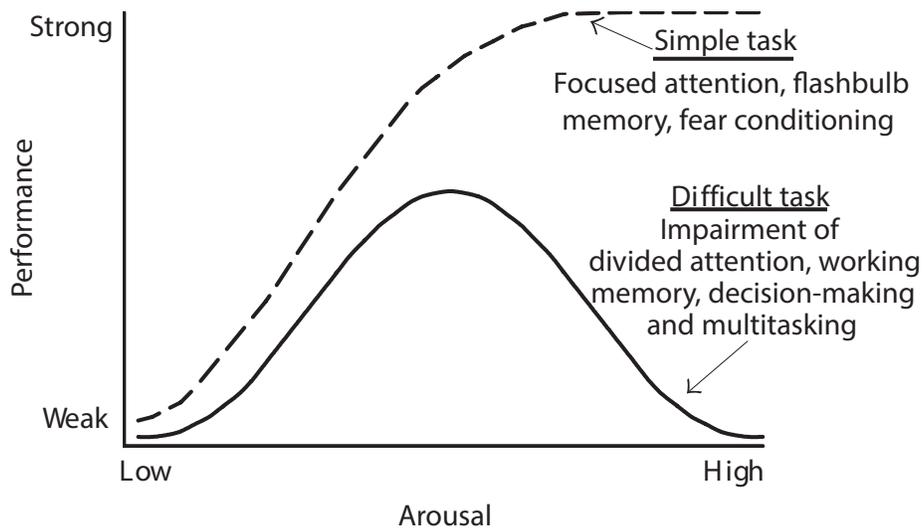


Figure 1.1: The actual version of Yerkes-Dodson law based on the findings and theorizing of Yerkes and Dodson

So stress monitoring is important for finding the optimal human performance envelope. To our advantage, stress is ubiquitous due to its physiological manifestation [37]. Previous work demonstrated that during emotional arousal physiological signs materialize on the face, such as increased blood flow in the peri-orbital area [23]

and transient perspiration on the perinasal area [29]. These signs have thermophysiological footprints and quantification methods have been proposed based on thermal imaging in [35] and [29]. Between the periorbital and perinasal signal, the latter is of particular interest to this work, because it is part of a cluster of sympathetic responses on sensory organs (tactile and olfactory) that are closely related to emotions [29]. Because the perinasal response is sympathetic in nature, it is non-specific to negative or positive arousal.

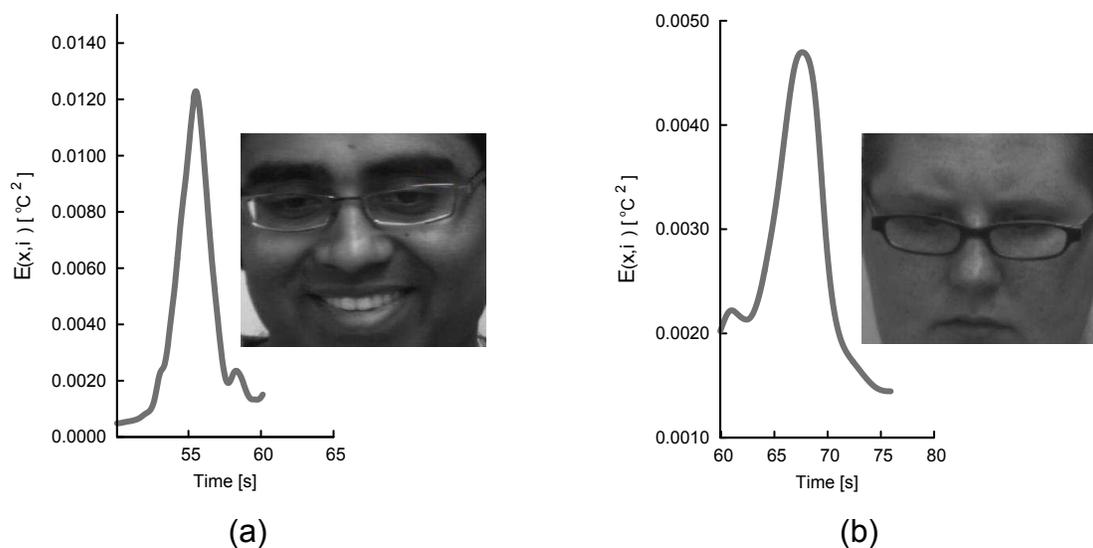


Figure 1.2: Perspiration signals for Eustress (a) and Distress (b)

Figure 1.2 depicts two different types of arousal — Eustress (positive stress) and Distress (negative stress). The physiological response is similar for these two different types of arousal. Physiological measurement (perinasal perspiration here) is an invaluable channel for stress study but one that is nonspecific. It meets the implicit measurement objective but not the greater detail objective. With respect

to stress, it provides a good measure of activation, but it does not provide a measure of valence. This shortcoming of the physiological response can be addressed by complementing it with a new dimension of the observation of facial expressions. Facial expressions are formed through coordinated muscle actions and can be classified using the Facial Action Coding System (FACS) [12]. FACS breaks down the development of expressions into sets of basic units. It was designed to measure visible facial behavior in any context, not just in emotions, and has become the gold standard for facial-measurement systems.

In my dissertation, I used the perspiration extraction method described in [29] to derive the perinasal signals. In addition, I have developed a new method to decode facial expressions in thermal modality rather than visual imagery. Hence, both physiological and observational analysis can be carried out under a single imaging modality.

The remainder of the dissertation features a thermal imaging method for facial expression recognition (TACS — Thermal Action Coding System). It is followed by a description of two field studies of stress. The objective of these stress studies was to investigate the role of stress on performance in as realistic conditions as possible and as objectively as possible. I then present the accuracy results of TACS on these field studies. Also, the collected neurophysiological data from these studies were analyzed by the proposed tandem approach. The stress analysis results brim with intriguing leads about human nature under stress.

Chapter 2

Background

2.1 Thermal Imaging

Our research mainly focuses on facial physiology 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 2.2). 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.

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

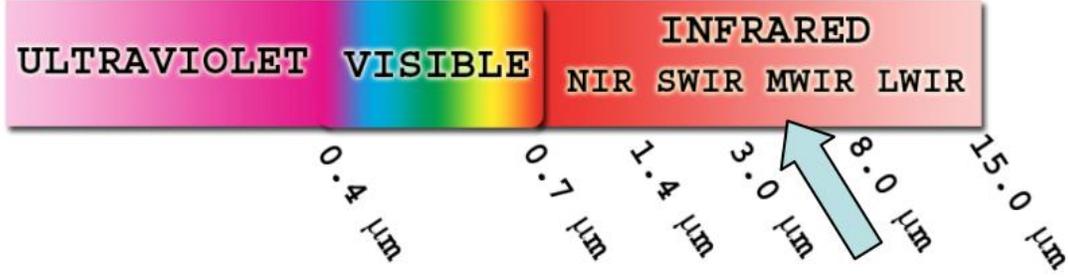


Figure 2.1: The electromagnetic spectrum. The mid-wave infrared spectrum is highlighted by the arrow.

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

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 2.2).

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.

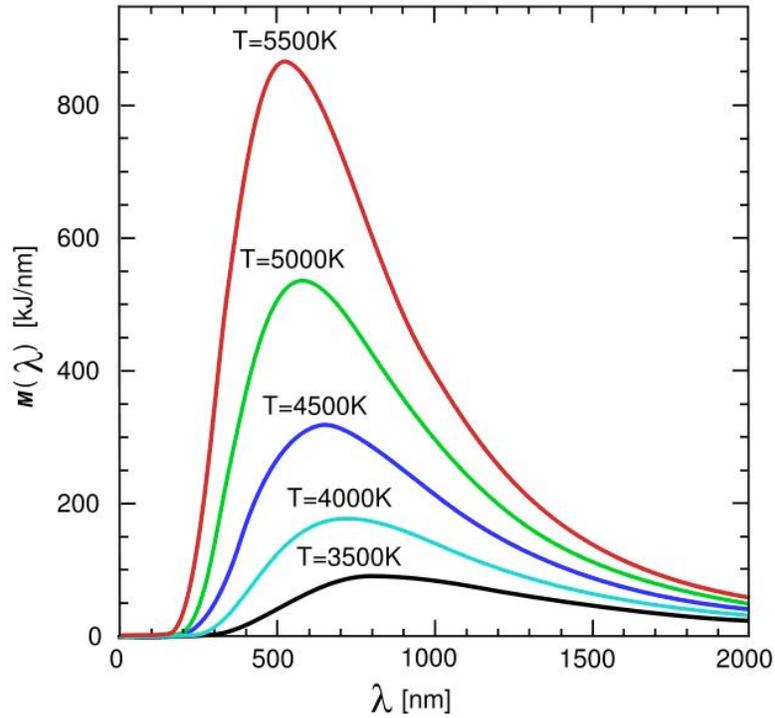


Figure 2.2: Black body spectrum. The graph shows a significant amount of energy radiation change in the thermal infrared band (300-1400 nm) due to change in temperature.

Thermal data collection in our research has been accomplished by ATHEMOS (Automatic THERmal MONitoring System), which we have developed in-house (see figure 2.3).

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



Figure 2.3: 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.

disk. The raw image must be mapped to useful color values in order to visualize it easily.

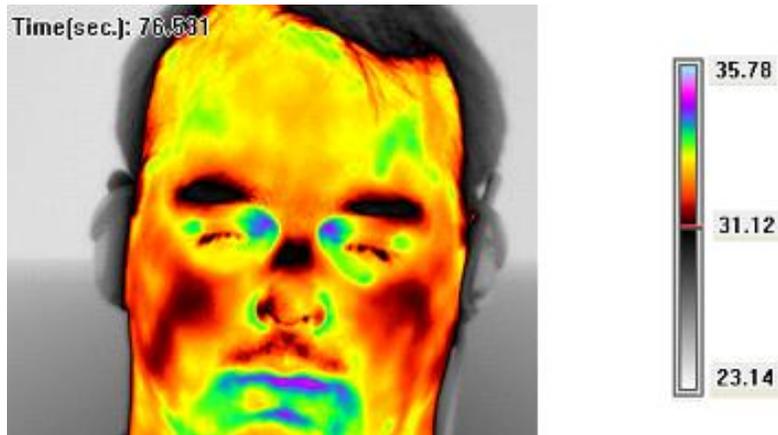


Figure 2.4: Thermal image. A sample thermal image of a subject. The bar on the right side shows the mapping between color and temperature in degrees centigrade.

2.2 Emotion Classification

When it comes to emotions/affect classification, there are two major theories, viz. Basic Emotions and Circumplex Model.

2.2.1 Basic Emotions



Figure 2.5: Six basic emotions

The basic emotions theory, as the name suggests, posits that humans are evolutionarily endowed with a discrete and restricted set of basic emotions [10]. The basic emotions are — anger, fear, disgust, surprise, happiness, and sadness. The theory states that each of these basic emotions is independent of the others and has a distinct neurophysiological signature. Each emotion is unique in its behavioral, psychological, and physiological manifestations. Each emotion is associated with a characteristic facial expression.

2.2.2 Circumplex Model of Emotions

The Circumplex Model is a dimensional model that regards emotions as a continuum of highly interrelated and often ambiguous states. The two dimensions are valence and arousal. Each and every emotion is the consequence of a linear combination of these two independent dimensions. (see figure 2.6)

2.2.3 Challenges and pitfalls of the existing emotion theories

Basic emotion theory has primarily relied on the expressive manifestation of emotions rather than the core physiological bases of emotions. Existing theories of emotions (including the above two) lack a quantitative model to describe a persons affective state. One of the major contributors to this shortcoming is the subjective evaluation of emotions by the participants. These subjective self-reports are usually acquired through the use of rating scales, or some form of imagery such as cartoon images or photographs of human faces. These self-reports rely on a persons introspective abilities to imagine or recall emotionally laden experiences. Research has shown that such self reports can often be inaccurate simply because of the inability of an individual to access their own emotions. Nisbett in his research [26] reported that humans do not have direct introspective access to many (if not most) of their mental processes. This suggests that emotions are not altogether distinct and independent from each other.

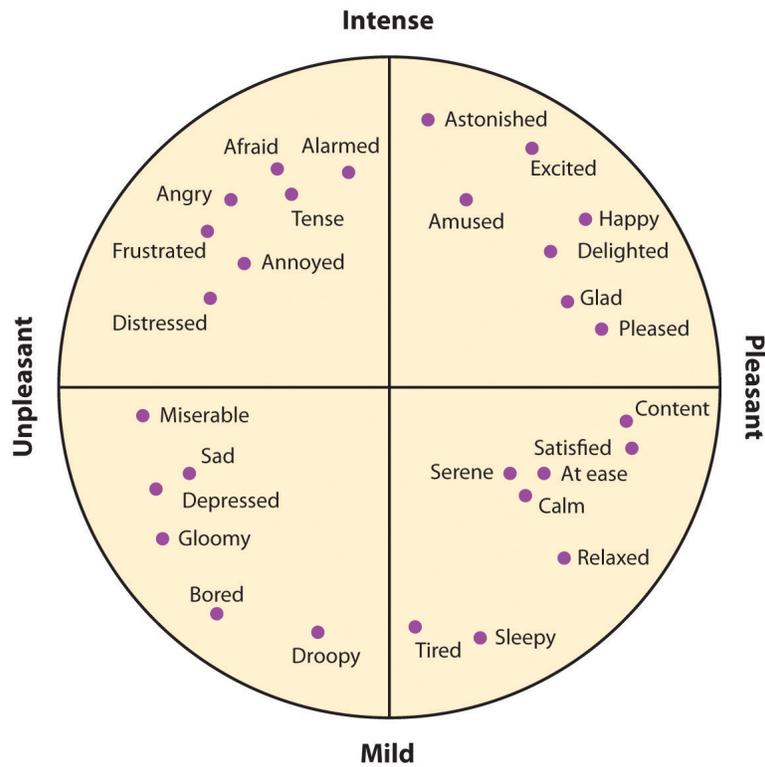


Figure 2.6: Circumplex model of emotions

2.2.4 Novel Stress Model Proposed

I have proposed a new stress model — Thermal Affective Stress Model (TASM). It derives its structure from the circumplex model where the dimensions are arousal and valence. Arousal is proxied by the quantitative measure of physiology (perspiration in our research) and the valence is measured by the quantitative measure of affective display (facial expressions extent in our research). Facial expression extent is the

portion of the frames where the subject makes an affective display. It is described as the percentage of the total frames.

Even though the proposed stress model takes the dimensionality aspect of emotions from the circumplex model, but unlike the latter, it is quantitative in its evaluation (see figure 2.7). The perspiration extraction method was developed by Dr. Shastri [30]. A brief description of the method is in the Stress Study chapter. I have developed facial expression recognition software using thermal imagery. The details for it are mentioned in the following chapter.

