VIDEO BASED DETECTION OF DRIVER FATIGUE

by
ESRA VURAL

**Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy**

**Sabanci University
Spring 2009**

VIDEO BASED DETECTION OF DRIVER FATIGUE

APPROVED BY

Assist. Prof. Dr. Mujdat CETIN (Thesis Supervisor)	.....
Prof. Dr. Aytul ERCIL (Thesis Co-Advisor)	.....
Prof. Dr. Javier MOVELLAN	.....
Prof. Dr. Marian Stewart BARTLETT	.....
Assist. Prof. Dr. Hakan ERDOGAN	.....
Assist. Prof. Dr. Selim BALCISOY	.....

DATE OF APPROVAL: .....

To the memory of Erdal Inonu

Acknowledgments

There are many people who have contributed to this work and have supported me throughout this journey. Thus my sincere gratitude goes to my advisers, mentors, all my friends and my family for their love, support, and patience over the last few years.

I am grateful to my supervisor Mujdat Cetin for his helpful discussions, motivating suggestions and guidance throughout the thesis. I would like to also thank him for being very patient, supportive and helpful throughout my graduate years. I would like to express my special thanks to my co-adviser Aytul Ercil for her support and guidance. I would like to also thank her for initiating Drive-Safe project and supporting me throughout this process. I am grateful to her for accepting me to the friendly and encouraging environment of Computer Vision and Pattern Analysis Laboratory, Sabanci University.

I would like to express my deep gratitude to my mentor Javier Movellan for his guidance, suggestions, invaluable encouragement and generosity throughout the development of the thesis. I owe acknowledgement to him for welcoming me to the Machine Perception Laboratory, University of California San Diego and making me feel at home. I am very grateful to him for being very generous of his time and energy for the project regardless of his very busy schedule. His enthusiasm, positiveness and great ideas made this journey a fascinating experience.

I would like to express my special thanks to my mentor Marian Stewart Bartlett for her kind suggestions, guidance and brilliant ideas in developing the thesis. I owe gratitude to her for guiding me in writing publications and for showing infinite patience in correcting many of my mistakes throughout the process. This work uses the output from the Computer Expression Recognition Toolbox (CERT) which is developed by Machine Perception Lab researchers as a result of many years of research and hard work. I would like

to especially thank her and all the colleagues in Machine Perception Lab for building CERT and making this work possible.

Many thanks to my committee members Hakan Erdogan and Selim Balcisoy for reviewing the thesis and providing very useful feedback.

I am grateful to my friends from Computer Vision and Pattern Analysis Laboratory for being very supportive. Many thanks to Rahmi Ficici and Gulbin Akgun for their support and help. I would like to thank my friends Diana Florentina Soldea, Soldea Octavian, Ozben Onhon, Serhan Cosar, Ozge Batu for their kindness.

I am grateful to my colleagues from Machine Perception Laboratory for being so helpful and nice. I would like to thank Gwen Littlewort for her effort in improving CERT and her great discussions and friendly approach. I warmly thank Tingfan Wu for his valuable advice and friendly help. His extensive discussions around my work and interesting explorations have been very helpful for this study. Many thanks to Luis Palacious for his great help in system administration and patient approach to my questions. I would like to thank Andrew Salamon for making CERT a reality. I would also like to thank Nick Butko, Paul Ruvolo and Jacob Whitehill for their helpful discussions. Finally I would like to thank Kelly Hudson in making my experience with administrative issues very smooth.

My biggest gratitude is to my family. I am grateful to my parents and sister for their infinite moral support and help throughout my life. I owe acknowledgment to them for their encouragement, and love throughout difficult times in my graduate years.

Abstract

This thesis addresses the problem of drowsy driver detection using computer vision techniques applied to the human face. Specifically we explore the possibility of discriminating drowsy from alert video segments using facial expressions automatically extracted from video. Several approaches were previously proposed for the detection and prediction of drowsiness. There has recently been increasing interest in computer vision approaches as it is a potentially promising approach due to its non-invasive nature for detecting drowsiness. Previous studies with vision based approaches detect driver drowsiness primarily by making pre-assumptions about the relevant behavior, focusing on blink rate, eye closure, and yawning. Here we employ machine learning to explore, understand and exploit actual human behavior during drowsiness episodes. We have collected two datasets including facial and head movement measures. Head motion is collected through an accelerometer for the first dataset (UYAN-1) and an automatic video based head pose detector for the second dataset (UYAN-2). We use outputs of the automatic classifiers of the facial action coding system (FACS) for detecting drowsiness. These facial actions include blinking and yawn motions, as well as a number of other facial movements. These measures are passed to a learning-based classifier based on multinomial logistic regression. In UYAN-1 the system is able to predict sleep and crash episodes during a driving computer game with 0.98 performance area under the receiver operator characteristic curve for across subjects tests. This is the highest prediction rate reported to date for detecting real drowsiness. Moreover, the analysis reveals new information about human facial behavior during drowsy driving. In UYAN-2 fine discrimination of drowsy states are also explored on a separate dataset. The degree to which individual facial action units can predict the difference between moderately drowsy to acutely drowsy is studied. Signal processing techniques and machine learning methods are employed to build a person independent acute drowsiness detection system. Temporal dynamics are captured using a bank of temporal filters. Individual action unit predictive power is explored with an MLR based classifier. Best performing five action units have been determined for a person independent system. The system is able to obtain

0.96 performance of area under the receiver operator characteristic curve for a more challenging dataset with the combined features of the best performing 5 action units. Moreover the analysis reveals new markers for different levels of drowsiness.

Keywords: Fatigue Detection, Driver Drowsiness Detection, Computer Vision, Automatic Facial Expression Recognition, Machine Learning, Multinomial Logistic Regression, Gabor Filters, Temporal Analysis, Iterative Feature Selection, Facial Action Coding System (FACS), Head Motion

Özet

Bu doktora tezinde yüze uygulanan bilgisayar görü teknikleri kullanılarak sürücüde uykululuğun sezimi problemi ele alınmıştır. Özellikle uykulu görüntü kesitlerinin uykusuz görüntü kesitlerinden yüz ifadeleri aracılığıyla ayrılabilirliği keşfedilmeye çalışılmıştır. Geçmişte uykululuğun sezimi ve tahmini icin ceşitli yaklaşımlar önerilmiştir. Uykulu sürücü seziminde bilgisayarla görü yaklaşımlarının umut vaad eden ve müdehaleci olmayan özellikleri son yıllarda bu yaklaşımlara ilgiyi arttırmaktadır. Bilgisayar görü yaklaşımıyla çalışan önceki calışmalar uykulu sürücü seziminde başlıca varsayımlar olan göz kırpma hızı, göz kapama, ve esneme gibi uygun davranışlara odaklanmaktadır. Burada makine öğrenme tekniklerini kullanarak uykululuk kesitlerinde gerçek insan davranışını araştırmayı, anlamayı ve kullanmayı hedeflemekteyiz. Bu çalışma icin yüz ölçümleri ve baş hareketleri ölçümlerini içeren iki veri kümesi toplanmıştır. Baş hareketi verileri ilk veri kümesinde bir ivmeölçer cihazi ile ikinci veri kümesinde ise otomatik görüntü tabanlı baş pozisyonu sezici yardımıyla toplanmıştır. Yüz hareket kodlama sistemi (FACS) otomatik sınıflandırıcılarının çıktıları uykulu sürücü seziminde kullanılmaktadır. Bu hareket birimleri göz kapama esneme ve de birkaç ek yüz hareketini barındırmaktadır. Bu ölçüler öğrenme tabanlı sınıflandırıcı olan Lojistik Bağlanım Sınıflandırıcılarına (MLR) geçirilmiştir. Sistem birinci veri kümesi icin bir bilgisayar sürüş simülasyonu kullanan deneklerin uykulu ve uykusuz kesitlerini kişi bağımsız testler icin ROC (Receiver Operating Characteristics) eğrisi altında kalan alan hesabında 0.98 başarı elde etmiştir. Bu uykululuğun seziminde en yüksek tahmin oranıdır. Ayrıca analiz uykululukta insan yüz davranışı icin yeni bilgiler ortaya koymaktadır. Uykulu hallerin ince ayrımı iki veri kümesinde araştırılmıştır. Bireysel yüz hareket birimlerinin ne derecede orta ve ileri dereceli uykululuk farkını tespit edebileceği çalışılmıştır. Sinyal işleme teknikleri ve makina öğrenme yontemleri kullanılarak kişi bağımsız ileri derecede uykululuk sezim sistemi kurulmustur. Zamandaki dinamik bilgi zamansal filtre bankası kullanılarak çıkarılmıştır. Bireysel hareket ünitelerinin tahmin gücü MLR tabanlı sınıflandırıcılar kullanılarak araştırılmıstır. En iyi performansı veren beş hareket birimi insan

bağımsız bir sistem icin belirlenmiştir. Sistem 5 hareket ünitesinin özniteliklerini birleştiren bir sınıflandırıcı için daha zorlu bir veri kümesinde ROC (Receiver Operating Characteristics) eğrisi altında kalan alan hesabında 0.96 başarı göstermektedir. Ayrıca analiz değişik seviyelerdeki uykululuk için yeni belirteçler ortaya koymaktadır.