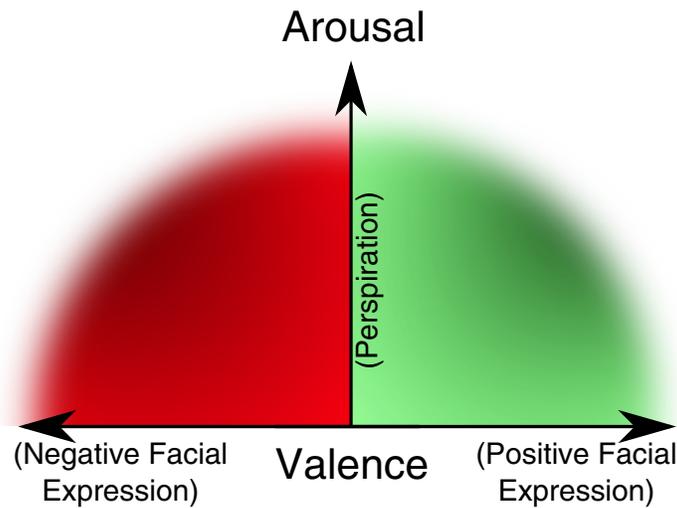


Figure 2.7: Thermal Affective Stress Model – TASM

Chapter 3

Facial Expression Recognition in Thermal Imagery

The detection and recognition of human facial expressions is a challenging task. Among different individuals the geometry, size, and color of the face vary greatly. Furthermore, a single expression can be formed at many different intensities and speeds, sometimes so subtle that it goes unnoticed to a human observer. This intense variance compounded with the subtlety of expressions necessitates more detailed and automated approaches to facial expression detection.

Visual cameras are most commonly used to capture facial data due to their low cost and ubiquitous availability. Several automated facial expression recognition algorithms were proposed in the recent years from visual imagery [14], [27], [6].

Bartlett et. al reported 93% accuracy of automated facial recognition on the Cohn-Kanade expression dataset [5], and recently Kotsia and Pitas reported classification accuracy of 99.7% and 95.1% on the same dataset [21]. Visual approaches, while shown to be quite effective on particular databases, have a few unaddressed obstacles. A major drawback is their tendency to lose accuracy when classifying subjects of darker skin tones. The OpenCV face detection system, which has become a basis for comparison shows a significant disparity in the accuracy of classifying dark- versus light-skinned subjects [41]. Furthermore, many databases used to test visual-based expression recognition systems, have a narrow variety of positions, textures, and intensities of light. This usually simplifies the task of classification and result in higher accuracy measurements. Hence, visual approaches tend to perform well under sterile lab conditions, but under varied light conditions, they may operate at lower accuracies [41]. This is one of the major reasons why most market available softwares perform sub-optimally in a free flowing experiment where participants make constant motion creating variations in reflectance. To illustrate my point, I processed visual videos from a dataset [8] using CERT [24] software. The classification accuracy was 20.45% on this dataset even though the author claims to have achieved 80% accuracy on a spontaneous facial expression dataset.

Thermal imaging is a well known alternative to visual imagery because of its illumination invariance [31]. A thermal camera measures the radiations emitted from the surface of the skin, which is a result of heat dissipation from the body's core due to blood flow, metabolic activities, subcutaneous tissue structure, and the

	Test	Classified AUs					No Detection
	AUs	1+2	4	9	10	12	NaN
	1+2	11	2	16	0	5	31
	4	0	6	20	0	3	17
	9	0	0	2	0	0	5
	10	0	0	0	0	0	3
	12	0	0	9	1	28	26

Table 3.1: Confusion matrix for facial expressions from deception study dataset [8]. CERT software was used for facial expression recognition in the visual imagery.

sympathetic nervous activities. Though study has been done in the area of thermal face recognition [20], few have attempted to explore facial expression recognition using this modality. Khan et al. explored and proved through statistical analysis, the feasibility of automated facial expression classification through thermal imaging [19]. Yoshitomi et al. reported success rates of 90% [45]. An unsupervised local and global feature localization algorithm for facial expression classification was proposed by Trujilo [34].

Thermal imagery stands as an alternative which may potentially eliminate the illumination problem encountered in the visual imagery. However, it may pose a possible challenge as facial thermograms may change depending on ambient temperature and the physical condition of the subject. This would render difficult, the task

of acquiring similar features for the same expressions.

In the next section of the dissertation, I will describe an experiment conducted to study the sensitivity of visual and thermal imagery to illumination and airflow variations, thus benchmarking thermal imagery against visual.

3.1 Benchmarking Thermal Imagery for Facial Expression Recognition

This was a feasibility study where I collected simultaneously visual and thermal imagery of participants in conditions of illumination and airflow variations. Spatial PCA was applied to the 13 fiducial regions of the human face recordings in both the thermal and visual imagery. I will present comparative results from both modalities. To the best of our knowledge, this is the first time such a comparative study is being reported. This work was published in International Symposium on Visual Computing (ISVC) [39].

FACS, developed by psychologists Ekman and Friesen [11], is most commonly used to classify human facial expressions through analysis of possible contortions of facial geometry. FACS breaks down the development of expressions into particular action units, each of which is derived from a muscle or muscle groups in the head. In this study, I classified a total of 8 action units (AU 1+2, 4, 6+12, 9, 10, 12, 15, 17). I selected these specific action units because they are the exemplary when forming

Action Unit	Principal Emotion
1 + 2	Suprise
4	Anger and Fear
6 + 12	Happiness
9	Disgust
10	Disgust
12	Happiness
15	Sadness
17	Disgust and Confusion

Table 3.2: General classification of the 8 selected action units.

any of the 6 universal emotions (see Table 3.1) [13].

Our automated facial expression recognition algorithm mainly contains three steps - face acquisition, facial feature extraction, and expression classification. In this section, I will explain in detail our experimental setup to collect simultaneous visual and thermal facial data, local facial feature extraction algorithm, and expression classification methods.

Experimental Setup. A snapshot of our experimental setup can be seen in the figure 3.1. A total of 8 subjects participated in our experiments with age range from 20 to 30 years, both genders, and varying ethnicities. To facilitate comparison, I collected simultaneous data from both midwave-thermal infrared and a monochrome-CCD visual cameras as shown in figure 3.1. The room is equipped with three lighting

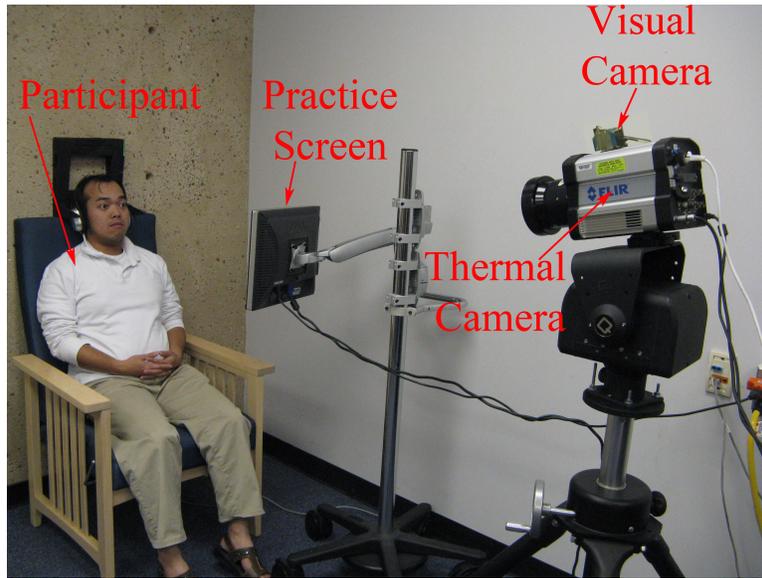


Figure 3.1: The experimental setup used to simultaneously collect thermal and visual facial data from subjects.

positions — subject’s right, top and front, to simulate the effect of illumination variation on visual imagery. We used a portable fan to simulate the effect of variable atmospheric air conditions on thermal imagery. The subjects were instructed to rinse their face and apply a small amount of 70% isopropyl alcohol. In order to ensure that the evaporation of the volatile alcohol mixture did not adversely affect the data, each subject waited a mandatory period of 15 minutes before beginning the data collection. A FACS encoder trained each subject regarding the facial expressions they were supposed to make during the data collection by showing them the videos of each expression. Each subject was allowed as much time as they needed to practice each expression, and they also have an option to skip any expression, if they so

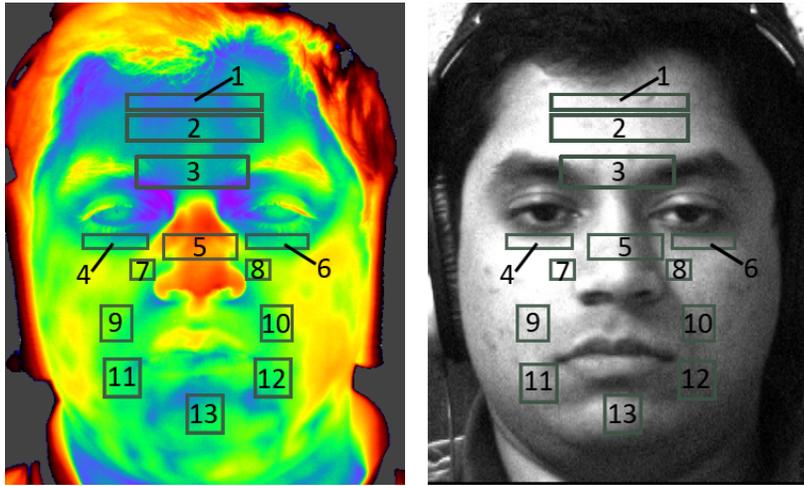


Figure 3.2: The 13 regions of interest used to capture facial movement and deformation.

desired. For each subject, I first recorded their relaxed and neutral expression for 25 seconds, followed by a visual instruction on a screen in front of them regarding the next expression they are supposed to make. The subjects were asked to repeat each expression 14 times at any intensity of their choice in order to simulate the variety of natural expression in everyday formulation.

Local Feature Extraction. The typical feature extraction algorithms in automated facial expression recognition can be categorized as holistic (where the face is processed as a whole), and local (where only facial features or areas that are prone to change with facial expressions are processed) [14]. Our feature extraction algorithm falls in the latter category with regions of interest (ROIs) placed at 13 fiducial points on the face (as shown in figure 3.2). The ROIs are carefully chosen according to various

facial muscles involved in different FACS action units as explained below:

ROIs 1 and 2: measures contraction of the frontalis muscles which raise the eyebrows. The raising of the eyebrows, present in FACS action units 1 and 2, are most commonly associated with expressions of surprise. The vertical placement of ROIs 1 and 2 distinguish between action unit combination 1+2 and 4. Action unit 4 affects mostly ROI 2 because the skin is only slightly stretched on the forehead, producing lower values in ROI 1.

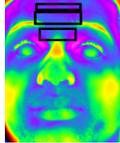
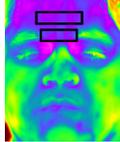
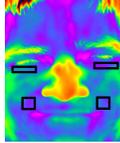
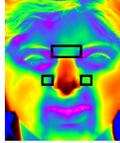
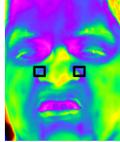
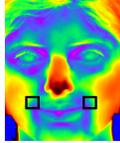
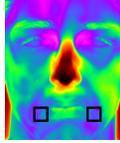
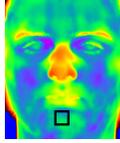
ROI 3: captures the translation of the tissue actuated by the corrugator and procerus muscles. These muscles are used to furrow the brow, action unit 4, when one is angry or sad. This ROI detects both the translation of the eyebrow and the deformation in the skin in between the eyebrows that generate signal.

ROIs 4 and 5: detects the orbicularis oculi. These are used to detect action unit 6, the critical difference between a Duchenne smile (AU 6 and 12) and a simple smile (AU 6). These ROIs detect the subtle raising of upper-cheek tissue and the wrinkling of the outer eye-edge.

ROI 6: measures the quadratus labii superioris which is responsible for scrunching the nose tissue. This is most commonly formed when a person is disgusted at something.

ROIs 7 and 8: additional measures to detect the lower set of elevator muscles, used to raise the tissue surrounding the nose. These attempt to measure action unit 10, a secondary expression of disgust. These measure the translation of new tissue

Table 3.3: Regions of interest based on the facial anatomy.

Example Image	Principal Emotions	Action Unit	Indicative ROIs	ROIs Placement
	Surprise	1 + 2	1, 2, 3	
	Anger and Fear	4	2, 3	
	Happiness	6+12	4, 6, 9, 10	
	Disgust	9	5, 7, 8	
	Disgust	10	7, 8	
	Happiness	12	9, 10	
	Sadness	15	11, 12	
	Disgust and Confusion	17	13	

from around the nose, just above the periorbital region.

ROIs 9 and 10: detects the contraction of the zygomaticus muscles, used most strongly in smiles. These measure action unit 12, the widening of the lips. These detect specifically the translation of cheek tissue as well as the crease formed at the edges of the mouth during a smile.

ROIs 11 and 12: measures the contraction of the triangularis, which lowers the other edges of the mouth into a frown. Action unit 15 is necessary for expressing sadness. These two ROIs measure both tissue translation and crease formation around the bottom edges of the mouth.

ROI 13: measures the change in the tissue attached to the mentalis. This allows for the measurement of any chin flexion, especially used to raise the lower lip.

The thermal and visual facial videos were recorded at 25 and 30 fps, respectively. We computed the neutral ROIs by computing the mean values in each ROI from first 25 seconds of the video, when the subjects made neutral expression. Then the principal components were computed for each ROI by treating each pixel within the ROI as a variable. The frames corresponding to greatest change from the neutral ROI will have the largest principal component values during the expression as depicted in figure 3.3.

After the principal components have been found for all ROIs, a profile for each expression is determined by computing the standard deviation of each ROI-principal component. To do this, we first annotate the onset (marking the start) and offset

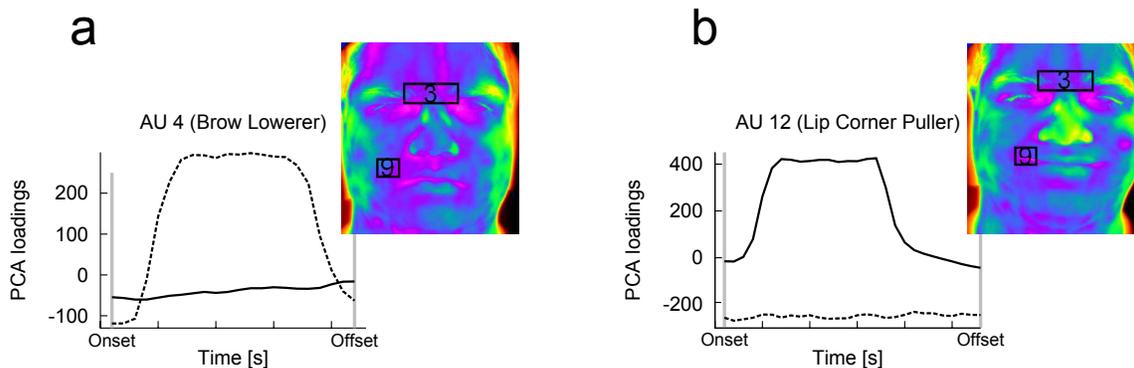


Figure 3.3: PCA values from ROIs 3 (depicted by dotted lines) and 9 (depicted by solid lines) while the subject is making (a) angry expression, AU 4 (Brow lowerer) and (b) happy expression, AU12 (Lip corner puller). It can be clearly seen that PCA values are larger for ROI 3 during angry expression, while it has large values for ROI 9 during happy expression.

(marking the end) frames for each expression as shown in figure 3.3. The standard deviation-expression profiles are generated by computing the standard deviation of each of the 13 ROIs during the window between the onset and offset. These expression profiles denote the amount of deviation found over the course of the expression, and hence are used to train the classifier.

Classification. The standard deviation-expression profiles computed in the feature extraction step are used to train feed-forward multilayer perceptrons [15] for both visual and thermal modalities. Each multilayer perceptron utilizes 14 input nodes, 10 sigmoid nodes in the hidden layer and 8 output nodes to classify expressions. Thermal and visual perceptron classifiers were generated separately by training them

with expressions that were coded by a certified FACs encoder to determine a ground truth.

Experimental Results and Discussion. In order to test the performance of each of the thermal and visual modalities during both ideal and challenging conditions, each subject was asked to participate in two sessions - Phase I and Phase II. In this section, I will present results from each of these sessions.

Phase I — Illumination Variance. In the first session (Phase I), I introduced variability in visual imagery by using different lighting positions (subject’s right, top and front) in the room for different subjects during the data collection. This resulted in a considerable variability in visual imagery (as shown in figure 3.4a) and hence posed a challenging condition for the visual perceptron classifier. However, the room temperature was maintained constant throughout the session, maintaining an ideal condition for thermal imagery. The Phase I dataset consisted of a total of 448 expressions from each of the thermal and visual modalities.

We used 10-fold cross validation and percentage split in order to test the classification accuracy. Table 3.4(left) shows the confusion matrix and Table 3.5 shows the accuracy for all the test-action units from thermal and visual modalities. As we expected, thermal modality performed better than visual modality because visual imagery was affected by the illumination variance introduced in the dataset.

Phase II — Temperature Variance. In the second session (Phase II), I introduced variability in the thermal imagery by blowing an air fan (from subject’s right,

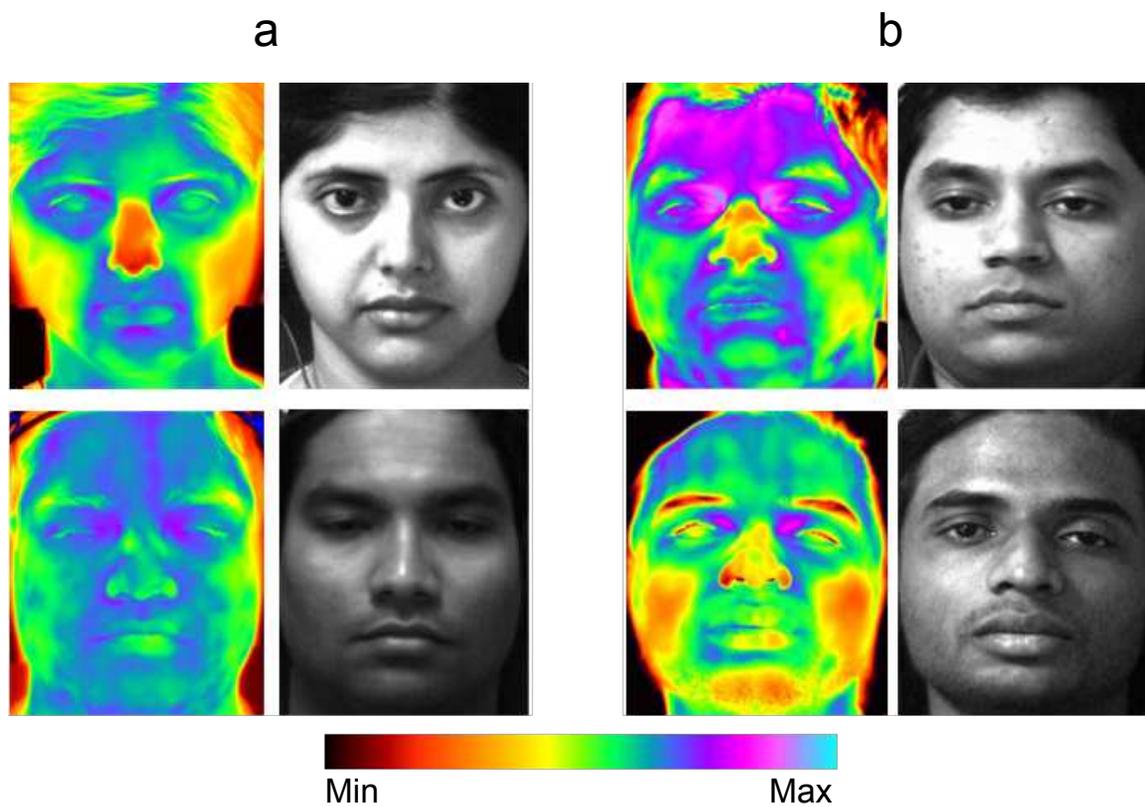


Figure 3.4: (a) A sample from Phase I (illumination variance) dataset. The top row shows the thermal and visual images acquired while lighting was from subject’s right, second row shows corresponding images from another subject when the lighting was from the top. (b) A sample from Phase II (temperature variance) dataset. The top row shows the thermal and visual images acquired while air is blown from subject’s right, second row shows corresponding images from another subject while air is blown from the front. The bottom row shows the color map used for thermal images.

left and front), affecting the subject’s thermal signature. This introduced a considerable variability in thermal imagery (as shown in Figure 3.4b), posing challenging conditions for the thermal perceptron classifier. However, the lighting in the room was maintained constant throughout the session — an ideal condition for visual imagery. The Phase II dataset consisted of a total of 448 expressions from each of the thermal and visual modalities.

Table 3.4 (right) shows the confusion matrix and Table 3.5 shows the accuracy for all the test action units from thermal and visual modalities. As we expected, the visual modality has better results in Phase II than in Phase I, since constant lighting is maintained during the data collection. However, an interesting observation is that despite the temperature variance introduced in the dataset, thermal modality remains unaffected in Phase II and has an almost-similar performance to that in Phase I.

The features fed to the classifiers are the principal components computed in each of the ROIs, which actually measures the change from neutral ROI during the expression. In the visual imagery, much of this change is a result of the formation of shadows on portions of face depending on the particular expression being made. It is possible that no new shadows are formed in the case of planar deformations or poor lighting. This is the reason why the classifier performance was poor on Phase I dataset where different lighting conditions were used during data collection. The thermal data, however, captured not only the translation, but also the deformation of the tissue due to the unique heat patterns generated on face during the expression.

Example Images	Test AUs	Modality	Phase I (Illumination Variance)								Phase II (Temperature Variance)							
			Classified AUs								Classified AUs							
			1+2	4	6+12	9	10	12	15	17	1+2	4	6+12	9	10	12	15	17
	1+2	T	50	1	0	1	0	0	0	0	40	0	0	0	0	0	0	0
		V	48	2	0	1	0	0	0	1	39	0	0	0	0	0	1	0
	4	T	1	50	0	0	0	0	0	0	0	34	0	0	0	0	1	0
		V	4	44	0	3	0	0	0	0	0	35	0	0	0	0	0	0
	6+12	T	0	0	49	0	1	0	0	0	0	0	38	0	0	1	0	0
		V	0	0	36	0	5	6	1	2	0	0	38	0	0	1	0	0
	9	T	0	1	0	48	1	1	0	1	0	0	0	54	0	0	0	0
		V	1	1	0	50	0	0	0	0	0	0	1	53	0	0	0	0
	10	T	0	0	1	0	54	0	1	0	0	0	0	0	45	0	0	0
		V	0	0	0	2	54	0	0	0	0	0	0	0	45	0	0	0
	12	T	0	0	3	0	1	45	0	1	0	0	3	0	0	40	4	2
		V	0	0	4	0	0	46	0	0	2	0	1	0	0	43	3	0
	15	T	0	2	0	0	1	1	35	0	0	0	1	0	0	2	35	1
		V	0	0	0	0	0	0	35	4	0	0	0	0	0	6	32	1
	17	T	0	0	1	0	0	0	1	53	0	0	1	0	0	0	0	51
		V	2	0	0	0	1	3	3	46	0	0	0	0	0	0	0	52

Table 3.4: Confusion matrices of Phase I (illumination variance) and Phase II (temperature variance) experiments; for thermal and visual modalities, and their fusion

	Thermal	Visual
Illumination Experiment Precision	94.81%	88.64%
Airflow Experiment Precision	94.6%	94.6%

Table 3.5: Accuracy of Phase I (illumination variance) and Phase II (temperature variance) experiments for thermal and visual modalities.

These deformations introduce variability that can always be measured by principal components. Hence, the classifier performance was same on both Phase I and Phase II datasets, even though considerable variability was introduced on the thermal data in Phase II by using an air fan.

There are a few challenges in classification of certain action units that were noticed in both modalities. The largest type of misclassification in the thermal domain is between action units 1+2 and 4. This error is caused largely by low intensity action unit 1+2, which develops a weak signal in the topmost ROI 1. This mostly resembles the low signal generated by action unit 4, and hence, confuses the perceptron classifier. In these cases, the perceptron misclassified the lower signal action unit 1+2 as action unit 4. Similarly, there is considerable misclassification between action units 1+2 and 4 in the visual approach, although the reason is slightly different. Medium to strong contraction of the frontalis (AU 1+2) creates wrinkles on the forehead, which casts shadows and in turn affects the PCA output. In a few instances, the intensity was so low that very few shadows were generated, and therefore, it was classified as action unit 4.

The second largest source of misclassification in both modalities is between action unit 12 and 15. AU 12 pulls the corner of the lips back and upwards (obliquely) creating a wide U shape to the mouth while AU 15, the lip corner depressor, pulls the lip corners down. Both of these action units produce strong signals in the ROIs placed in the buccal region (ROIs 9, 10, 11 and 12), which in turn confuses the perceptron classifier in some cases, and hence leads to misclassification.

The third largest source of misclassification in thermal imagery is between action unit combination 6+12 and 12. This error is caused when the two ROIs measuring the orbicularis oculi do not detect the subtle deformation of the skin around the eye socket.

Conclusion. The visual approach has long been considered the most powerful approach to facial expression recognition. We have shown through pilot experiments that the thermal modality can be an alternative to visual modality that can overcome some of its shortcomings, such as illumination dependency. We have collected two sessions of simultaneous thermal and visual facial expression datasets, with each session comprising a challenging variability in each modality. We noticed that the visual modality has best performance when the lighting conditions are kept constant, but the performance degraded considerably when illumination variance was introduced in the dataset. However, the thermal modality performed equally well even in the presence of heat variability in the dataset. To the best of our knowledge this is the first comparative study between the two modalities for automated facial expression recognition.

The results from the above study led to choosing thermal imagery for facial expression recognition. In the following section, I describe the actual algorithm developed.

3.2 Thermal Action Coding System — TACS

Thermal Action Coding System (TACS) is the facial expression recognition algorithm that I have developed for the thermal imagery. The methodology of this automation tool and its application in the surgical training study was published in the proceedings of the 2012 ACM Conference on Human Factors in Computing Systems (CHI) [40]. We have tracked 7 regions of interest (ROIs) on the thermal imagery of the face (figure 3.5). The tracking algorithm used is described in [46]. The ROIs were carefully chosen to align with facial muscles heavily involved in emotional action units (see table 3.6). Each ROI was abstracted by its centroid that was tracked over time forming a trajectory. The centroid of ROI-5 (nose) was used as a reference, because the nose is the most stable part of the face and is largely invariant under expressions.

Evolving Euclidean distances between centroid trajectories were used as indicators of muscle actions. Specifically, the algorithm computed the Euclidean distances $d(\mathbf{x}, \mathbf{5})$ between each ROI- \mathbf{x} ($\mathbf{x} \neq 5$) and ROI-5 from the onset till the offset of every expression (figure 3.6A). A feature vector for each expression was then formed by computing the standard deviations of these Euclidean-distance signals (figure 3.6B).

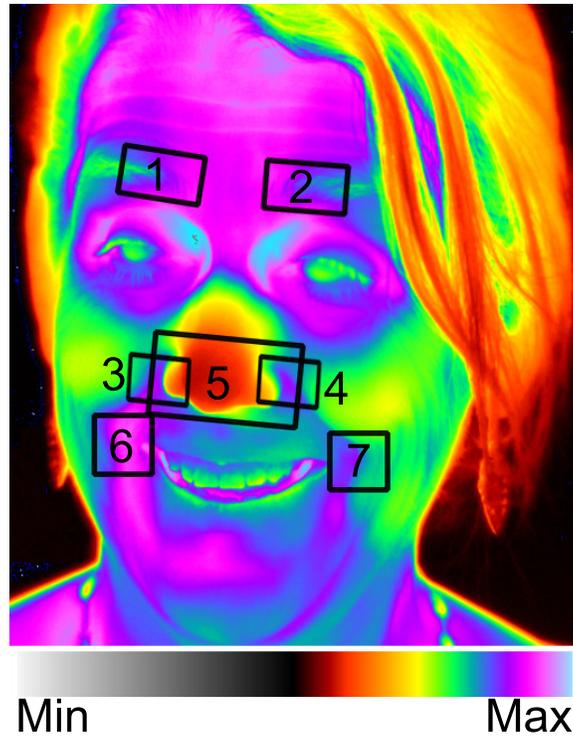


Figure 3.5: The seven ROIs used to capture facial muscle movement and deformation

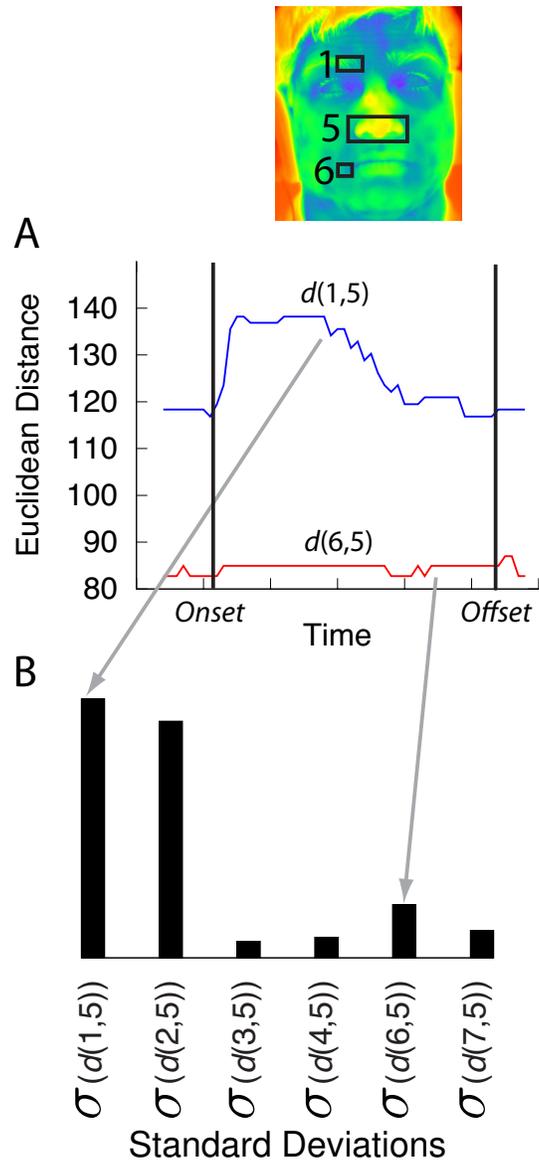


Figure 3.6: Feature vector formation for expression AU 1+2 (Inner+ Outer Eyebrow Raise)

These feature vectors capture the characteristic inter-muscle deformations over the course of expressions, and hence can be used to train a classifier. The training was done on expressions made out of 5 action unit combinations (AU 1+2, 4, 9, 10, and 12). I chose a feed-forward multilayer perceptron for classification. The multilayer perceptron featured 13 input nodes, 12 sigmoid nodes in the hidden layer, and 5 output nodes to classify expressions. The accuracy results of TACS for the two stress studies have been mentioned in the results chapter.

3.3 System Architecture

As we recall, the proposed stress model has two dimensions — arousal and valence. Perspiration measure is a proxy for arousal with higher perspiration indicating higher arousal. Similarly, the extent of the affective display via the facial expressions is a proxy for valence. Figure 3.7 depicts the system architecture for the stress model. Using the Flir thermal camera SDK, the system acquires the real-time thermal frames. On these thermal frames, one tracker can be placed on the peri-nasal area for perspiration extraction and seven trackers are placed for tracking facial muscle movement. The output of these trackers go to the perspiration extraction and facial expression recognition methods. Finally, the output from these methods are sent to the stress model module which uses the information from the valence to disambiguate the type of arousal.