Anahtar Sozcükler: Yorgunluğun Sezimi, Sürücüde Uykululuğun Sezimi, Bilgisayar Görü Sistemleri, Otomatik Yüz İfadeleri Tanıma Sistemi, Makina Öğrenmesi, Lojistik Bağlanım Sınıflandırıcıları, Gabor Filtreleri, Zamansal Analiz, Öznitelik Seçimi, Yüz Hareket Kodlama Sistemi, Baş Hareketleri

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Definition

This thesis addresses the problem of drowsy driver detection using computer vision techniques applied to the human face. Specifically we explore the possibility of discriminating drowsy from alert video segments using facial expressions automatically extracted from video. In order to objectively capture the richness and complexity of facial expressions, behavioral scientists have found it necessary to develop objective coding standards. The facial action coding system (FACS)[23] is the most widely used expression coding system in the behavioral sciences. A human coder decomposes facial expressions in terms of 46 component movements or action units which roughly correspond to the individual facial muscle movements. FACS provides an objective and comprehensive way to analyze all the different facial expressions that a human face can make into elementary components, analogous to decomposition of speech into phonemes. Because it is comprehensive, FACS has proven useful for discovering facial movements that are indicative of cognitive and affective states [22]. In this thesis facial expressions in a video segment are extracted using an automated facial expression recognition toolbox, called Computer Expression Recognition Toolbox (CERT) [10], that operates in real-time and is robust to the video conditions in real applications. CERT codes facial expressions in terms of 30 actions from the facial action coding system (FACS). CERT assigns a continuous value for each of the 30 action units it considers. These continuous values represent the estimated intensities (muscle activations) of the action units observed in

that frame.

In this thesis we use the CERT system to address several questions: First we investigate the hypothesis of whether or not automatically detected facial behaviour is a good source of information for detecting drowsiness. If so, our second goal is to investigate what aspects of the morphology and dynamics of facial expressions are indicative of drowsiness. Our third goal is to understand the possibilities and challenges of automatic drowsiness detection based on facial expression analysis and develop classification algorithms. Finally our fourth goal is to understand the facial expressions occurring at fine states of drowsiness such as moderate drowsiness and acute drowsiness.

1.2 Solution Approach

The approach we take to answer this problem is as follows.

(1) Data sets are collected from subjects showing spontaneous facial expressions during the state of fatigue.

(2) We analyze the degree to which individual facial action units can predict the difference between alert and drowsy or moderately drowsy and acutely drowsy

(3) Temporal dynamics are captured using a bank of temporal filters. How to extract the relevant feature set of filters for a person independent drowsiness detector is studied.

1.3 Significance of the Problem

The US National Highway Traffic Safety Administration (NHTSA) estimates that in the US alone approximately 100,000 crashes each year are caused primarily by driver drowsiness or fatigue [36][5]. According to statistics gathered by the federal government each year, at least 1500 people die and 40,000 people get injured in crashes related to sleepy, fatigued or drowsy drivers in the United States of America. These numbers are most likely an underestimate. Unless someone witnesses or survives the crash and can testify the driver's condition, it is difficult to determine if the driver fell asleep[5]. In a 2003 interview with 4010 drivers in the U.S.A., 37% of the drivers reported having nodded off while driving at some point in their lives and 29% of these drivers reported having experienced this problem within the past year [20][32]. Sim-

ilarly in a 2006 survey with 750 drivers in the province of Ontario, Canada, nearly 60% of the drivers admitted driving while drowsy or fatigued at least sometimes, and 15% reported falling asleep while driving during the past year[32] [55]. A questionnaire study participated by 154 truck drivers to assess the relationship between prior sleep, work and individual characteristics and drowsiness found out that prior sleep aspects contributed the most to sleepiness while driving [52]. The National Safety Traffic Board (NTSB) concluded that 52 % of 107 single-vehicle accidents involving heavy trucks were fatigue-related; in nearly 18 per cent of the cases, the driver admitted to falling asleep [1].

Tiredness and fatigue can often affect a person's driving ability long before he/she even notices that he/she is getting tired. Fatigue related crashes are often more severe than others because driver's reaction times are delayed or the drivers have failed to make any maneuvers to avoid a crash. The number of hours spent driving has a strong correlation to the number of fatigue-related accidents. Figure 1.1 displays the relationship between number of hours driven and the percent of crashes related to driver fatigue [4]. A study conducted by the Adelaide Centre for Sleep Research has shown that drivers who have been awake for 24 hours have an equivalent driving performance to a person who has a BAC (blood alcohol content) of 0.1 g/100ml, and is seven times more likely to have an accident[1]. In fact, NHTSA has concluded that drowsy driving is just as dangerous as drunk driving. Thus methods to automatically detect drowsiness may help save many lives and contribute to the well-being of the society.

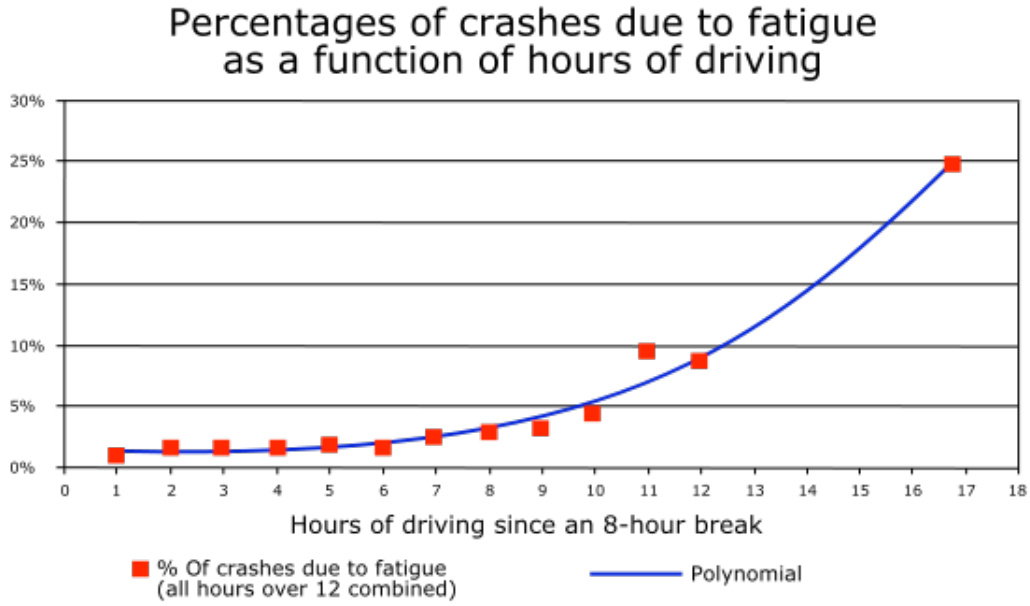

Figure 1.1: The figure displays the relationship between number of hours driven and the percent of crashes related to driver fatigue [4].

Current state of the art technologies focus on behavioral cues to detect drowsiness. Behavioral technologies detect drowsiness based on physiological signals or computer vision methods. Brain waves, heart rate and respiration rate are some of the physiological signals exploited for the detection of drowsiness[14][38][34]. Physiological signals usually require physical contact with the driver and may cause disturbance. Hence there has recently been increasing interest in computer vision as it is a prominent and a non-invasive approach for detecting drowsiness. Computer vision approaches use facial expressions to infer drowsiness[30][58]. Previous approaches to drowsiness detection primarily make pre-assumptions about the relevant behavior, focusing on blink rate, eye closure, and yawning [30] [48]. Here we employ machine learning methods to explore actual human behavior during drowsiness episodes. Computer vision based expression analysis systems can use several inputs ranging from low-level inputs such as raw pixels, to higher level inputs i.e facial action units or basic facial expressions to detect the facial appearance changes. For drowsiness detection since large sets of data from different subjects is not available, using higher levels of input such as action units helps to increase the performance of the system. FACS also

provides versatile representations of the face. FACS does not apply interpretive labels to expressions but rather a description of physical changes in the face. This enables studies of new relationships between facial movement and internal state, such as the facial signals of stress or drowsiness[9]. Developing technologies and methods to automatically recognize internal states, like drowsiness, from objective behavior, has a revolutionary effect in the brain and behavioral sciences. Moreover, the problem of automatic recognition of facial behavior from video is currently a recognized research area within the machine perception and computer vision communities[41][40].

This thesis contributes to understand how to build better vision machines with potential practical applications. It also helps us understand from a computational point of view the problems that the human visual system solves seamlessly.

.

1.4 Contributions

A common dataset of non-posed, spontaneous facial expressions during drowsiness is not available for the research community. Hence for this thesis we created our own spontaneous drowsiness dataset. Capturing spontaneous drowsiness behavior is a challenging and laborious task. We preferred to collect drowsiness data during midnight as it is of lesser chance to observe drowsiness during the day. A unique dataset of spontaneous facial expressions are collected from 20 subjects during driving in alert and drowsy conditions. Spontaneous facial expressions have not been studied for drowsiness until now and this is the first study that explores spontaneous facial expressions occurring during drowsiness to our knowledge. We analyzed what aspects of the morphology and dynamics of facial expressions are informative about drowsiness and to what degree. Machine learning methods are developed and evaluated for a person independent drowsiness detection system. Different classification and feature extraction methods are explored for a more accurate drowsiness detector. How to detect fine states of drowsiness like acute and moderate drowsiness is also explored in this thesis. Facial expressions informative about these two states are explored. Our analysis with this limited dataset discovered new expressions indicative of acute and moderate drowsiness states. We also obtained a better performing classifier by including features capturing temporal dynamics of facial expressions.