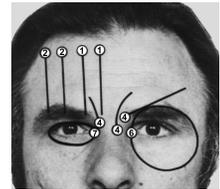
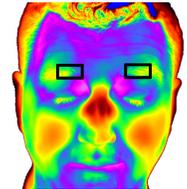
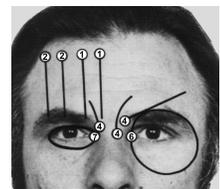
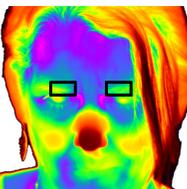
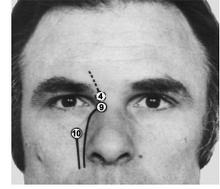
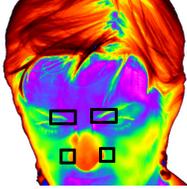
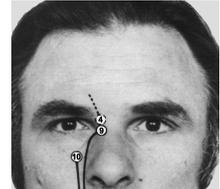
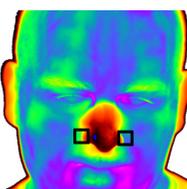
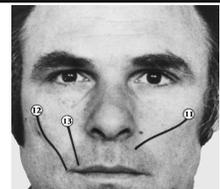
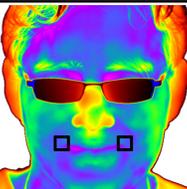
Example Image	Muscular Action	Facial Muscles	AU	ROI Placement
		Frontalis, Pars Medialis, Pars Lateralis	1+2	
		Corrugator Supercilii, Depressor Supercilii	4	
		Levator Labii Superioris Alaqua Nasi	9	
		Levator Labii Superioris	10	
		Zygomaticus Major	12	

Table 3.6: Dynamic trackers placement based on facial anatomy.

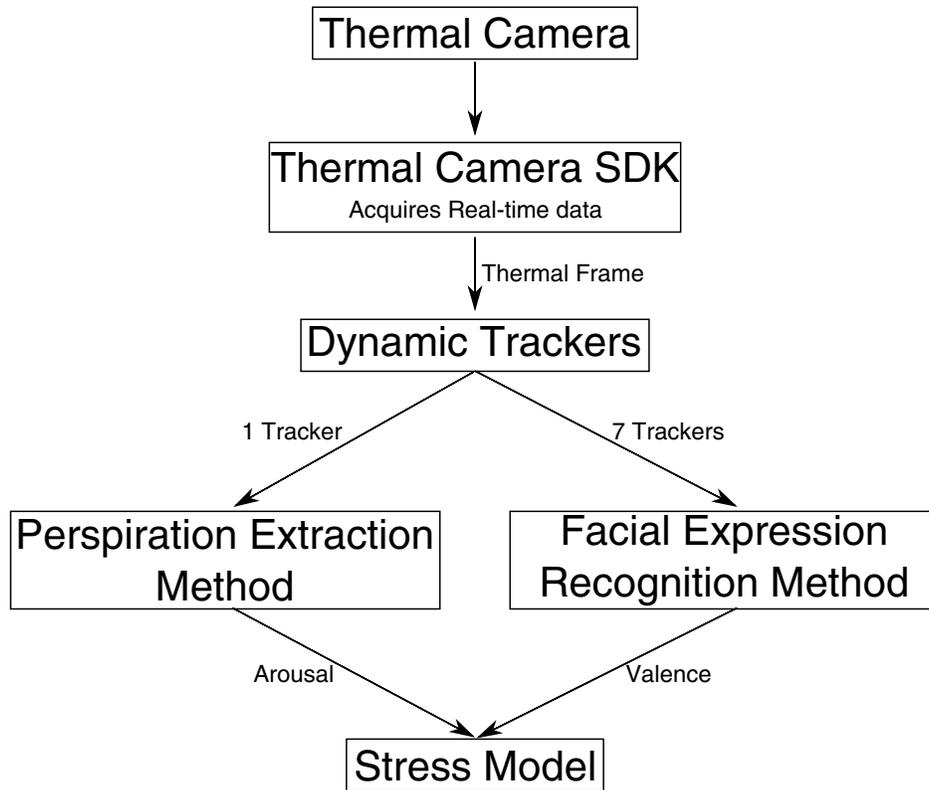


Figure 3.7: System architecture for the stress model.

Chapter 4

Conducting Stress Studies

The very first hurdle in studying stress is generating it. Stress should be studied in an environment where it arises naturally. A very important part of my dissertation was to conduct stress studies. The premise was to have stress studies with different types of stressors (see Table 4.1) and with minimal interference (without wiring the subjects). Conducting these studies took a lot of work. It involved writing an IRB, designing the experimental setup, subject scheduling, lots of practices, data collection and data management.

In each stress study, a participant's face was imaged with a thermal and visual camera throughout the experimental session. We used a Thermo Vision SC6000 Mid-Wave Infrared (MWIR) camera from FLIR [1] with a MWIR 100 mm lens and a CCD monochrome visual camera from ImagingSource [18]. The thermal imaging data were used to extract the perinasal perspiration signal and for facial expression

Study	Type of Stressor
Surgeon Stress	Dexterous Challenge + Loss of Proprioception
Operator Overloading	Cognitive Load + Motor Conflict + Emotional Stress

Table 4.1: Stress Studies designed for different types of stressors.

recognition (see figure 4.1 for more details). The visual imaging data were used to decode the participant’s facial expressions, done manually, for generating ground truth for validating my facial expression recognition algorithm (TACS).

The perspiration extraction method developed by Dr. Shastri was applied to both the stress studies. The results from these stress studies have been mentioned in the results chapter.

Perspiration Extraction Method. To quantify sympathetic responses, we extracted the transient perspiratory signal in the perinasal area from the thermal imaging data. This signal extraction was performed according to the method described in [30]. Specifically, on the initial frame of every thermal clip, we selected the Measurement Region of Interest (MROI) on the lower part of the perinasal area (see figure 4.2). This MROI was tracked over time via a tracker that was specifically designed to track facial tissue in thermal imaging [46]. The tracker was based on particle-filtering driven by a probabilistic template mechanism with spatial and temporal smoothing components.

Thermal imaging system

- FLIR (model SC6000)
- Mid-wave (MWIR) infrared camera
- Operating range 3-5 μm
- 640 X 512 pixels
- Sensitivity 0.025 $^{\circ}$ C
- 100 mm lens
- 25 fps



Visual imaging system

- FireWire CCD monochrome zoom camera
- 1024 X 768 pixels
- 15 fps



Figure 4.1: Sensing instruments

A morphology-based segmentation algorithm was applied on the MROI of each frame to quantify the perspiratory response [30]. In sympathetic excitement, perspiration pores are activated appearing as small cold spots in thermal imagery. These spots feature a distinct morphological pattern – “cold” inner area transitioning to a “hot” surrounding background. The segmentation algorithm delineated the perspiration spots, thus computing the perspiration intensity for every thermal frame.

The tracking step along with the physiological computation step were iteratively applied until the end of the thermal clip. Thus, we obtained a 1D perspiration signal from a sequence of 2D thermal frames.

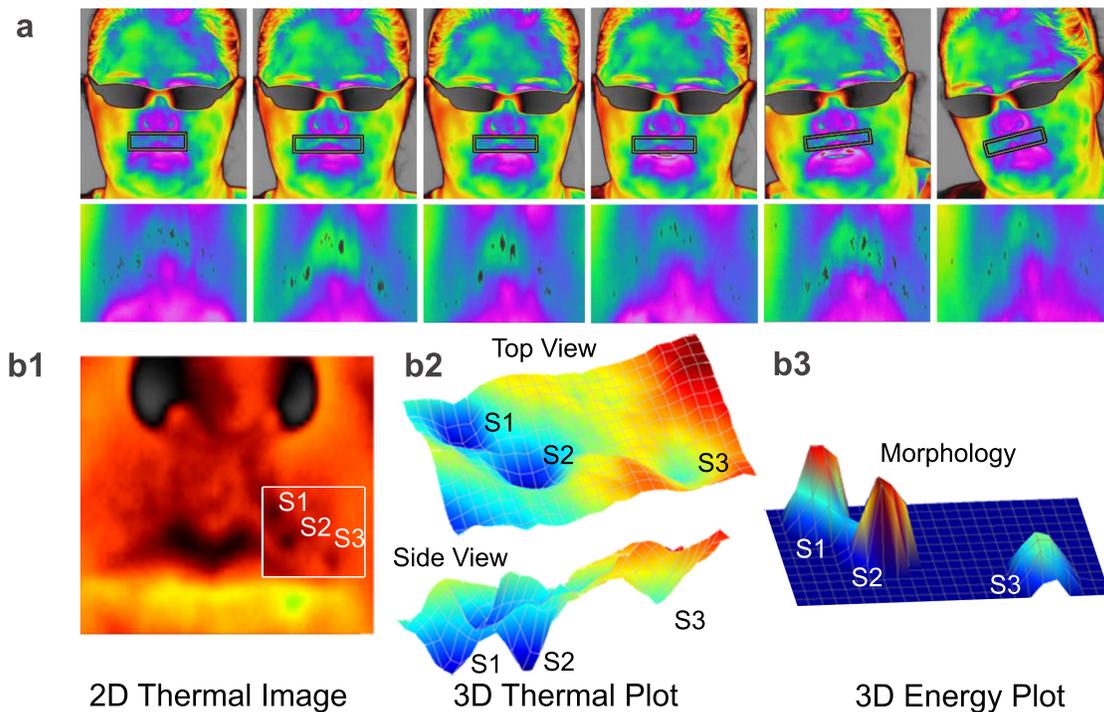


Figure 4.2: **(a)** First Image Row: Virtual tissue tracker (black rectangle) at work as a subject exhibits head motion during execution of Task 3 in Stress Study I. Second Image Row: Motion-corrected perinasal area snapshots throughout the performance period. The end effect could be considered virtual tethering of a virtual probe, so that the measurement area remains as consistent as in conventional measurements with tethered physical probes. **(b1)** Thermal image of the face. Spots S1, S2, and S3 are ‘cold’ spots indicative of perspiration. **(b2)** 3D thermal plot of the area surrounding perspiration spots S1, S2, and S3. The conic shape of the spot profiles denotes the gradual transition from a ‘cold’ core to a ‘hot’ surrounding background. **(b3)** Outcome of the morphological extraction algorithm as 3D energy plot.

4.1 Stress Study I — Surgical Training

The Surgeon Stress Study focuses on stress generated due to dexterous challenge. When it comes to stress due to dexterity, specific experimental studies focused overwhelmingly on aviation, where the effect of stress on performance is deemed paramount [25]. There have also been some studies on the effect of stress on surgical performance [3, 4, 16]. Both the aviator and surgeon professions are critical to society and involve dexterity. Due to the introduction of new technologies, such as laparoscopy in surgery and unmanned aerial vehicles in aviation, required skills in the two professions look increasingly similar (e.g., maintaining dexterity despite loss of proprioception). Emerging professions, such as robot tele-operators and actors controlling avatars, fall under the same skilled category. While this convergence of skilled professions takes place, the literature on addressing issues of stress versus performance in dexterous tasks remains fragmented (per profession) and lacks appropriate methods and unifying abstractions. Indeed, common threads in many published studies are the use of subjective or snapshot stress indicators and the reliance on non-orthogonal performance measures that are often culturally defined. The key aim of our investigation is to use our proposed objective stress-measurement method that is unobtrusive and real-time. We monitored stress and performance patterns among surgeons during training in an inanimate laparoscopic skills lab. The selected activity locus merely serves as a sample window through which we can observe the human behaviors of interest.

Subjects. Grouping was consistent with the standard categorization of surgical-skill level [36]. Specifically, $n_{Total} = 17$ surgeons randomly volunteered from: (1) a pool of novices ($n_N = 7 : 5$ male/2 female) comprised of surgical residents or technicians with no surgical-practice record and limited training in laparoscopic surgical skills; (2) a pool of experienced surgeons ($n_E = 10 : 7$ male/3 female) with an extensive surgical-practice record and at least some experience with the tested laparoscopic-surgical skills.

Table 4.2: Make-up of surgeon pool.

Level	$n(\text{Male/Female})$	Age[yr]
(1) Novices	7 (5/2)	26.0 ± 2.6
(2) Experts	10 (7/3)	35.8 ± 9.0

Data shown as mean \pm s.d.

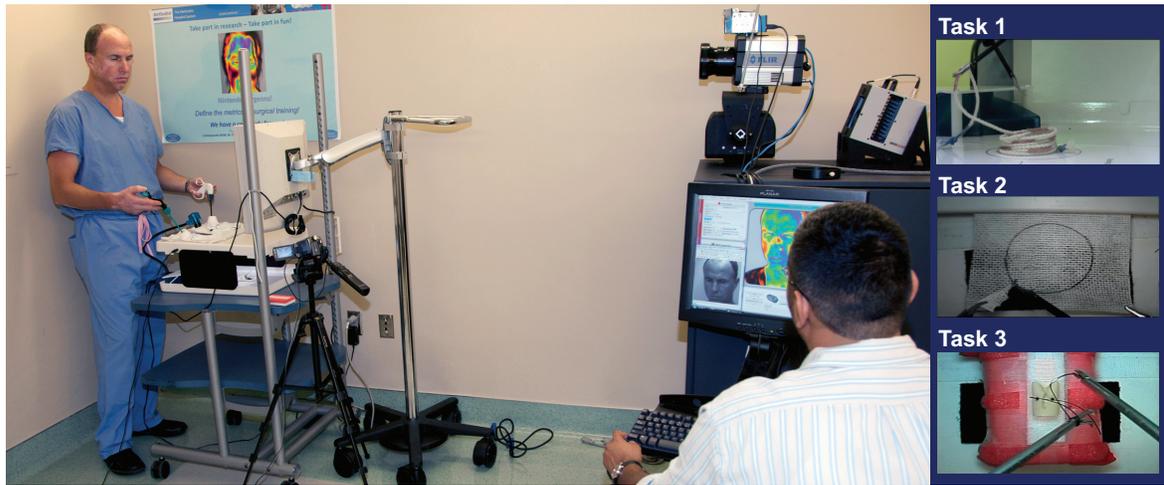


Figure 4.3: Surgeon imaged thermally and visually during trial execution in the inanimate-laparoscopic skills lab of the Methodist Institute for Technology, Innovation, and Education (MITTIESM). The panels to the right of the figure show details inside the surgical training box during the execution of the three different tasks.

The surgeons were controlled (analysis of variance, $P > 0.05$) for general psychological traits such as, anxiety [33], positive affect [38], and shyness [7] that could bias the experimental results. All surgeons were recruited from the Methodist Hospital. All training took place in the inanimate laparoscopic-skills lab of the Methodist Institute for Technology, Innovation, and Education (MITIESM) in Houston, Texas. The Institutional Review Boards of the University of Houston and the Methodist Hospital approved the study and all subjects signed informed-consent forms, including publication statements.

Study Design. The surgeons trained on three laparoscopic drills that were chosen to cover the full spectrum of difficulty according to conventional wisdom (see figure 4.3).

Task1: A simple, ad hoc, drill where a string is manipulated from one end to the other via its colored sections.

Task 2: A more challenging drill that requires the cutting of a circular pattern on a piece of gauze. It is part of the Fundamentals of Laparoscopic Surgery (FLS), a widely-accepted educational-module in laparoscopic surgery [32].

Task 3: A highly complex drill that requires precise suturing on a fine-rubber tube. This is also part of FLS.

Training was longitudinal, with repeat sessions spread over the course of a few months; every session included multiple trials of each task. In our analysis, we studied the stress indicators which included neurophysiological (via thermal imaging) and observational (via visual imaging).

Neurophysiologically, stress was tracked through the perinasal response. Specifically, in every trial i of a task j in session k for a surgeon l ($\mathbf{x} \equiv (j, k, l)$), we quantified the entire perinasal-perspiratory signal $\mathbf{E}(\mathbf{x}, i)$ and represented it via its mean intensity $\bar{\mathbf{E}}(\mathbf{x}, i)$. Then, we tracked stress by computing the mean-signal intensity $\mu_{\mathbf{E}}(\mathbf{x}) \equiv \sum_{i=1}^I \bar{\mathbf{E}}(\mathbf{x}, i)/I$ over all trials $i = 1, \dots, I$ of task j in session k for surgeon l . With the aid of observational variable (facial expressions), we were able

to disambiguate instances of negative (distress) versus positive (eustress) excitation in a sympathetic signal, such as the perinasal.

Before each session, every surgeon completed a State Anxiety Inventory (SAI) sheet [33]. Scoring of SAI gave an indication of the surgeons stress level prior to the execution of the protocol.

4.2 Stress Study II — Operator Overloading

Although concurrent performance of multiple tasks is part of human life, insufficient research has been done to understand its effect on human emotional states and performance. The purpose of this study was to develop an effective tool to gauge stress-load of operators engaging in multi-tasking. The study design focused on cell-phone and texting communication during driving simulation — a classic example of operator overloading.

Driver distraction has been a topic of discussion since the last century. The hypnotic effect of windshield wipers on driving performance was debated in 1905. Around the 1930s, the focus was on the impact of the radio programs on the primary driving task. In the current era, the debate has gained attention for cell phone usage in the vehicle. There are several studies devoted to designing a monitoring system for vehicle-drivers distraction. Yamaguchi et al. proposed the monitoring of drivers stress using biomarkers [42]. Healey et al. used four types of physiological sensors, electrocardiogram (EKG), electromyogram (EMG), skin conductivity (EDA, GSR),

and respiration (through chest cavity expansion) for monitoring driving stress [17]. Yamakoshi and his group used differential skin temperature as a driving stress index [43]. Previous work has primarily relied on on-body sensors for use in cars. On-body sensors may not be practical for continuous monitoring. In addition, they restrict users' motion and increase their awareness of being monitored. Therefore, it is not an effective way of continuous monitoring. With our current stress measurement method, we can now monitor participants passively and contact-free and thus, is suitable for continuous monitoring. We used a popular car racing game as a testing platform and introduced two ubiquitous disruptions: an involved-phone conversation with some emotional content and a text exchange. We monitored each participant's behavior during a gaming session via three complementary information channels: (a) The evolution of the participant's sympathetic responses. (b) The evolution of the participant's facial expressions. (c) The evolution of the participant's game performance. The first two channels quantified the participant's affect while the third was monitored his/her output. Once again, all three channels were real-time, objective, and unobtrusive (thermal imaging for sympathetic quantification, visual imaging for decoding of facial expressions, and screen recording for quantification of performance).

Subjects. We recruited $n = 23$ participants for the experiment. The dataset included participants of both genders (9 females/14 males) with mean age 24.56 (24.56 ± 5.45 (s.d.)) and substantive gaming experience (more than year). We excluded people without a driver's license.

Study Design. We selected a car racing game called *Test Drive: Unlimited* for this experiment. It represents a popular game genre, which requires minimal learning period, as it draws on people’s driving experience. The game runs on a XBOX-360 game console and comes complete with a steering wheel, break pedal, and gas pedal. As gaming interruptions, we introduced two cell phone activities, talking and texting, using an iPhone.

Upon arrival, each participant completed a biographic questionnaire. Next, the participant familiarized herself with the game setup for 10 min. She was also given the opportunity to familiarize herself with the iPhone operations for calling and texting. After this learning phase, the participant relaxed for 3 min (BL) during which she listened to soothing music. This relaxation phased aimed to reduce excitation built during the earlier activities, so that the main experiment starts with the participant’s sympathetic system close to the tonic level.

During the main experiment, each participant played the car racing game. Part of her gaming session featured two multitasking disruptions: phone calling (P) and texting (T). Specifically, the timeline of the gaming session was as follows: 1 min pure gaming, about 4 min gaming while talking over the phone, 1 min pure gaming, 5 min relaxation period, 1 min pure gaming, about 4 min gaming while texting, and 1 min pure gaming. In fact, we randomized the order of calling and texting in the trials. This randomization and the intervening relaxation between the multitasking periods aimed to minimize the effect of confounding factors. At the end of the experiment, the participant was debriefed regarding her game experience. This feedback



Figure 4.4: Experimental setup for the driving simulation game.

was later used for subjective analysis. The questions posed to the participant over the phone were designed to build-up her cognitive distraction and also charge her emotionally. The question-set included a combination of simple, cognitive, and uncomfortable questions:

Simple

Q1: Do not hang up until you are told so. Do you understand?

Q2: Is there anyone else in the room, yes or no?

Q3: Are you single or married?

Cognitive

Q4: Who won the American civil war, the north or the south?

Q5: What is the sum of $16 + 28$?

Q6: How many 'e' letters are in the word embarrassment?

Uncomfortable

Q7: Did you ever take anything from a place where you worked?

Q8: Did you ever tell a lie to make yourself look good?

Q9: If I told you I don't believe what you are saying would that upset you? Why?

Q10: What do you consider to be your greatest strengths?

Q11: Think about people that really irritate you. Why do they bother or annoy you?

Q12: What is the worst job you ever had and why did you dislike it?

Cognitive

Q13: My grandfather's daughter hit her daughter. How are the daughters related?

Q14: What is the sum of $58 + 74$?

Closing

Q15: What is good about talking over the phone while driving?

Q16: You may now hang up the phone and pay attention to the game

In the texting session we posed 5 simple questions to each participant. Each successive texting question required a lengthier response, thus building up the visual and motor distraction to the gamer.

Q1: Are you male or female?

Q2: Where were you born?

Q3: Type your university name and address.

Q4: What are your plans for the weekends?

Q5: What is good about texting while driving?

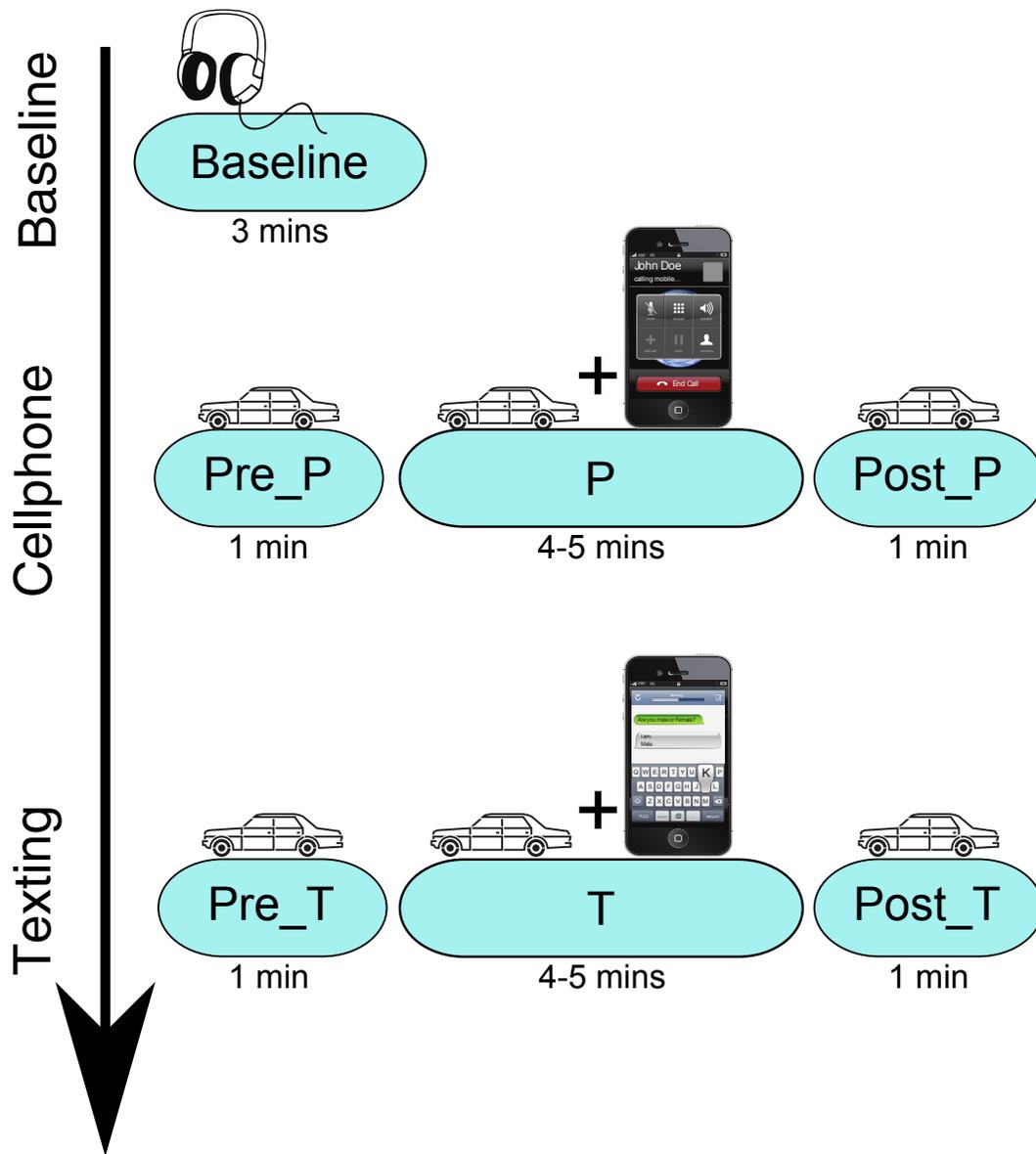


Figure 4.5: Experimental protocol for operator overloading study.

Chapter 5

Results

5.1 TACS Validation Results

A certified FACS expert decoded facial expressions in the visual stream. These were used as ground-truth to assess the accuracy of the computational-facial expression-recognition method in the thermal stream. We used multilayer perceptron with 10-fold cross-validation and percentage-split in order to test the classification accuracy. In the surgical training study, from 977 training trials (1-4 min each), I found 244 expressions made out of the five action unit combinations (AU 1+2, 4, 9, 10, and 12). These were the five AUs on which I trained my classifier. For stress study II, I found 95 such facial expressions. For study II there wasn't enough expressions of AUs 9 and 10 to train the classifier. I also applied my facial recognition software on an extra dataset from the deception study conducted by the CPL lab in the University

of Houston. In this extra dataset, there were a total of 185 expressions from of the above 5 types of AU combinations.

The recognition rate for the facial expressions from the two stress studies and the extra dataset were 81.55%, 80% and 79.46% respectively — an excellent performance if one takes into account the realism of these studies (see tables 5.1 5.2 5.3). The method accuracy on the conglomeration of facial expressions from all the stress studies was 75% (see table 5.4). The above accuracy results are a testament to the robustness and scalability of TACS. Hence, one can use this automated method in field practice (as we did) to disambiguate physiological signals. The gain is tremendous savings in labor and expense at the cost of some accuracy. If the data set is large, such as in the above studies, one can argue that the benefits far outweigh any losses.

5.2 Stress Analysis for the Two Stress Studies

5.2.1 Study Variables

Our stress model requires the observational annotation of the neurophysiological response in its arousal dimension and the extent of the facial expression as the valence dimension. This resulted in a more detailed level of stress analysis. Specifically, we quantified just the portions of the perinasal perspiratory signal where the participants showed facial expressions manifesting negative feelings (distress); let us denote this

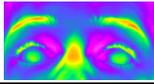
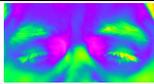
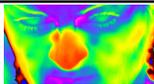
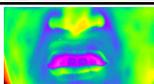
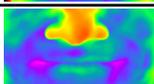
	Test	Classified AUs				
	AUs	1+2	4	9	10	12
	1+2	6	5	3	3	1
	4	3	39	0	3	0
	9	2	1	5	7	5
	10	0	2	1	69	3
	12	1	1	0	4	80

Table 5.1: Confusion matrix for facial expressions from Stress Study I — Surgical Training.

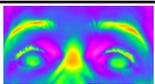
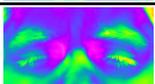
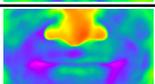
	Test	Classified AUs		
	AUs	1+2	4	12
	1+2	7	9	2
	4	1	38	5
	12	1	1	31

Table 5.2: Confusion matrix for facial expressions from Stress Study II — Operator Overloading.

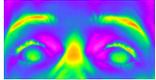
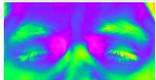
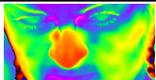
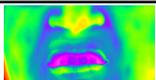
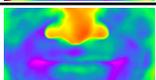
	Test	Classified AUs				
	AUs	1+2	4	9	10	12
	1+2	53	9	1	0	2
	4	9	34	0	0	3
	9	2	0	4	0	1
	10	1	0	0	0	2
	12	4	4	0	0	56

Table 5.3: Confusion matrix for facial expressions from an extra dataset from the deception study conducted by the CPL lab in the University of Houston.

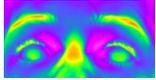
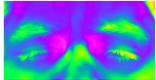
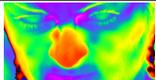
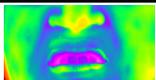
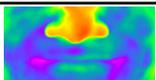
	Test	Classified AUs				
	AUs	1+2	4	9	10	12
	1+2	52	29	3	10	8
	4	14	103	0	7	11
	9	7	3	5	10	3
	10	0	3	0	71	7
	12	3	9	1	4	166

Table 5.4: Confusion matrix for facial expressions from the two studies and the extra dataset.

negative affect signal as $E_N(x)$ (with mean $\overline{E}_N(x)$) and its extent (percent of total frames in the trial) as $N(x)$. In this case, we tracked stress by computing the mean signal intensity $\mu_{E_N}(x) \equiv \sum_{i=1}^I \overline{E}_N(x)/I$ over all the negative affect signal portions $i = 1, \dots, I$. We also computed the mean extent $\mu_N(x) \equiv \overline{N}(x)$ of the negative affect signal portions. Therefore, at this level of analysis, distress changes were evident not only via the changes of $\mu_{E_N}(x)$, but also via the changes of $\mu_N(x)$ (see Figure 5.1).

At the same time, we tracked positive excitation by quantifying the portions of the perinasal perspiratory signal where the surgeon had facial expressions manifesting positive feelings (eustress); let us denote this positive affect signal as $E_P(x)$ (with mean $\overline{E}_P(x)$) and its extent (percent of total frames in the trial) as $P(x)$. These positive-affect signal portions were characterized by mean intensity, $\mu_{E_P}(x)$, as well as mean extent, $\mu_P(x)$, similarly to the negative affect signal portions. Therefore, eustress changes were evident either via the changes of $\mu_{E_P}(x)$ or $\mu_P(x)$.

5.2.2 Stress Analysis on Stress Study I — Surgical Training

5.2.2.1 Qualitative Analysis

Figure 5.2 shows a characteristic example of eustress from the study set. About 130 seconds into the drill, the perinasal signal of the surgeon exhibited elevation, which is characteristic of arousal. The signal by itself is not informative as to the type of arousal (positive or negative). However, once the observational channel is included in the analysis, it can be safely inferred that this is a bout of eustress.