1.5 Outline

In Chapter 2 we describe prior work on fatigue detection and prediction technologies. We also introduce some of the methods employed for processing the signal, developing automatic classifiers, and evaluating performance : e.g. ROC, Adaboost, Multinomial Logistic Regression, Gabor Filters. In Chapter 3 we describe Study I that predicts sleep versus crash episodes from facial expressions of subjects performing a driving simulator task. We also describe some preliminary results obtained from head movement measures. In Chapter 4 we present the results for detecting fine states of drowsiness like acute drowsiness and moderate drowsiness. A new dataset, UYAN-2, has been collected for this study which consists of 11 subjects using the driving simulator while their faces are captured with a DV Camera and the brain dynamics and upper torso movements are measured using EEG and Motion Capture facilities respectively. The details about the experimental setup and the subject-wise differences in comparison with the UYAN-1 are also presented in Chapter 4. We discuss how different signal processing approaches and machine learning methods perform on generalization to novel subjects. The discriminative power of individual filters for predicting drowsiness is studied and how to select the prominent features is analyzed in the same chapter. Finally in Chapter 5 we present our conclusions together with some potential topics for future work.

Chapter 2

Background

2.1 Background on Fatigue Detection and Prediction Technologies

Dinges and Mallis [18] identified 4 different categories of fatigue detection technologies : (1) Fitness for Duty Technologies, (2) Ambulatory Alertness Prediction Technologies, (3) Vehicle-based Performance Technologies and (4) In-vehicle Online Operator Status technologies.

2.1.1 Fitness for Duty Technologies

The goal of fitness-for-duty technologies is to assess the vigilance or alertness capacity of an operator before a high risk type of work such as mining or driving is performed. Performance of the subject at a chosen task is used as a measure to detect existing fatigue impairment. Eye hand coordination [45] or driving simulator tasks are some of the previously used methods in detecting fatigue using this approach. This technology is potentially useful for measuring existing fatigue impairment [33]: an operator who fails the chosen test task lacks the vigilance for the work. Note that even if the operator passes the test, his/her state will change during the course of duty. The predictive validity, the task's predictive power of future fatigue, is still not well established[33]: it is not known how long an operator, that passes the test at a chosen task, will keep vigilant during work.

2.1.2 Ambulatory Alertness Prediction Technologies

The goal of ambulatory alertness prediction technologies is to predict operator alertness/performance at different times based on interactions of sleep, circadian rhythm, and related temporal antecedents of fatigue. Note that these technologies are different from our work as they do not assess fitness online as the work is performed. This technology predicts alertness using devices that monitor sources of fatigue, such as how much sleep an operator has obtained (via wrist activity monitor, defined below), and combine this information with mathematical models that predict performance and fatigue over future periods of time[33]. As an example to such a system US Army medical researchers have developed a mathematical model to predict human performance on the basis of prior sleep [11]. They integrated this model into a wrist-activity monitor based sleep and performance predictor system called "Sleep Watch". The Sleep Watch system includes a wrist-worn piezo electric chip activity monitor and recorder which will store up records of the wearer's activity and sleep obtained over several days. While this technology shows potential to predict fatigue in operators, more data and possible fine tuning of the models are needed before they can be fully accepted [33].

2.1.3 Vehicle-based Performance Technologies

Vehicle-based performance technologies place sensors on standard vehicle components, e.g., steering wheel, gas pedal, and analyzes the signals sent by these sensors to detect drowsiness [51]. Some of the previous studies use driver steering wheel movements and steering grip as an indicator of fatigue impairment. Microcorrections for steering are necessary for environmental factors and the reduction in number of microcorrections to steering indicate an impaired state [9]. Some car companies, Nissan[56] and Renault[7], adopted this technology however the main problem with steering wheel input is that it works in very limited situations [37]. Such monitors are too dependent on the geometric characteristics of the road (and to a lesser extent the kinetic characteristics of the vehicle), thus they can only function reliably on motorways [7].

Simple systems that purport to measure fatigue through vehicle-based performance are currently commercially available. However, their effectiveness in terms of reliability, sensitivity and validity is uncertain (i.e. formal validation tests either have not been undertaken or at least have not been

made available to the scientific community) [33].

A commercial product, AAlert (AA), is a flexible rubber device that uses motion combined with reaction time to determine whether or not the driver is in a drowsy state. The device vibrates when a driver is tired and should take a break from the wheel. If a driver, while driving, doesn't move his/her wrist for more than 15 seconds, a vibration is sent to the bracelet. To stop the vibration, the person needs to move his/her wrist. The slower the reaction to the vibration, the more likely it is that the driver is tired and should take a break from the wheel. The device communicates with an RFID tag positioned in the car and only starts detecting drowsiness when the driver is in the car. The picture of the device is shown in Figure 2.1.

Figure 2.1: AAlert wristband driver drowsiness detection device developed by Dan Ruffle. The device uses motion combined with reaction time to determine whether or not the driver is in a drowsy state.

2.1.4 In-vehicle, On-line, Operator Status Monitoring Technologies : Behavioral Studies using Physiological Signals

These techniques estimate fatigue based on physiological signals such as heart rate variability (HRV), pulse rate, breathing and Electroencephalography (EEG) [15][57] measures. Time series of heart beat pulse signal can be used to calculate the heart rate variability (HRV) – the variations of beat-to-beat intervals in the heart rate [6], and HRV has established differences between waking and sleep stages from previous psycho-physiological studies [24][57]. The frequency domain spectral analysis of HRV shows that typical HRV in human has three main frequency bands: high frequency band (HF) that lies in 0.15 – 0.4 Hz, low frequency band (LF) in 0.04 – 0.15 Hz, and very low frequency (VLF) in 0.0033 – 0.04 Hz [6] [57]. A number of psycho-physiological researches have found that the LF to HF power spectral density ratio (LF/HF ratio) decreases when a person changes from waking into drowsiness/sleep stage, while the HF power increases associated with this status change [24] [57].

EEG is the recording of electrical activity along the scalp produced by the firing of neurons within the brain. In clinical contexts, EEG refers to the brain's spontaneous electrical activity as recorded from multiple electrodes placed on the scalp. There are five major brain waves distinguished by their different frequency ranges. These frequency bands from low to high frequencies respectively are called alpha, theta, beta, delta and gamma. The alpha and beta waves lie between 8-12 Hz and 12-30 Hz respectively (Berger et al. 1929). Alpha waves tend to occur during relaxation or keeping the eyes closed. Beta is the dominant wave representing alertness, anxiety or active concentration. Gamma refers to the waves of above 30 Hz (Jasper and Andrews (1938)). Gamma waves are thought to represent binding of different populations of neurons together into a network for the purpose of carrying out a certain cognitive or motor function[3]. The delta waves designate all frequencies between 0-4 Hz (Walter et al, 1936). Theta waves have frequencies within the range of 4-7.5 Hz. Theta waves represent drowsiness in adults.

In the literature power spectrum of EEG brain waves is used as a measure to detect drowsiness [38]. It has been reported by researchers that as the alertness level decreases EEG power of the alpha and theta bands increases

[34]. Hence providing indicators of drowsiness. However using EEG as a measure of drowsiness has drawbacks in terms of practicality since it requires a person to wear an EEG cap while driving. Moreover motion related artifacts are still an unsolved research problem.

One important problem in EEG is that it is very easy to confuse artifact signals caused by the large muscles in the neck and jaw with the genuine delta response [49]. This is because the muscles are near the surface of the skin and produce large signals, whereas the signal that is of interest originates from deep within the brain and is severely attenuated in passing through the skull [49]. In general EEG recordings are extremely sensitive to motion artifacts. Motion related signals are actually 3 orders of magnitude larger than signals due to neural activity and this is still a big unsolved problem for EEG analysis.

2.1.5 In-vehicle, On-line, Operator Status Monitoring Technologies : Behavioral Studies using Computer Vision Systems

Computer vision is a prominent technology in monitoring the human behavior. The advantage of computer vision techniques is that they are non-invasive, and thus are more amenable to use by the general public. In recent years machine learning applications to computer vision had a revolutionary effect in building automatic behavior monitoring systems. The current technology provides us imperfect but reasonable tools to build computer vision systems that can detect and recognize the facial motion and appearance changes occurring during drowsiness [30] [58].

Most of the published research on computer vision approaches to detection of fatigue has focused on the analysis of blinks [53]. Percent closure (PERCLOS), which is the percentage of eyelid closure over the pupil over time and reflects slow eyelid closures ("droops") rather than blinks, is analyzed in many studies [16] [28]. Some of these studies used infrared cameras to estimate the PERCLOS measure [16]. It is worth pointing out that infrared technology for PERCLOS measurement works fairly well at night, but not very well in daylight, because ambient sunlight reflections make it impractical to obtain retinal reflections of infrared waves[33]. Other studies used the video frames for estimating the PERCLOS measure [50]. One example of such commercial products is the Driver State Sensor (DSS) device developed

by SeeingMachines [2]. DSS is a robust, automatic and nonintrusive sensor platform that uses cutting edge face tracking techniques to deliver information on operator fatigue and operator distraction. In cars DSS is located on the dashboard and it uses the eyelid opening and Percent Closure (PERCLOS), which is the the percentage of eyelid closure over the pupil over time, as a measure to derive the drowsiness state. A snapshot of the system is displayed in Figure 2.2.