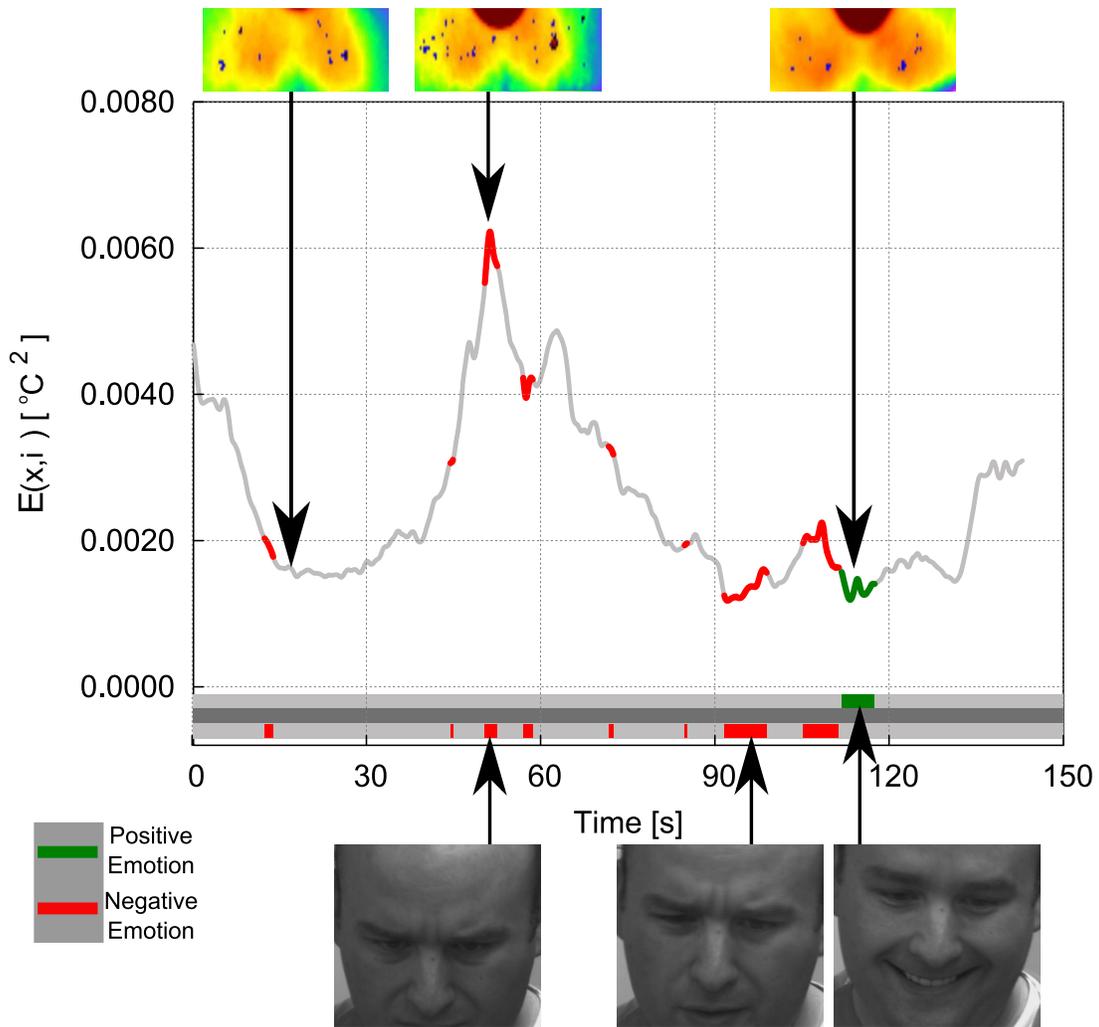


Figure 5.1: The gray 1D signal is the perspiration signal. The horizontal red and green bars at the bottom are the annotated time intervals for negative and positive facial expressions. With the aid of observations, the perspiration signal can be disambiguated into distress or eustress signal.

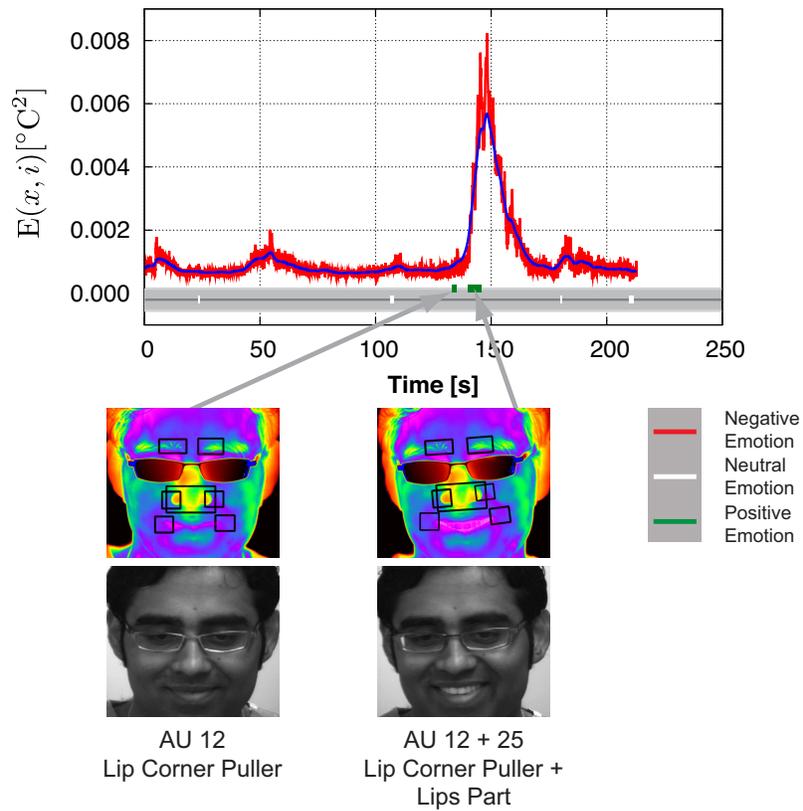


Figure 5.2: A bout of eustress for an experienced surgeon (D005), as indicated by the locally elevated perinasal signal and the positive facial expressions.

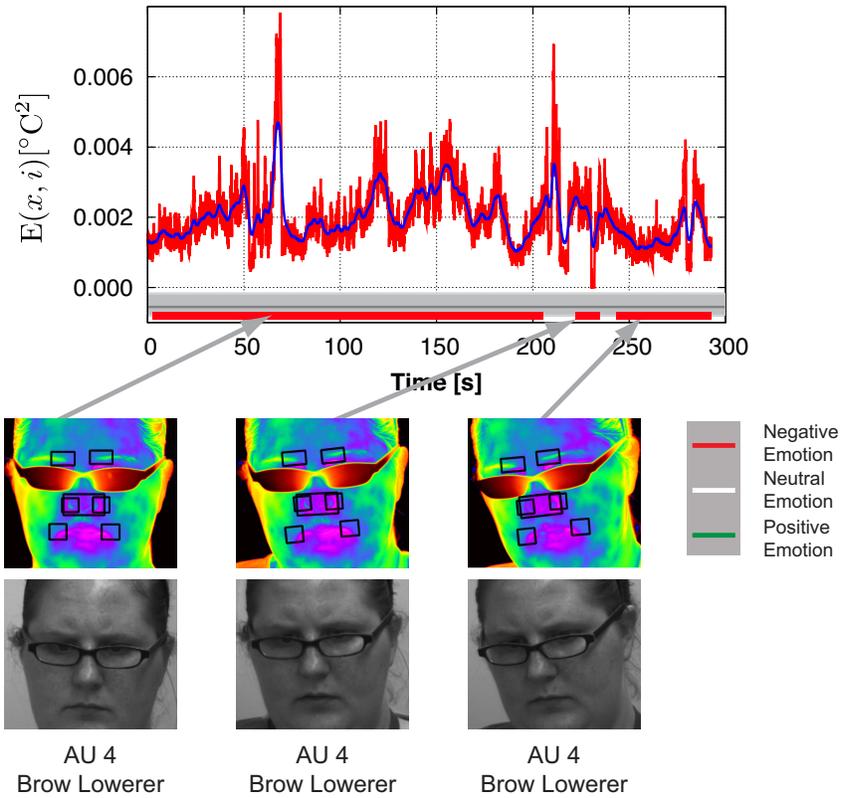


Figure 5.3: Bouts of distress for a novice surgeon (D025), as indicated by the fluctuating perinasal signal and the negative facial expressions.

Indeed, the experimental context supports this conclusion, as this is an experienced surgeon who successfully addressed a technical challenge towards the end of the drill, causing the physiological arousal to be accompanied by a sense of accomplishment.

Figure 5.3 shows a characteristic example of distress from the experimental set. The surgeon seems to be undergoing a roller-coaster of emotional arousal. The distressing type of arousals is informed by the observational channel. Indeed, this is a

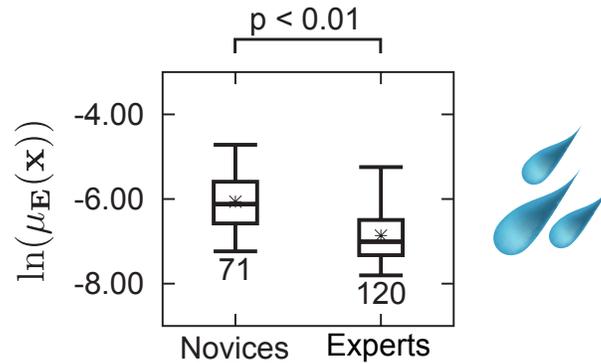


Figure 5.4: Distribution per skill level of mean perspiration intensity $\mu_E(x)$ ($\ln(\cdot)$ transformation to comply with the ANOVA assumptions). $\ln(\mu_E(x))$ is -6.0626 ± 0.6926 (s.d.) for novices and -6.8614 ± 0.6319 (s.d.) for expert surgeons.

novice surgeon who performed multiple errors during the execution of the drill, causing a sequence of physiological arousals to be accompanied by a sense of foreboding.

This unobtrusive and highly automated quantification and qualification of emotional states may prove a powerful tool in large field studies of human-machine interactions.

5.2.2.2 Quantitative Analysis

Looking purely at the unguided perspiration signal, we notice that novice surgeons perspire significantly more than the expert surgeons (see figure 5.4). And since these signals are agnostic to the types of stress, we are restricted by this limitation to infer further.

However for the same study, after using the proposed stress model, whereby we focus on the portions of the perspiration signals marked by the negative facial expressions we can infer that novice surgeons experience significantly higher distress. The boxplots in figure 5.5 depicts the distress arousal as well as its extent. Novice surgeons have significantly higher distress arousal and significantly higher distress extent. Since the stress model is a dimensional model we see in the scatter plot distress for each surgeon for different trials 5.6. The stress profile for the novices (represented by the red points) can be seen clustered around the upper left part of the quadrant while the experts (blue points) have lower arousal and lower negative valence. The average stress profile for the novice and expert surgeon cohorts is depicted by the solid colored red and blue points respectively.

These results look intuitive, however prior to this research there wasn't any quantifiable means to study stress with its nuances.

Another auxiliary finding was that positive expressions were extremely rare. In fact, the experienced surgeons were almost exclusively the ones who exhibited brief positive feelings during task execution (analysis of variance, $P < 0.01$ - 5.8). It was particularly intriguing that some of these positive feelings were associated with occasional errors (figure 5.7) - an indication that experienced surgeons positively embraced slight misfortune, perhaps as an antidote to boredom. These results further consolidate our study design which was primarily about stress generation — we were expecting higher stress for the novice surgeons and little positive expressions. These results were published in the National Center for Human Performance (NCHP,

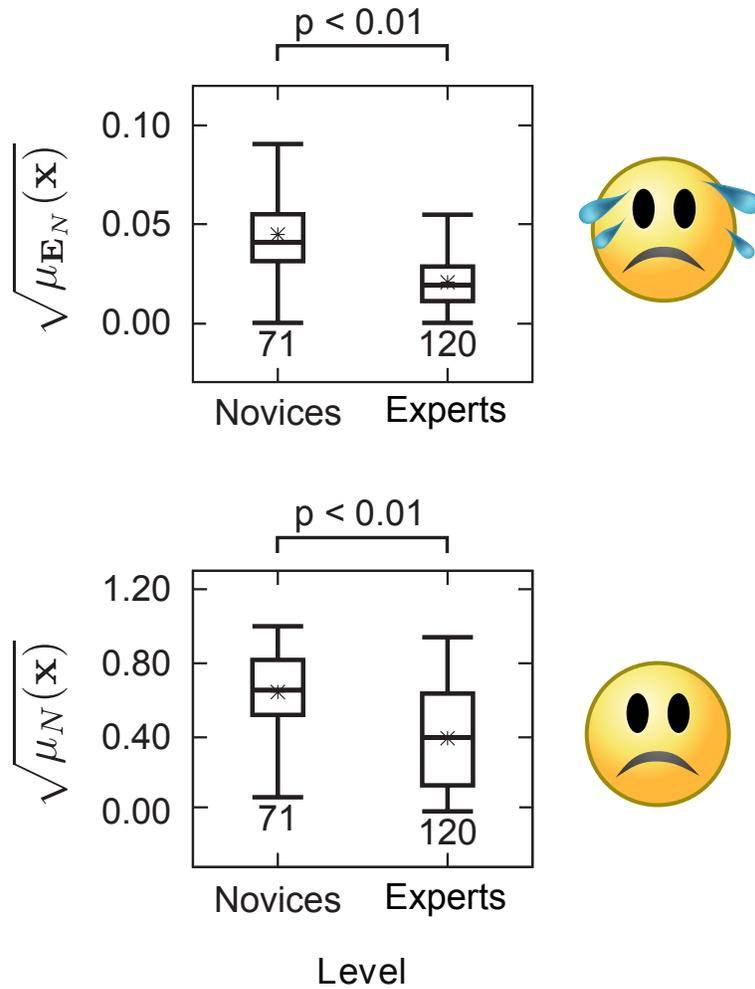


Figure 5.5: **(a)**. Distribution per skill level of mean stress responses during expressed negative feelings $\mu_{E_N}(x)$ ($\sqrt{\cdot}$ transformation to comply with the ANOVA assumptions). $\sqrt{\mu_{E_N}(x)}$ is 0.0447 ± 0.0204 (s.d.) for novices and 0.0205 ± 0.0169 (s.d.) for experienced surgeons. **(b)** Distribution per skill level of mean distress extent $\mu_N(x)$ ($\sqrt{\cdot}$ transformation to comply with the ANOVA assumptions). $\sqrt{\mu_N(x)}$ is 0.6457 ± 0.2279 (s.d.) for novices and 0.3963 ± 0.2928 (s.d.) for experienced surgeons.

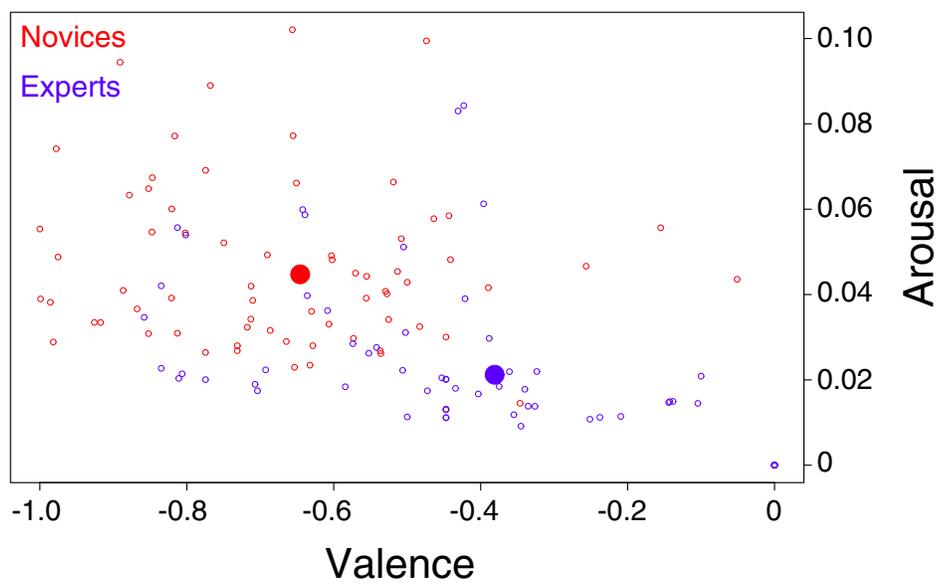


Figure 5.6: The stress model representing distress for all the trials of all the surgeons. The red points are distress levels for the novice surgeons and the red points are for the expert surgeons. The bigger solid points represents an average distress level for the each surgeon level.

2010) [2] and also in the Scientific Reports [28].

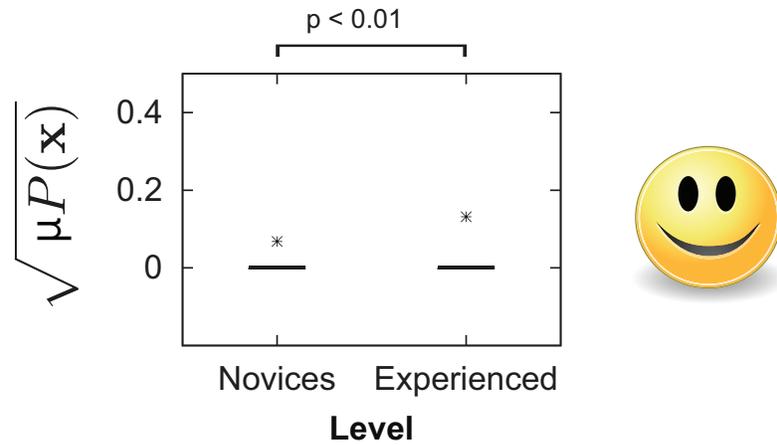


Figure 5.8: Distribution of extent of expressed positive feelings $\mu_P(x)$ during tasking per skill level. For novices $\sqrt{\mu_P(x)}$ is 0.0682 ± 0.1938 (s.d.) while for experienced surgeons is 0.1316 ± 0.2946 (s.d.). We used $\sqrt{\cdot}$ transformation to comply with ANOVA assumptions.

5.2.3 Stress Analysis on Stress Study II — Operator Overloading

Protocol. Upon arrival each participant completed a biographic questionnaire. Next, the participant familiarized herself with the game setup for 10 min. She was also given the opportunity to familiarize herself with the iPhone operations for calling and texting. After this learning phase, the participant relaxed for 3 min (BL) during which she listened to smooth music. This relaxation phase aimed to reduce

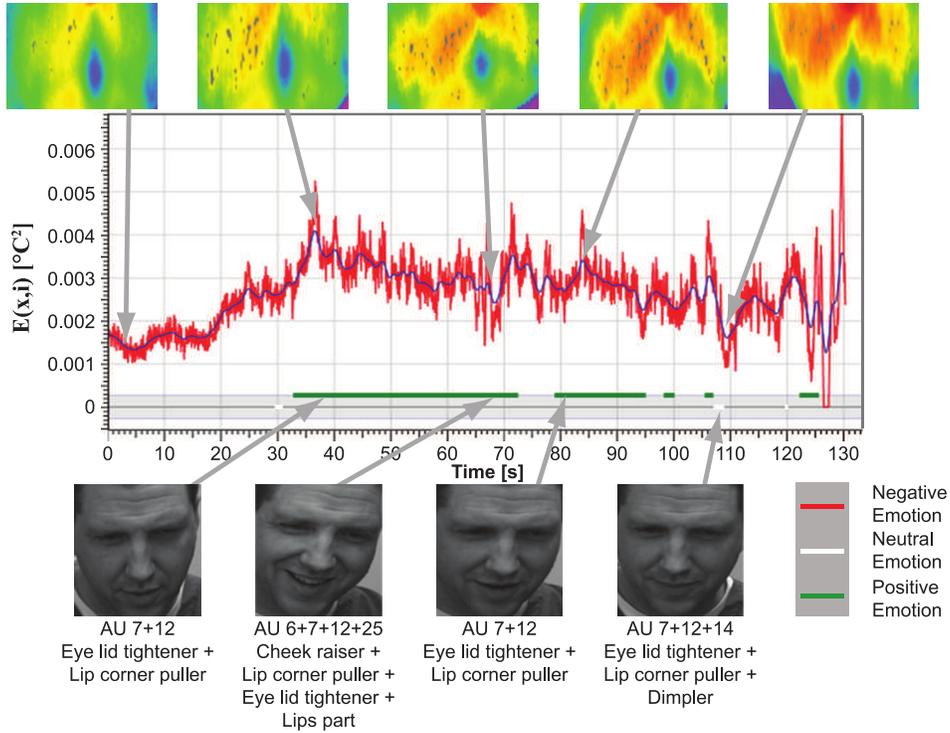


Figure 5.7: Experienced surgeon’s (subject ID: D001) thermo-physiological (peri-nasal) and observational (facial) images during execution of Task 3, Session 4, Trial 2. The corresponding perspiration (stress) signal is shown in the middle. There is an elevation in the signal due to some excitation, about 30 sec into the execution of the trial. This excitation is positive, as the FACS-decoding [12] of facial expressions indicates along the timeline (bottom). This response was due to an error committed by the surgeon (needle drop), from which he quickly recovered.

excitation built during the earlier activities, so that the main experiment starts with the participants sympathetic system close to the tonic level.

During the main experiment, each participant played the car racing game. Part of the gaming session featured two multitasking disruptions: phone calling (P) and texting (T). Specifically, the timeline of the gaming session was as follows: 1 min pure gaming, about 3 min gaming while talking over the phone, 1 min pure gaming, 5 min relaxation period, 1 min pure gaming, about 3 min gaming while texting, and 1 min pure gaming. The order of calling and texting was randomized 5.9.

Thus, for each subject we start with the Baseline period and we have two sessions: P and T, each on consisting of three segments: Pre, Activity, Post, where in pre and post we have pure game play, while in the activity segment the subject plays the phone and simultaneously is texting/talking on the iPhone. Using thermal camera, we measure the perinasal perspiration intensity for which we obtain the mean value over each segment of each session of the experiment.

5.2.3.1 Quantitative Analysis

Using thermal camera, we measure the perinasal perspiration intensity for which we obtain the mean value over each segment of each session of the experiment (see figure 5.10). To comply with the ANOVA assumptions, we transformed the mean intensity using the natural logarithm function (to achieve equal variance among the groups). For each of the 23 participating subjects, we have a total of 7 repeated

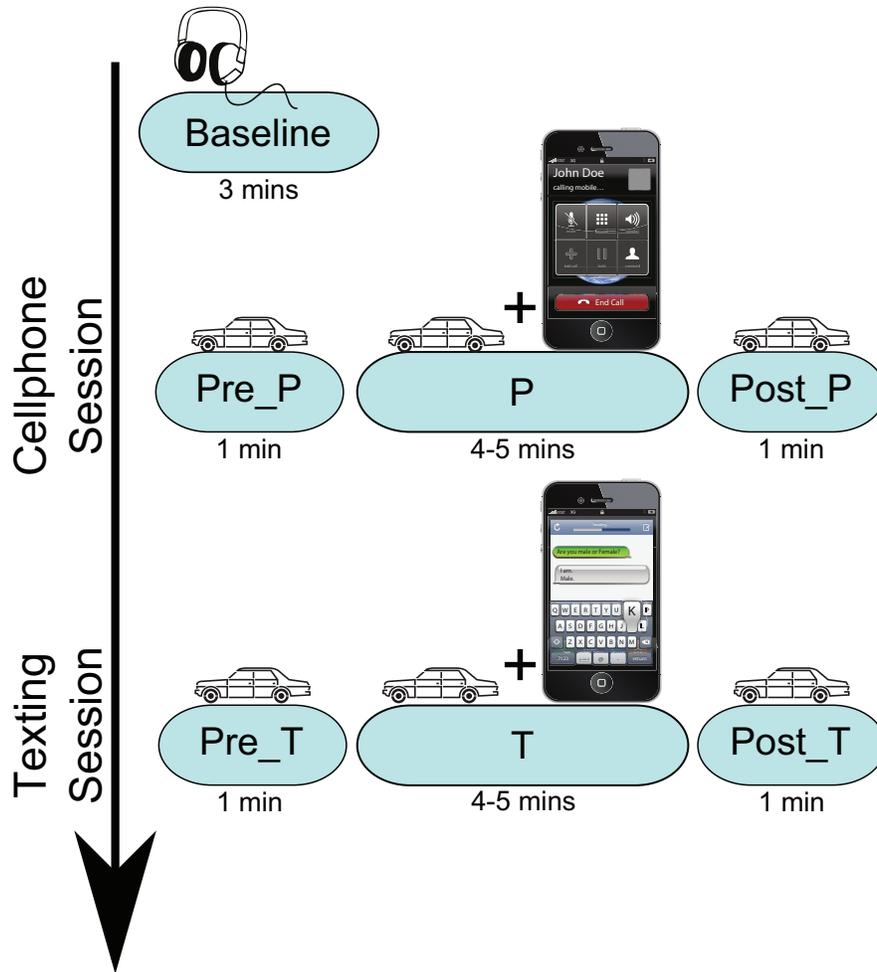


Figure 5.9: Protocol for the operator overloading. There are three sessions — Baseline, Cellphone, and Texting. Cellphone and Texting sessions are further divided into three segments — solo game play, interrupted game play, and post solo game play.

measurements of the $\ln(\text{mean perspiration})$ during each of the 7 segments of the experiment (BL, Pre_P, P Post_P, Pre_T, T, Post_T). In this repeated measure design, the segment is a fixed Factor (taking 7 values) while the Subject (taking 23 values for each of the participants) is a random factor.

We ran the repeated measure ANOVA for this data and observed that there seem to exist significant differences among the different segments of the experiment (ANOVA, $p < 0.01$). Furthermore, this significance is lost when we remove the multi-tasking segments (ANOVA, $p > 0.01$). These results consolidate the study design of generating sufficient arousal for the multitasking sessions while maintaining uniform arousal otherwise (see figure 5.10).

Next, we compared the physiological responses for the multi-tasking segments only and found that there was a significant difference in the mean responses. The physiological arousal for the P segment was significantly higher than the T segment.

Finally, we applied our proposed stress model to the multi-tasking segments. We observed no statistical difference in the distress levels. Both in the distress arousal and the distress extent, there was no significant difference (see figures 5.11). We also observe no statistical difference in the amounts of error committed in the two multi-tasking segments (figure 5.13). Based on the above results, we can conclude that emotionally loaded phone conversations can be as detrimental to the driver as sending text messages while driving. This quantitative conclusion about driver's distress levels was only made possible by the proposed stress model. Figure 5.12 plots the distress profile for all the subjects in the multi-tasking segments. The red

points are for the phone conversation segment and the blue points are for the texting segment. In the stress model's distress quadrant, the blue and the red points are in proximity to each other.

When we applied the stress model in the operator overloading study, we noticed that the distress levels were not significantly different in the two multitasking segments. However, the overall unguided perspiration levels were significantly different in those two multi-tasking segments. In other words, the specific analysis was different than the unguided analysis. This was not the case in the surgical training study. There we noticed the novice surgeons had significantly higher perspiration level and also significantly higher distress level. This differences in the results from the two stress studies can be partly understood by plotting the percentage of times participants make negative and positive facial expressions — since the overall unguided signal contains both the distress and eustress portions. Figure 5.14 indicates the potential reason for the varying results in the two stress studies. The operator overloading study had a higher percentage of positive facial expressions and this affected the overall perspiration signal that included the eustress portions. In the surgical training study, since there were fewer positive facial expressions, the contribution of the eustress portion to the overall perspiration was minimal.

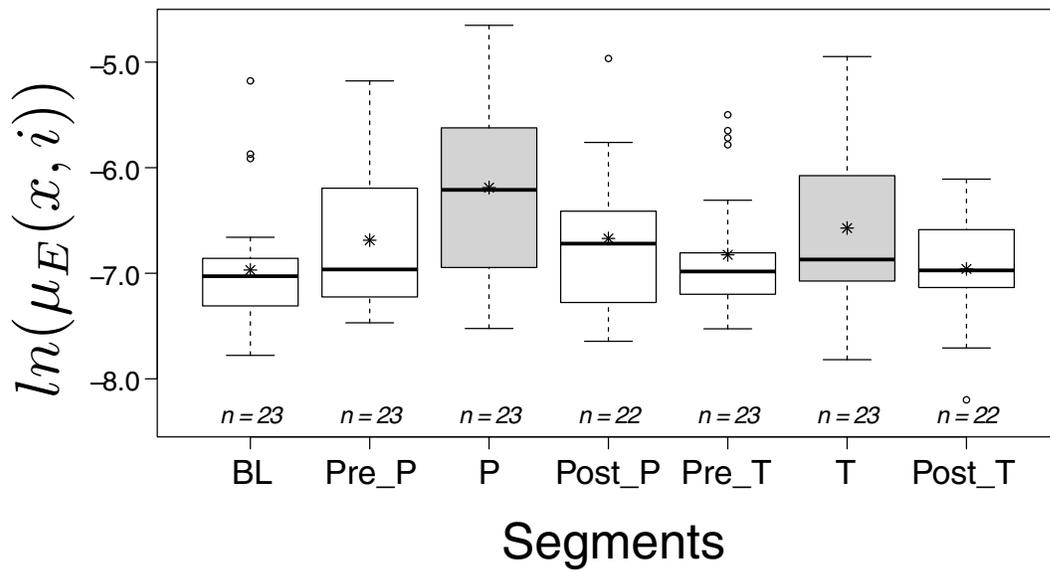


Figure 5.10: Box-plots for the overall perspiration response for all the subjects across the segments. We used the $\ln(\cdot)$ transformations to comply with the analysis of variance assumptions.