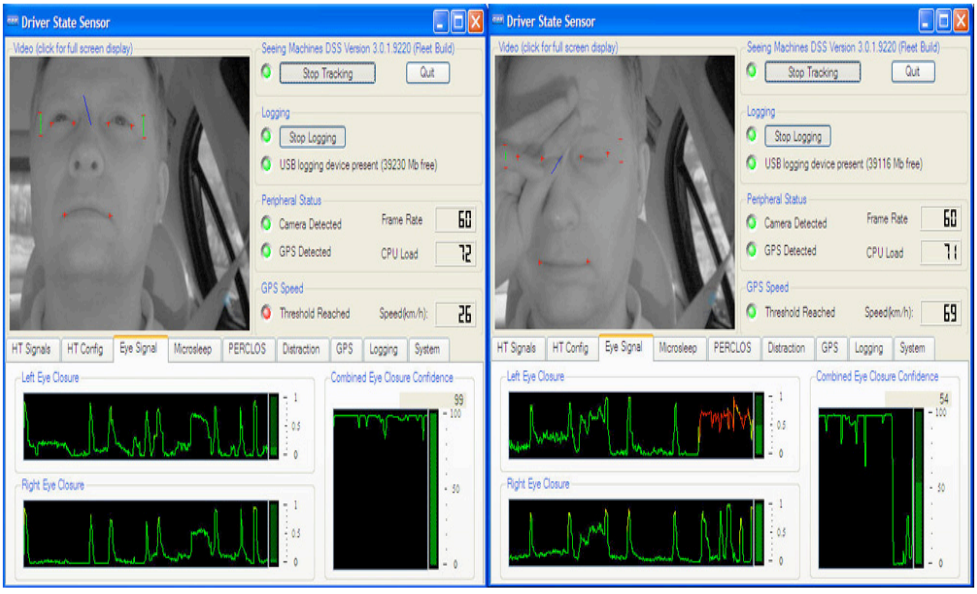

Figure 2.2: Driver State Sensor (DSS) device developed by SeeingMachines. DSS uses eyelid opening as a measure to infer the drowsiness state.

Head nodding [48] and eye closure[50][48] have been studied as indicators of fatigue but there are other facial expressions and not much is known about facial behavior during the state of fatigue. Until now tools have not been available to study these expressions and manual coding of facial expressions is extremely difficult.

Computer vision has advanced to the point that scientists are now beginning to apply automatic facial expression recognition systems to important research questions in behavioral science: Lie detection, differentiating real pain from faked pain, understanding emotions such as happiness, surprise etc are all possible applications of facial expression recognition systems

[40][8][39].

Gu & Ji [31] presented one of the first fatigue studies that incorporated certain facial expressions other than blinks. Their study fed action unit information as an input to a dynamic Bayesian network. The network was trained on subjects posing a state of fatigue. The video segments were classified into three stages: inattention, yawn, or falling asleep. For predicting falling-asleep, head nods, blinks, nose wrinkles and eyelid tighteners were used. While this was a pioneering study, its value is limited by the use of posed expressions. Spontaneous expressions have a different brain substrate than posed expressions. They also typically differ in dynamics and morphology in that different action unit combinations occur for posed and spontaneous expressions. In addition, as we have observed during the work, it is very difficult for people to guess the expressions they would actually make when drowsy or fatigued. Using spontaneous behavior for developing and testing computer vision systems is highly important given the fact that the spontaneous and posed expressions have very different brain substrate, morphology and dynamics [22]

Previous approaches to drowsiness detection primarily make pre-assumptions about the relevant behavior, focusing on blink rate, eye closure, and yawning. Here we employ machine learning methods to data-mine actual human behavior during drowsiness episodes. The objective of this thesis is to investigate whether there are facial expression configurations or facial expression dynamics that are predictors of fatigue and to explain methods for analyzing automatic facial expression signals to effectively extract this information. In this thesis, facial motion was analyzed automatically from video using a fully automated facial expression analysis system based on the Facial Action Coding System (FACS) [10]. In addition to the output of the automatic FACS recognition system we also collected head motion data either through an accelerometer placed on the subject's head, or a computer vision-based head pose tracking system, as well as steering wheel data.

Computer vision based expression analysis systems can use several inputs ranging from low-level inputs such as raw pixels to higher level inputs i.e facial action units or basic facial expressions to detect the facial appearance changes. For special purpose systems designed to detect only a particular expression or a particular state it may be beneficial to avoid intermediate representations such as FACS, provide a large database is available. For example Whitehill et. al presents a smile analyzer system [54] that can discern smiles from non-smiles by training the system with a set of 20,000 different subject's

face data. The system is able to detect smile versus non-smiles with a high performance. On the other hand when the dataset is relatively small, it may be beneficial to use systems that provided a rich intermediate representation, such as FACS codes. In addition the use of a FACS based representation has the advantage of being anatomically interpretable. For drowsiness detection large sets of data from different subjects is not available as capturing spontaneous drowsiness behavior is a challenging and laborious task. Hence using higher levels of input such as action units might increase the performance of the system. FACS also provides versatile representations of the face. Thus for all the above reasons action unit outputs from CERT[10], which is a user independent fully automatic system for real time recognition of facial actions from the Facial Action Coding System (FACS), is used as an input to the automated drowsiness detector.

2.1.5.1 Facial Action Coding System

The facial action coding system (FACS) [23] is one of the most widely used methods for coding facial expressions in the behavioral sciences. The system describes facial expressions in terms of 46 component movements, which roughly correspond to the individual facial muscle movements. An example is shown in Figure 2.3. FACS provides an objective and comprehensive way to analyze expressions into elementary components, analogous to decomposition of speech into phonemes. Because it is comprehensive, FACS has proven useful for discovering facial movements that are indicative of cognitive and affective states. See Ekman and Rosenberg (2005) [22] for a review of facial expression studies using FACS. The primary limitation to the widespread use of FACS is the time required to code. FACS was developed for coding by hand, using human experts. It takes over 100 hours of training to become proficient in FACS, and it takes approximately 2 hours for human experts to code each minute of video. Researchers have been developing methods for fully automating the facial action coding system [10][19]. In this thesis we apply a computer vision system trained to automatically detect FACS to data mine facial behavior under driver fatigue.

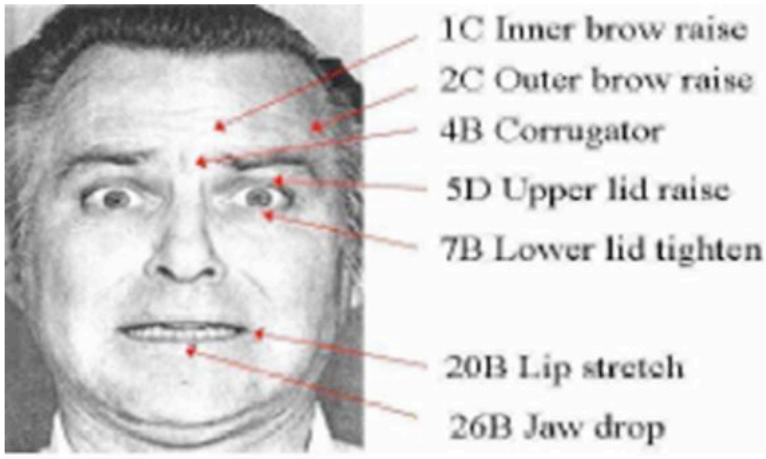

Figure 2.3: Example facial action decomposition from the Facial Action Coding System [23].

2.1.5.2 Spontaneous Expressions

The machine learning system presented in this thesis was trained on spontaneous facial expressions. The importance of using spontaneous behavior for developing and testing computer vision systems becomes apparent when we examine the neurological substrate for facial expression. There are two distinct neural pathways that mediate facial expressions, each one originating in a different area of the brain. Volitional facial movements originate in the cortical motor strip, whereas spontaneous facial expressions originate in the sub-cortical areas of the brain (see [47], for a review). These two pathways have different patterns of innervation on the face, with the cortical system tending to give stronger innervation to certain muscles primarily in the lower face, while the sub-cortical system tends to more strongly innervate muscles primarily in the upper face [42]. The facial expressions mediated by these two pathways have differences both in which facial muscles are moved and in their

dynamics [21][22]. Subcortically initiated facial expressions (the spontaneous group) are characterized by synchronized, smooth, symmetrical, consistent, and reflex-like facial muscle movements whereas cortically initiated facial expressions (posed expressions) are subject to volitional real-time control and tend to be less smooth, with more variable dynamics [47]. Given the two different neural pathways for facial expressions, it is reasonable to expect to find differences between genuine and posed expressions of states such as pain or drowsiness. Moreover, it is crucial that the computer vision model for detecting states such as genuine pain or driver drowsiness be based on machine learning of expression samples when the subject is actually experiencing the state in question. It is very difficult for people to imagine and produce the expressions they would actually make when they are tired or drowsy.

2.1.5.3 The Computer Expression Recognition Toolbox (CERT)

This study uses the output of CERT as an intermediate representation to study fatigue and drowsiness. CERT, developed by researchers at Machine Perception Laboratory UCSD [10], is a user independent fully automatic system for real time recognition of facial actions from the Facial Action Coding System (FACS). The system automatically detects frontal faces in the video stream and codes each frame with respect to 20 Action units. An overview of the system can be found in Figure 2.4.