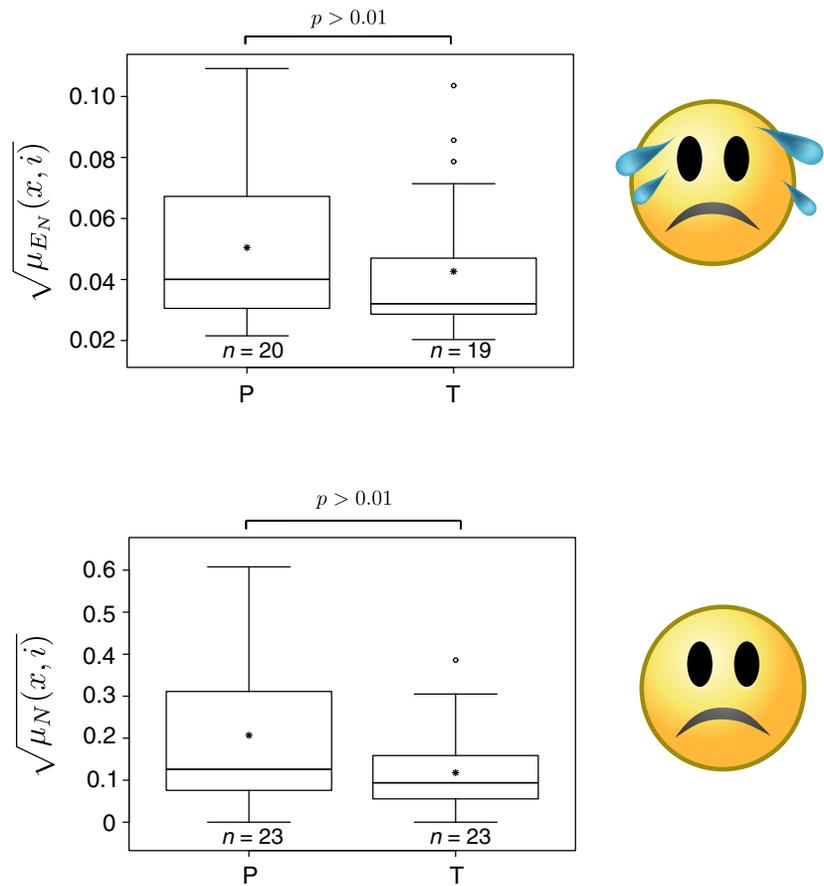


Figure 5.11: Box-plots denoting the distress intensity and distress extent for the multi-tasking segments. We used the sqrt (.) transformations to comply with the analysis of variance assumptions.

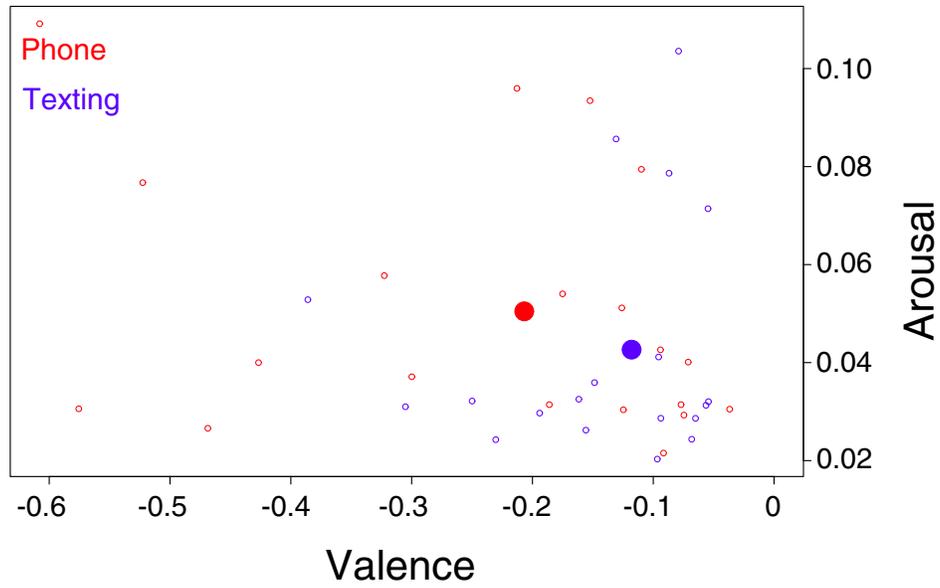


Figure 5.12: The stress-model representing distress for all the subjects in the multi-tasking segments. The red points are distress levels for the phone segment and the blue points are for the texting segment. The bigger solid points represent an average distress level for each segment.

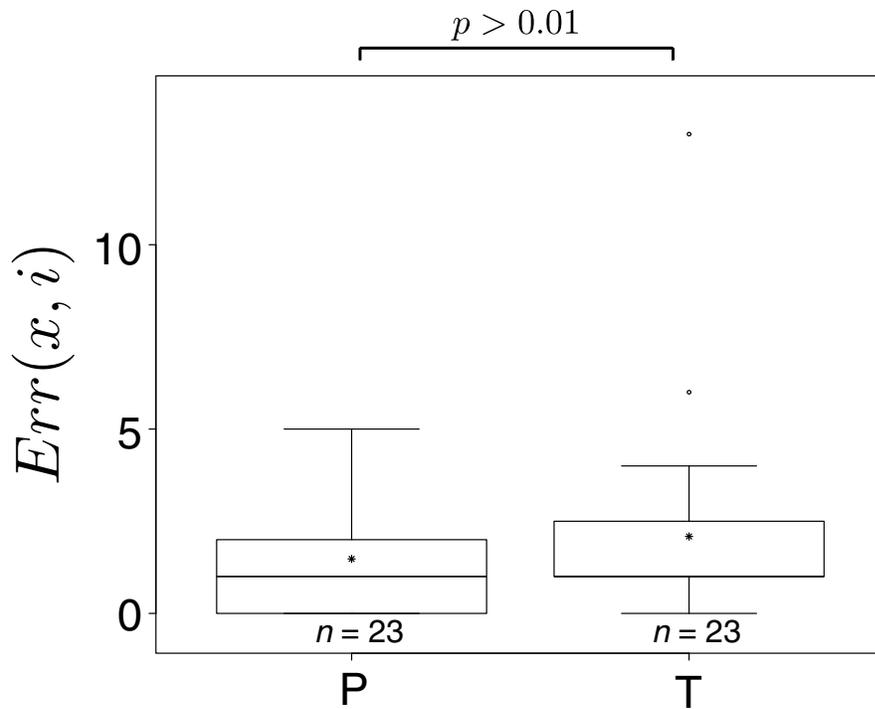


Figure 5.13: Box-plots for errors committed in the multi-tasking segments. We used the $\text{sqrt}(\cdot)$ transformations to comply with the analysis of variance assumptions.

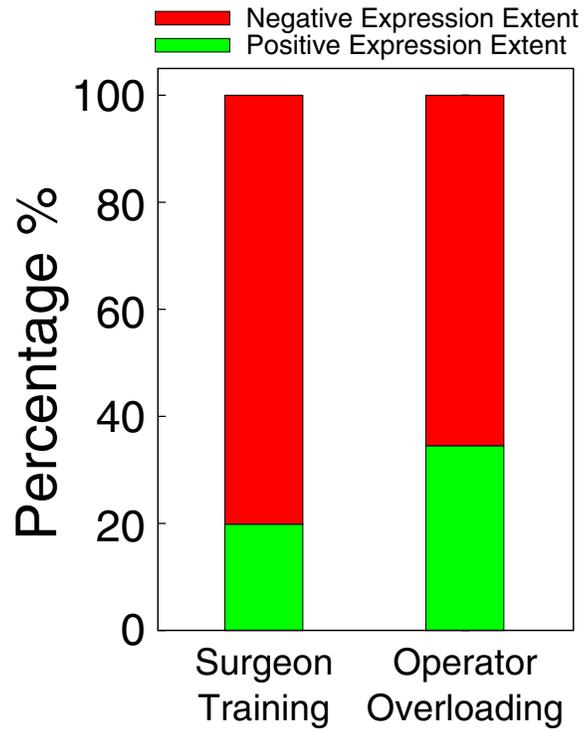


Figure 5.14: The plot depicts the percentage of the negative and positive facial expressions made by the participants in each study. Green region represents the positive facial expression percentage and the red region represents the negative facial expression percentage.

Chapter 6

Conclusion

The detailed stress analysis on the two stress studies was made possible by the introduction of a new stress model of combining physiology (perinasal perspiration in our case) with the observation of facial expressions. Using the suite of perspiration extraction method developed by Dr. Shastri along with my proposed facial expression recognition method (TACS), we are now capable of unobtrusively quantifying human stress responses in natural settings. Instead of using the traditional visual imagery for expression recognition, I used thermal imagery. The benchmarking experiment conducted to compare the two modalities for expression recognition showed that thermal imagery has an added benefit of being minimally affected by illumination variance. This enhanced the robustness of TACS and made it possible to be used in field studies. If the result of my current work is any guide, the method is not only a valuable tool for scientific discovery, but also a technology that may be used in

practice to objectively assess human performance under stress.

Future studies may benefit from the proposed stress model that is not only comprehensive (quantitative and qualitative), but also economical (single imaging modality with no labor). A future work to my dissertation would be to compare distress with eustress for participants in the two stress studies.

Bibliography

- [1] FLIR Systems, <http://www.flir.com>.
- [2] Fast by nature-how stress patterns define human experience and performance. In *National Center for Human Performance*, 2010.
- [3] S. Arora, N. Sevdalis, R. Aggarwal, A. Sirimanna, A. Darzi, and R. Kneebone. Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surgical Endoscopy*, 24:2588–2593, 2010.
- [4] S. Arora, N. Sevdalis, D. Nestel, M. Woloshynowych, A. Darzi, and R. Kneebone. The impact of stress on surgical performance: A systematic review of the literature. *Surgery*, 147(3):318–330, 2010.
- [5] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Recognizing facial expression: Machine learning and application to spontaneous behavior. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR '05, pages 568–573, San Diego, California, 2005. IEEE Computer Society.
- [6] M. S. Bartlett, G. Littlewort, C. Lainscsek, I. Fasel, and J. Movellan. Machine learning methods for fully automatic recognition of facial expressions and facial actions. In *Proc. IEEE Int'l Conf. Systems, Man and Cybernetics*, pages 592–597, 2004.
- [7] K. Bortnik, L. Henderson, and P. Zimbardo. The shy q, a measure of chronic shyness: Associations with interpersonal motives, interpersonal values and self-conceptualizations. In *36th Annual Conference of the Association for the Advancement of Behavior Therapy*, 2002.

- [8] M. Dcosta, D. Shastri, and I. Pavlidis. Perinasal indicators of malevolence. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 1, pages 1–4, May 2015.
- [9] D. Diamond, A. Campbell, C. Park, J. Halonen, and P. Zoladz. The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the yerkes-dodson law. *Neural Plasticity*, (60803):1–33, 2007.
- [10] P. Ekman. An argument for basic emotions. In *Cognition and Emotion*, 6, 169–200, 1992.
- [11] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, 1978.
- [12] P. Ekman, W. Friesen, and J. C. Hager. *Facial Action Coding System Investigators Guide*. 2002.
- [13] P. Ekman, W. Friesen, and J. C. Hager. *Facial Action Coding System Investigator’s Guide*. 2002.
- [14] B. Fasel and J. Luetttin. Automatic facial expression analysis: A survey. *PATTERN RECOGNITION*, 36(1):259–275, 1999.
- [15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: An update. *SIGKDD Explorations*, 11, 2009.
- [16] I. Hassan, P. Weyers, K. Maschuw, B. Dick, B. Gerdes, M. Rothmund, and A. Zielke. Negative stress-coping strategies among novices in surgery correlate with poor virtual laparoscopic performance. *British Journal of Surgery*, 93:1554–1559, 2006.
- [17] J. Healey and R. Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166, 2005.
- [18] ImagingSource. <http://www.theimagingsource.com>.
- [19] M. M. Khan, M. Ingleby, and R. D. Ward. Automated facial expression classification and affect interpretation using infrared measurement of facial skin temperature variations. *ACM Trans. Auton. Adapt. Syst.*, 1:91–113, 2006.

- [20] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi. Recent advances in visual and infrared face recognition - a review. *Computer Vision and Image Understanding*, 97:103–135, 2005.
- [21] I. Kotsia and I. Pitas. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *Image Processing, IEEE Transactions on*, 16(1):172–187, 2007.
- [22] R. Lazarus, J. Deese, and S. Osler. The effects of psychological stress upon performance. *Psychological Bulletin*, 49:293–317, 1952.
- [23] J. Levine, I. Pavlidis, and M. Cooper. The face of fear. *Lancet*, 357(9270):1757, 2001.
- [24] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett. The computer expression recognition toolbox (cert). In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 298–305, 2011.
- [25] M. Morphew. The future of human performance and stress research: A new challenge. In P. Hancock and P. Desmond, editors, *Stress, Workload, and Fatigue*, pages 249–262. Lawrence Erlbaum Associates, 2001.
- [26] R. E. Nisbett and T. D. Wilson. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84:231–259, 1977.
- [27] M. Pantic, S. Member, and L. J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1424–1445, 2000.
- [28] I. Pavlidis, P. Tsiamyrtzis, D. Shastri, A. Wesley, Y. Zhou, P. Lindner, P. Budharaju, R. Joseph, Mandapati, A., B. Dunkin, and B. Bass. Fast by nature — how stress patterns define human experience and performance in dexterous tasks. *Scientific Reports*, 2, 2012.
- [29] D. Shastri, A. Merla, P. Tsiamyrtzis, and I. Pavlidis. Imaging facial signs of neurophysiological responses. *IEEE Transactions on Biomedical Engineering*, 56(2):477–484, 2009.
- [30] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. Pavlidis. Perinasal imaging of physiological stress and its affective potential. *IEEE Transactions on Affective Computing.*, 3(3), 2012.

- [31] D. A. Socolinsky, L. B. Wolff, J. D. Neuheisel, and C. K. Eveland. Illumination invariant face recognition using thermal infrared imagery. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:527, 2001.
- [32] N. Soper and G. Fried. Fundamentals of laparoscopic surgery: Its time has come. *Bulletin of the American College of Surgeons*, 93(9):30–32, 2008.
- [33] C. Spielberger, R. Gorsuch, and R. Edward. *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press, 1970.
- [34] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez. Automatic feature localization in thermal images for facial expression recognition. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops - 03*, 14–21, Washington, DC, USA, 2005. IEEE Computer Society.
- [35] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. Pavlidis, M. Frank, and P. Ekman. Imaging facial physiology for the detection of deceit. *International Journal of Computer Vision*, 71(2):197–214, 2007.
- [36] S. Tsuda, D. Scott, J. Doyle, and D. Jones. Surgical skills training and simulation. *Current Problems in Surgery*, 46(4):261–371, 2009.
- [37] A. Uncini, S. Pullman, R. Lovelace, and D. Gambi. The sympathetic skin response: Normal values, elucidation of efferent components and application limits. *Journal of the Neurological Sciences*, 87:299–306, 1988.
- [38] D. Watson, L. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: The panas scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070, 1988.
- [39] A. Wesley, P. Buddharaju, R. Pienta, and I. Pavlidis. A comparative analysis of thermal and visual modalities for automated facial expression recognition. In *International Symposium on Visual Computing (ISVC)*, Crete, Greece, 2012.
- [40] A. Wesley, P. Lindner, and I. Pavlidis. Eustressed or distressed? combining physiology with observation in user studies. In *Proceedings of the 2012 ACM Conference on Human Factors in Computing Systems (CHI)*, 2012.
- [41] J. Whitehill, G. Littlewort, I. Pasel, M. Bartlett, and J. Movellan. Toward practical smile detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(11):2106–2111, 2009.

- [42] M. Yamaguchi, J. Wakasugi, and J. Sakakima. Evaluation of driver stress using biomarker in motor-vehicle driving simulator. In *Engineering in Medicine and Biology Society, IEEE*, 1834–1837, 2006.
- [43] T. Yamakoshi, K. Yamakoshi, S. Tanaka, M. Nogawa, M. Shibata, Y. Sawada, P. Rolfe, and Y. Hirose. A preliminary study on driver’s stress index using a new method based on differential skin temperature measurement. In *Engineering in Medicine and Biology Society, IEEE*, 722–725, 2007.
- [44] 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.
- [45] Y. Yoshitomi, N. Miyawaki, S. Tomita, and S. Kimura. Facial expression recognition using thermal image processing and neural network. *Robot and Human Communication*, 380–385, 1997.
- [46] Y. Zhou, P. Tsiamyrtzis, and I. Pavlidis. Tissue tracking in thermo-physiological imagery through spatio-temporal smoothing. *Proc. of the 12th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2009)*, 5762:1092–1099, 2009.