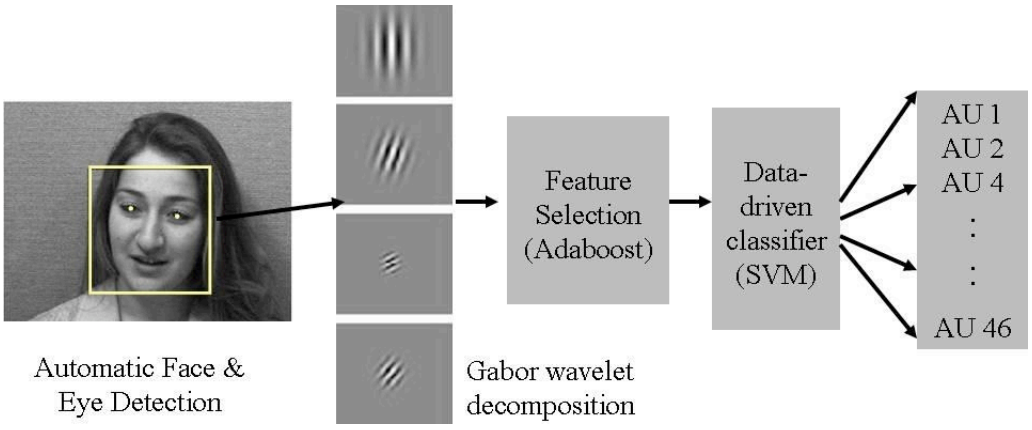

Figure 2.4: Overview of fully automated facial action coding system

Real Time Face and Feature Detection CERT uses a real-time face detection system that uses boosting techniques in a generative framework (Fasel et al.) and extends work by Viola and Jones (2001). Enhancements to Viola and Jones include employing Gentleboost instead of AdaBoost, smart feature search, and a novel cascade training procedure, combined in a generative framework. Source code for the face detector is freely available at http://kolmogorov.sourceforge.net. Accuracy on the CMU-MIT dataset, a standard public data set for benchmarking frontal face detection systems (Schneiderman & Kanade, 1998), is 90% detections and 1/million false alarms, which is state-of-the-art accuracy. The CMU test set has unconstrained lighting and background. With controlled lighting and background, such as the facial expression data employed here, detection accuracy is much higher. The system presently operates at 24 frames/second on a 3 GHz Pentium IV for 320x240 images. The automatically located faces are rescaled to 96x96 pixels. The typical distance between the centers of the eyes is roughly 48 pixels. Automatic eye detection [26](Fasel et al., 2005) is employed to align the eyes in each image. In the CERT system the images are then passed to a filtering stage through a bank of 72 Gabor filters 8 orientations and 9 spatial frequencies (2:32 pixels per cycle at 1/2 octave steps). Output magnitudes are then passed to the action unit classifiers.

Automatic Facial Action Classification The AU classifiers in the CERT system were trained using three posed datasets and one dataset of spontaneous expressions. The facial expressions in each dataset were FACS coded by certified FACS coders. The first posed dataset was the Cohn- Kanade DFAT-504 dataset [35] (Kanade, Cohn & Tian, 2000). This dataset consists of 100 university students who were instructed by an experimenter to perform a series of 23 facial displays, including expressions of seven basic emotions. The second posed dataset consisted of directed facial actions from 24 subjects collected by Ekman and Hager. Subjects were instructed by a FACS expert on the display of individual facial actions and action combinations, and they practiced with a mirror. The resulting video was verified for AU content by two certified FACS coders. The third posed dataset consisted of a subset of 50 videos from 20 subjects from the MMI database (Pantic et al., 2005). The spontaneous expression dataset consisted of the FACS-101 dataset collected by Mark Frank (Bartlett et. al. 2006). 33 subjects underwent an interview about political opinions on which they felt strongly. Two

minutes of each subject were FACS coded. The total training set consisted of posed databases and 3000 from the spontaneous set.

Twenty linear Support Vector Machines were trained for each of 20 facial actions. Separate binary classifiers, one for each action, were trained to detect the presence of the action in a one versus all manner. Positive examples consisted of the apex frame for the target AU. Negative examples consisted of all apex frames that did not contain the target AU plus neutral images obtained from the first frame of each sequence. Eighteen of the detectors were for individual action units, and two of the detectors were for specific brow region combinations: fear brow (1+2+4) and distress brow (1 alone or 1+4). All other detectors were trained to detect the presence of the target action regardless of co-occurring actions. A list is shown in Table 1A. Thirteen additional AU's were trained for the Driver Fatigue Study. These are shown in Table 1B.

In general the output of a classifier is thought as discrete, rather than real-valued. Here the output of the system is the distance to the separating hyperplane of an SVM classifier. The distance is a real number representing the output of an AU classifier. Previous work showed that the distance to the separating hyperplane (the margin) contained information about action unit intensity [10](e.g. Bartlett et al., 2006). A vector of real-valued numbers is output by the system each number representing the output of an AU classifier.

In this thesis we will be using the output of CERT as our basic representation of facial behavior. Classifiers will be built on top of the CERT output to investigate which facial expressions and facial expression dynamics are informative of driver drowsiness.

2.2 Background on Machine Learning Techniques

Here we will give a brief introduction to machine learning concepts that have been used for the context of this thesis.

2.2.1 System Evaluation : Receiver operating characteristic (ROC)

In signal detection theory, a receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot of the sensitivity vs. (1 - specificity) for a binary classifier system as its discrimination threshold is varied [29]. In

this thesis, area under the ROC curve (A') used to assess performance most frequently rather than overall percent correct, since percent correct can be an unreliable measure of performance, as it depends on the proportion of targets to non-targets, and also on the decision threshold. Notice that A' will refer to the area under the ROC curve for the context of the thesis. Similarly, other statistics such as true positive and false positive rates depend on decision threshold, which can complicate comparisons across systems. The ROC curve is obtained by plotting true positives against false positives as the decision threshold shifts from 0 to 100% detections. The area under the ROC (A') ranges from 0.5 (chance) to 1 (perfect discrimination). Figure 2.5 shows the true positive rate (TPR) and false positive rate (FPR) for positive and negative instances for a certain threshold. The figure also shows a plot for the ROC curve. A' is equivalent to the theoretical maximum percent correct achievable with the information provided by the system when using a 2-Alternative Forced Choice testing and paradigm [13]. 2-Alternative Forced Choice (abbreviated to 2AFC) testing is a psycho-physical method for eliciting responses from a person about his or her experiences of a stimulus. For example, a researcher might want to decide on every trial which of two locations A or B contains the stimulus [25]. On any trial, the stimulus might be presented at location A or location B. The subject then has to choose whether the stimulus appeared in location A or B. The subject is allowed only to choose two of these locations; he or she is not allowed to say "Not sure", or "I don't know". Thus the subject's choice is forced in this sense. The area below an ROC curve corresponds to the fraction of correct decisions in a two-alternative forced choice task. For this thesis we will use the term "A"' to refer to the area under the response operating curve.

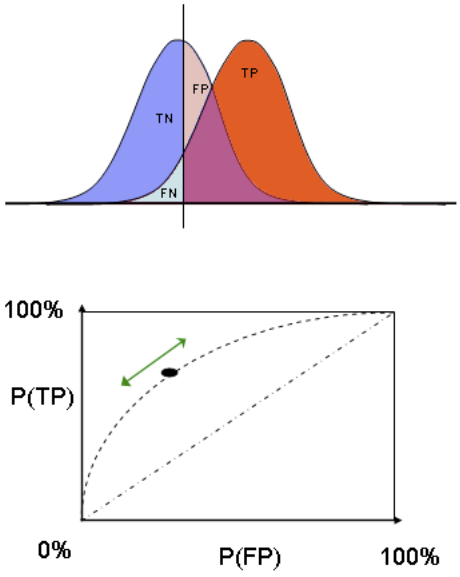

Figure 2.5: The true positive rate (TPR) and false positive rate (FPR) of positive and negative instances for a certain threshold. ROC plot is obtained by plotting true positives against false positives as the decision threshold shifts from 0 to 100% detections.

2.2.2 Signal Processing

2.2.2.1 Gabor Filter

A Gabor filter is a linear filter whose impulse response is defined by a complex sinusoid multiplied by a Gaussian function [43]. In this thesis, we use two different types for Gabor Filters. Spatial Gabor filters are used by the CERT system to extract features from images to detect facial action units. A bank of 72 Gabor filters 8 orientations and 9 spatial frequencies (2:32 pixels per cycle at 1/2 octave steps) are employed for filtering face images. Output magnitudes are then passed to the action unit classifiers. In this thesis we are employing temporal Gabor filters to analyze the temporal patterns of action units. A set of complex Gabor [17] filters is used for analyzing temporal patterns of action unit signals. Gabor filters can serve as excellent band-pass

filters for uni-dimensional signals (e.g., speech). Uni-dimensional temporal Gabor Filters are employed for capturing temporal properties of the action unit signals for detecting drowsiness. A complex Gabor filter is defined as the product of a Gaussian kernel times a complex sinusoid, i.e.

$$g(t) = ke^{j\theta}w(at)s(t) \tag{2.1}$$

where

$$w(t) = e^{-\pi t^2} \tag{2.2}$$

$$s(t) = e^{j2\pi f_o t} \tag{2.3}$$

$$e^{j\theta}s(t)e^{j2\pi f_o t+\theta} = (sin(2\pi f_o t + \theta), jcos(2\pi f_o t + \theta)) \tag{2.4}$$

Here a, k, θ, f_o are filter parameters that correspond to a bandwidth, amplitude constant, phase and peak frequency respectively. We can think of the complex Gabor filter as two out-of-phase filters conveniently allocated in the real and complex part of a complex function, the real part holds the filter in equation 5[43].

$$g_r(t) = w(t)sin(2\pi f_o t + \theta) \tag{2.5}$$

and the imaginary part holds the filter

$$g_i(t) = w(t)cos(2\pi f_o t + \theta) \tag{2.6}$$

Frequency Response

Frequency response is obtained by taking the Fourier Transform,

$$\hat{g}(f) = ke^{j\theta} \int_{-\infty}^{\infty} e^{-j2\pi ft}w(at)s(t)\ dt = ke^{j\theta} \int_{-\infty}^{\infty} e^{-j2\pi(f-f_o)t}w(at)dt \tag{2.7}$$

$$= \frac{k}{a}e^{j\theta}\hat{w}(\frac{f-f_o}{a}) \tag{2.8}$$

where

$$\hat{w}(f) = F\{w(t)\} = e^{-\pi f^2} \tag{2.9}$$

Gabor Energy Filters

The real and imaginary components of a complex Gabor filter are phase sensitive, i.e., as a consequence their response to a sinusoid is another sinusoid [43]. By getting the magnitude of the output (square root of the sum of squared real and imaginary outputs) we can get a response that is phase insensitive and thus gives unmodulated positive response to a target sinusoid input . In some cases it is useful to compute the overall output of the two out-of-phase filters. One common way of doing so is to add the squared output (the energy) of each filter, equivalently we can get the magnitude. This corresponds to the magnitude (more precisely the squared magnitude) of the complex Gabor filter output. In the frequency domain, the magnitude of the response to a particular frequency is simply the magnitude of the complex Fourier transform, i.e.

$$\|g(f)\| = \frac{k}{a}\hat{w}(\frac{f - f_o}{a}) \tag{2.10}$$

Note this is a Gaussian function centered at f_o and with width proportional to a.

Bandwidth and Peak Response

Thus the peak filter response is at f_o. To get the half-magnitude bandwidth Δ_f note

$$\hat{w}(\frac{f - f_o}{a}) = e^{-\pi \frac{f-f_o}{a^2}} = 0.5 \tag{2.11}$$

Thus the half peak magnitude is achieved for

$$f - f_o \pm \sqrt{a^2 \log 2\pi} = 0.4697a \approx 0.5a \tag{2.12}$$

Thus the half-magnitude bandwidth is $(2)(0.4697)a$ which is approximately equal to a. Thus a can be interpreted as the half-magnitude filter bandwidth.

2.2.3 Adaboost

AdaBoost calls a weak classifier repeatedly in a series of rounds . For each call a distribution of weights is updated such that it indicates the importance of

examples in the data set for the classification. On each round, the weights of each incorrectly classified example are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focuses more on those examples. It is shown that Adaboost can be interpreted as a method of sequential maximum likelihood estimation [27].

In this thesis AdaBoost is used only for within subject drowsiness prediction of the UYAN-1 dataset. For this study AdaBoost selected the facial action detector among a set of 30 Facial Action units that minimized the training error. Since multinomial logistic regression (MLR) obtained better performance MLR was employed for the rest of the thesis.

2.2.4 Support Vector Machines (SVM)

Support vector machines (SVMs) [12] are supervised learning methods used for classification and regression. Viewing input data as two sets of vectors in an n-dimensional space, an SVM will construct a separating hyperplane in that space, one which maximizes the margin between the two data sets. To calculate the margin, two parallel hyperplanes are constructed, one on each side of the separating hyperplane, that maximixes the minimum distance from the hyperplane to the closest training point. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the neighboring datapoints of both classes, since in general the larger the margin the lower the generalization error of the classifier. In this thesis Support Vector Machines are used by the CERT system as a classifier for determining the action unit intensity value. In general the output of a classifier is thought as discrete, rather than real-valued. Here the output of the system is the distance to the separating hyperplane of an SVM classifier. The distance is a real number representing the output of an AU classifier.

2.2.5 Multinomial Logistic Regression (MLR)

In statistics, logistic regression is used for prediction of the probability of occurrence of an event by fitting data to a logistic curve. Multinomial logistic regression (MLR) is an extension of logistic regression with two or more classes. Our goal is to train a well defined model based on examples of input-output pairs. For this thesis we use MLR with two classes (drowsy and alert) of dependents. The inputs to MLR are n-dimensional vectors and the outputs are 2-dimensional vectors representing drowsy and alert classes.

The training samples consist of m input output pairs. We organize the example inputs as an $m \times n$ matrix $\boldsymbol{x}$. The corresponding example outputs are organized as a $2 \times m$ matrix $\boldsymbol{y}$. The rows in $\boldsymbol{y}$ matrix add up to 1. For example for a given training sample the first row may have the value of 0 and the second row has the value of 1 for a drowsy sample or vice verse for an alert sample. The MLR makes predictions $h(\hat{\boldsymbol{y}})$ where h is defined in equation 2.15 and $\hat{\boldsymbol{y}}$ is a linear transformation of the data $\hat{\boldsymbol{y}}= \boldsymbol{\theta x}$, and $\boldsymbol{\theta}$ is a $2 \times n$ weight matrix.

The optimality of $\hat{\boldsymbol{y}}$, and thus of θ, is evaluated using the following criterion in L2 regularization norm [44]. L2 imposes a Gaussian prior over the weights and forces the weights to be small.

$$\Phi(\theta) = -\sum_{j=1}^{m} \rho(\boldsymbol{y}_j, \widehat{\boldsymbol{y}}_j) + \frac{\alpha}{2} \sum_{k=1}^{2} \boldsymbol{\theta}_k^T \boldsymbol{\theta}_k \tag{2.13}$$

Informally in the above formula, the first term can be seen as a negative log-likelihood function, capturing the degree of match between the data and the model, the second term can be interpreted as a negative log prior over $\boldsymbol{\theta}$ [44].

$$\rho(\boldsymbol{y}_j, \widehat{\boldsymbol{y}}_j) = \sum_{k=1}^{2} y_{jk} \log h_k(\hat{\boldsymbol{y}}_j) \tag{2.14}$$

where

$$h_k(\hat{y}_j) = \frac{e^{\hat{y}_{jk}}}{\sum_{i=1}^{2} e^{\hat{y}_{ji}}} \tag{2.15}$$

Newton Raphson algorithm is employed to minimize Φ [44]. There are many possible solutions to $\boldsymbol{\theta}$, we choose the one for which the last row is all zeros. For this thesis MLR training algorithm was used with different L2 regularization parameters (weight decay parameter). Once the model is trained and the weight vector is found the weighted data, $\widehat{\boldsymbol{y}}_{j1} = \sum_{l=1}^{n} \theta_{1l} x_{jl}$, is used as a measure to estimate the area under the ROC curve A' for the two classes.

Chapter 3

Study I : Detecting Drowsiness

The goal in this study is to investigate whether or not automatically detected facial behaviour is a reliable source of information for detecting drowsiness. We employ machine learning methods to datamine actual human behavior during drowsiness episodes. Automatic classifiers for 30 facial actions from the Facial Action Coding system were passed to classifiers such as Adaboost and Multinomial Logistic Regression(MLR). A block diagram of the system is shown in Figure 3.1. The system was able to predict sleep and crash episodes during a driving computer game with 98% accuracy across subjects. Moreover, the analysis revealed new information about human facial behavior during drowsy driving.

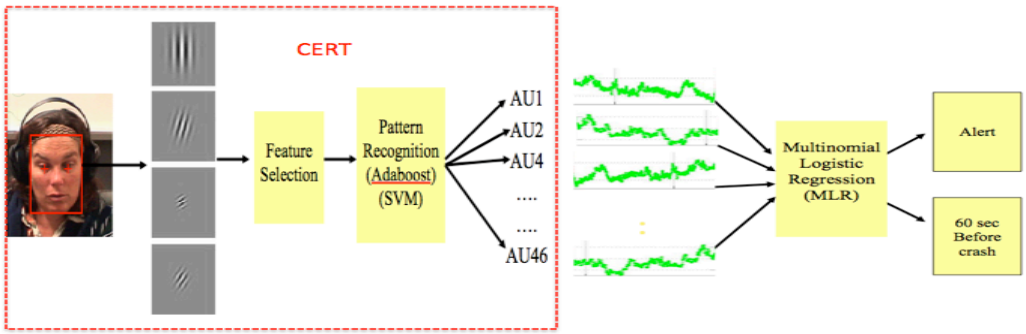

Figure 3.1: Outline of the Fatigue Detection System

3.1 UYAN-1 Dataset

Subjects were asked to drive a virtual car simulator. The simulator displayed the driver's view of a car through a computer terminal. The interface with the simulator was a steering wheel [1] and an open source multi-platform video game[2] (See Figure 3.2). The windows version of the video game was maintained such that at random times, a wind effect was applied that dragged the car to the right or left, forcing the subject to correct the position of the car. This type of manipulation had been found in the past to increase fatigue [46]. Driving speed was held constant. Four subjects performed the driving task over a three hour period beginning at midnight. During this time subjects fell asleep multiple times thus crashing their vehicles. Episodes in which the car left the road (crash) were recorded. Video of the subjects face was recorded using a DV camera for the entire 3 hour session. Subject data was partitioned into drowsy (non-alert) and alert states as follows. The one minute preceding a sleep episode or a crash was identified as a non-alert state. There was a mean of 24 non-alert episodes per subject with a minimum of 9 and a maximum of 35. Fourteen alert segments for each subject were collected from the first 20 minutes of the driving task.[3]

3.2 Head movement measures

Head movement was measured using an accelerometer that has three one dimensional accelerometers mounted at right angles measuring accelerations in the range of 5g to +5g where g represents earth gravitational force. A preliminary analysis of the correlation between head movement measure and the steering signal is employed.

3.3 Facial Action Classifiers

In this chapter we investigate whether there are action units that are predictive of the levels of drowsiness observed prior to the subjects falling sleep.

[1] ThrustMaster steering wheel

[2] Torcs

[3] Several of the drivers became drowsy very quickly which prevented extraction of more alert segments.

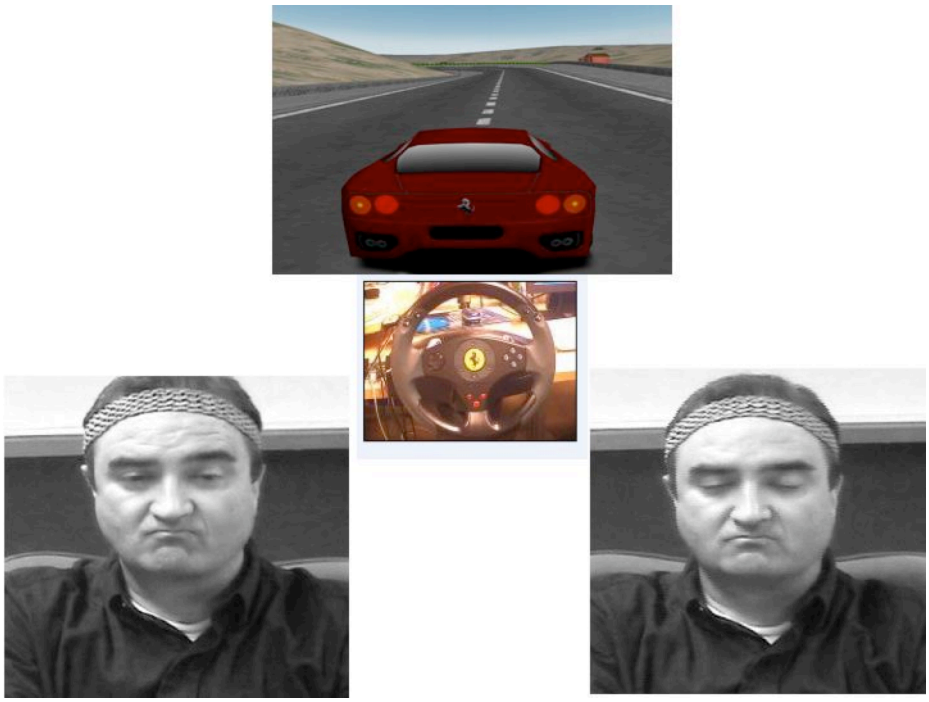

Figure 3.2: Driving simulation task

In Chapter 2 a computer expression recognition toolbox, named CERT, is presented for fully automated detection of facial actions from the facial action coding system [10]. Previously for 20 facial action units, a mean of 93% correct detection under controlled posed conditions, and 75% correct for less controlled spontaneous expressions with head movements and speech was reported for CERT.

For this study an improved version of CERT is used which was retrained on a larger dataset of spontaneous as well as posed examples. In addition, the system was trained to detect an additional 11 facial actions for a total of 31 (See Table 3.1). The facial action set includes blink (action unit 45), as well as facial actions involved in yawning (action units 26 and 27). The selection of this set of 31 out of 46 total facial actions was based on the availability of labeled training data. Figure 3.3 shows sample facial actions from the Facial Action Coding System incorporated in CERT.

Table 3.1: Full set of action units used for predicting drowsiness in Study I

AU	Name
1	Inner Brow Raise
2	Outer Brow Raise
4	Brow Lowerer
5	Upper Lid Raise
6	Cheek Raise
7	Lids Tight
8	Lip Toward
9	Nose Wrinkle
10	Upper Lip Raiser
11	Nasolabial Furrow Deepener
12	Lip Corner Puller
13	Sharp Lip Puller
14	Dimpler
15	Lip Corner Depressor
16	Lower Lip Depress
17	Chin Raise
18	Lip Pucker
19	Tongue show
20	Lip Stretch
22	Lip Funneller
23	Lip Tightener
24	Lip Presser
25	Lips Part
26	Jaw Drop
27	Mouth Stretch
28	Lips Suck
30	Jaw Sideways
32	Bite
38	Nostril Dilate
39	Nostril Compress
45	Eye Closure

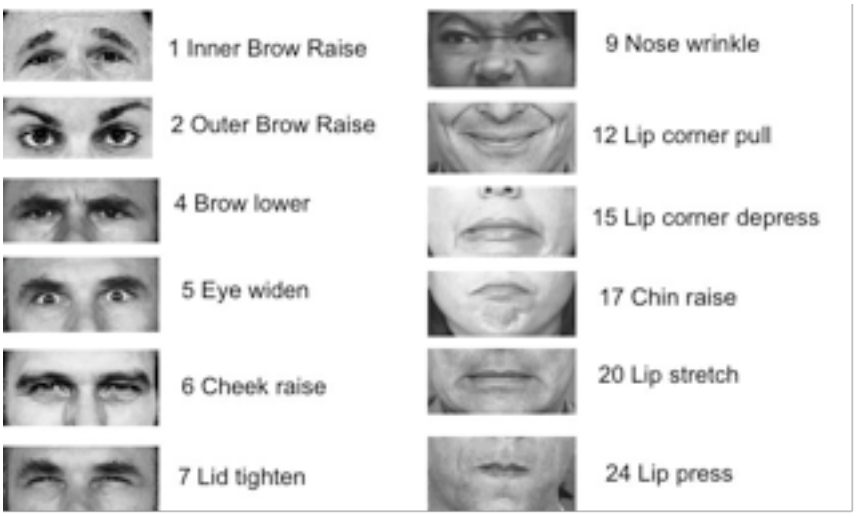

Figure 3.3: An Improved version of CERT is used for this study. The figure displays the sample facial actions from the Facial Action Coding System incorporated in CERT

3.4 Facial action signals

Our initial analysis focused on drowsiness prediction within-subjects. The output of the facial action detector consisted of a continuous value for each frame which was the distance to the separating hyperplane, i.e., the margin. Histograms for two of the action units in alert and non-alert states for two different subjects are shown in Figure 3.4. ROC curve was obtained by plotting false positive versus true positive rate for the intensity of an action unit for a subject. The area under the ROC (A') was computed for the outputs of each facial action detector to see to what degree the alert and non-alert output distributions were separated for all possible thresholds. The A' measure is derived from signal detection theory and characterizes the discriminative capacity of the signal, independent of decision threshold. Area under the ROC A', as discussed in the previous chapter, can be interpreted as equivalent to the theoretical maximum percent correct achievable with the information provided by the system when using a 2-Alternative Forced Choice testing paradigm. Table 3.2 shows the actions with the highest A'

for each subject. As expected, the blink/eye closure measure was overall the most discriminative for most subjects. However note that for Subject 2, the outer brow raise (Action Unit 2) was the most discriminative.

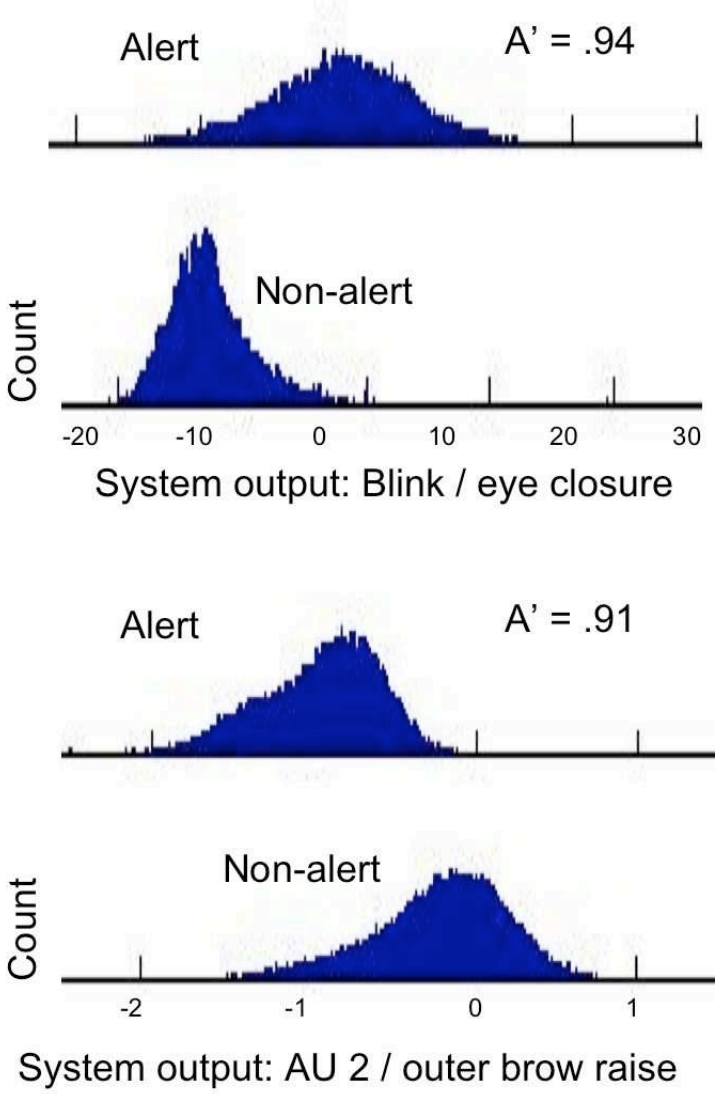

Figure 3.4: Histograms for blink and Action Unit 2 in alert and non-alert states. A' is area under the ROC.

Table 3.2: The top 5 most discriminant action units for discriminating alert from non-alert states for each of the four subjects. A' is area under the ROC curve.

	AU	Name	A'
Subj1	45	Blink	.94
	17	Chin Raise	.85
	30	Jaw sideways	.84
	7	Lid tighten	.81
	39	Nostril compress	.79
Subj2	2	Outer brow raise	.91
	45	Blink	.80
	17	Chin Raise	.76
	15	Lip corner depress	.76
	11	Nasolabial furrow	.76
Subj3	45	Blink	.86
	9	Nose wrinkle	.78
	25	Lips part	.78
	1	Inner brow raise	.74
	20	Lip stretch	.73
Subj4	45	Blink	.90
	4	Brow lower	.81
	15	Lip corner depress	.81
	7	Lid tighten	.80
	39	Nostril Compress	.74

3.5 Drowsiness prediction

The facial action outputs were passed to a classifier for predicting drowsiness based on the automatically detected facial behavior. Two learning-based classifiers, Adaboost and multinomial logistic regression are compared. Within-subject prediction of drowsiness and across-subject (subject independent) prediction of drowsiness were both tested.

3.5.1 Within subject drowsiness prediction.

For the within-subject prediction, 80% of the alert and non-alert episodes were used for training and the other 20% were reserved for testing. This resulted in a mean of 19 non-alert and 11 alert episodes for training, and 5 non-alert and 3 alert episodes for testing per subject.

The features for the Adaboost classifier consisted of each of the 30 Facial Action units of each frame of video. The classifier was trained to predict alert or non-alert from each frame of video. There was a mean of 43,200 training samples, $(24 + 11) \times 60 \times 30$, and 1440 testing samples, $(5 + 3) \times 60 \times 30$, for each subject. On each training iteration, Adaboost selected the facial action detector that minimized prediction error given the previously selected detectors. Adaboost obtained 92% correct accuracy for predicting driver drowsiness based on the facial behavior.

Classification with Adaboost was compared to that using multinomial logistic regression (MLR). Performance with MLR was slightly better, obtaining 94% correct prediction of drowsy states. The facial actions that were most highly weighted by MLR also tended to be the facial actions selected by Adaboost. 85% of the top ten facial actions as weighted by MLR were among the first 10 facial actions to be selected by Adaboost. Table 3.3 shows the within subject drowsiness prediction performances for Adaboost and MLR. Means and standard deviations in this table are shown across subjects. For the rest of the study we continued with MLR for the classification task.

Table 3.3: Performance for drowsiness prediction, within subjects. Means and standard deviations are shown across subjects.

Classifier	Percent Correct	Hit Rate	False Alarm Rate
Adaboost	.92 ±.03	.92±.01	.06±.1
MLR	.94 ±.02	.98±.02	.13±.02

3.5.2 Across subject drowsiness prediction.

The ability to predict drowsiness in novel subjects was tested by using a leave-one-out cross validation procedure. The data for each subject was first normalized to zero-mean and unit standard deviation before training the classifier. MLR was trained to predict drowsiness from the AU outputs several ways. Performance was evaluated in terms of area under the ROC. For all of the novel subject analysis, the MLR output for each feature was summed over a temporal window of 12 seconds (360 frames) before computing A'. MLR trained on all features obtained an A' of .90 for predicting drowsiness in novel subjects.

Action Unit Predictiveness: In order to understand the action unit predictiveness in drowsiness MLR was trained on framewise outputs of each facial action individually. Examination of the A' for each action unit reveals the degree to which each facial movement is associated with drowsiness in this study. The A's for the drowsy and alert states are shown in Table 3.4. Performance was evaluated in terms of area under the ROC. For all of the novel subject analysis, the MLR output for each feature was summed over a temporal window of 12 seconds (360 frames) before computing A'. Cross validation was performed with with MLR trained on 3 subjects and tested on 1 subject at a time. The average of the five facial actions that were the most predictive of drowsiness by *increasing* in drowsy states were 45, 2 (outer brow raise), 15 (frown), 17 (chin raise), and 9 (nose wrinkle). The five actions that were the most predictive of drowsiness by *decreasing* in drowsy states were 12 (smile), 7 (lid tighten), 39 (nostril compress), 4 (brow lower), and 26 (jaw drop). The high predictive ability of the blink/eye closure measure was expected. However the predictability of the outer brow raise (AU 2) was previously unknown.

Table 3.4: MLR model for predicting drowsiness across subjects. Predictive performance of each facial action individually is shown.

More when critically drowsy

AU	Name	A'
45	Blink/eye closure	0.94
2	Outer Brow Raise	0.81
15	Lip Corner Depressor	0.80
17	Chin Raiser	0.79
9	Nose wrinkle	0.78
30	Jaw sideways	0.76
20	Lip stretch	0.74
11	Nasolabial furrow	0.71
14	Dimpler	0.71
1	Inner brow raise	0.68
10	Upper Lip Raise	0.67
27	Mouth Stretch	0.66
18	Lip Pucker	0.66
22	Lip funneler	0.64
24	Lip presser	0.64
19	Tongue show	0.61

Less when critically drowsy

AU	Name	A'
12	Smile	0.87
7	Lid tighten	0.86
39	Nostril Compress	0.79
4	Brow lower	0.79
26	Jaw Drop	0.77
6	Cheek raise	0.73
38	Nostril Dilate	0.72
23	Lip tighten	0.67
8	Lips toward	0.67
5	Upper lid raise	0.65
16	Lower lip depress	0.64
32	Bite	0.63

We observed during this study that many subjects raised their eyebrows in an attempt to keep their eyes open, and the strong association of the AU

2 detector is consistent with that observation. Also of note is that action 26, jaw drop, which occurs during yawning, actually occurred *less* often in the critical 60 seconds prior to a crash. This study suggests the prediction that yawning does not tend to occur in the final moments before falling asleep.

Finally, a new MLR classifier was trained using a sequential feature selection approach, starting with the most discriminative feature (AU 45), and then iteratively adding the next most discriminative feature given the features already selected. These features are shown at the bottom of Table 3.5. Best performance of .98 was obtained with five features: 45, 2, 19 (tongue show), 26 (jaw drop), and 15. This five feature model outperformed the MLR trained on all features. Note that using all features did not improve the performance.

Table 3.5: Drowsiness detection performance for novel subjects, using an MLR classifier with different feature combinations. The weighted features are summed over 12 seconds before computing A'.

Feature	A'
AU45	.9468
AU45,AU2	.9614
AU45,AU2,AU19	.9693
AU45,AU2,AU19,AU26	.9776
AU45,AU2,AU19,AU26,AU15	.9792
all the features	.8954

Effect of Temporal Window Length: We next examined the effect of the size of the temporal window on performance. The five feature model was employed for this analysis. The performances shown to this point in this chapter were for temporal windows of one frame, with the exception of the novel subject analysis (Tables 3.4 and 3.5), which employed a temporal window of 12 seconds. The MLR output in the 5 feature model was summed over windows of N seconds, where N ranged from 0.5 to 60 seconds. Figure 3.5 shows the area under the ROC for drowsiness detection in novel subjects over time periods. Performance saturates at about 0.99 as the window size exceeds 30 seconds. In other words, given a 30 second video segment the system can discriminate sleepy versus non-sleepy segments with 0.99 accuracy across subjects.

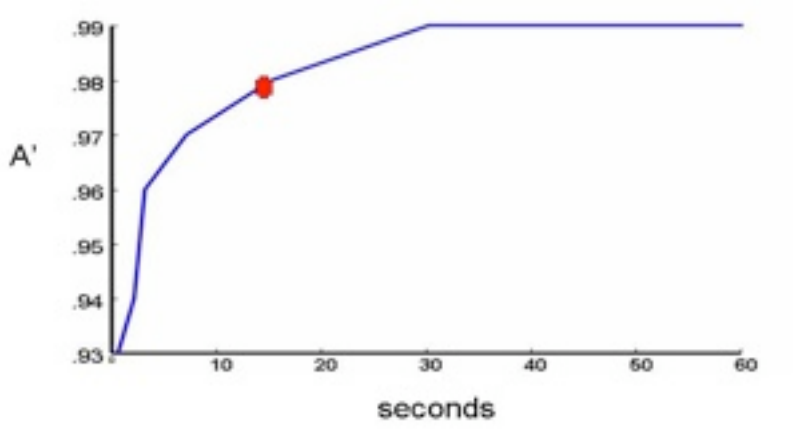

Figure 3.5: Performance for drowsiness detection in novel subjects over temporal window sizes.

3.6 Coupling of Steering and Head Motion

Observation of the subjects during drowsy and nondrowsy states indicated that the subjects head motion differed substantially when alert versus when the driver was about to fall asleep. Here the accelerometer's z axis is aligned with the gravitational force and the subject's face is looking towards the y direction of the accelerometer. Thus roll is measured as the accelerometer's measure in the x direction (roll direction). Surprisingly, head motion increased as the driver became drowsy, with large roll motion coupled with the steering motion as the driver became drowsy. Just before falling asleep, the head would become still.

We also investigated the coupling of the head and arm motions. Steering wheel motion is measured as a continuous value ranging between (-1,+1) corresponding to angular values (-pi,0). Correlations between head motion as measured by the roll dimension (of the accelerometer output and the steering wheel motion are shown in Figure 3.6. For this subject (subject 2), the correlation between head motion and steering increased from 0.33 in the alert state to 0.71 in the non-alert state. For subject 1, the correlation between head motion and steering similarly increased from 0.24 in the alert state to 0.43 in the non-alert state. The other two subjects showed a smaller

coupling effect. Future work includes combining the head motion measures and steering correlations with the facial movement measures in the predictive model. Next chapter describes a predictive model that also includes head movement measures combined with the other action units.

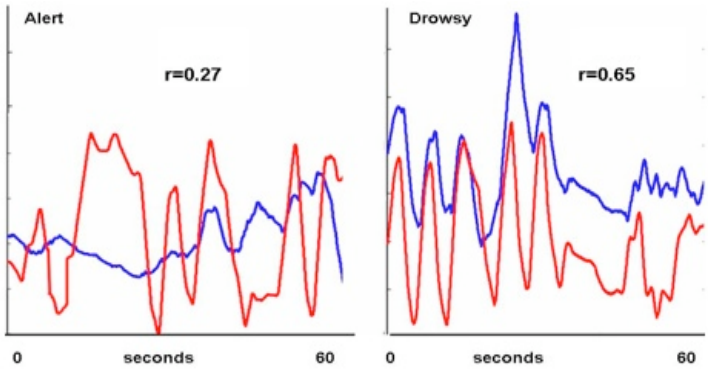

Figure 3.6: Head motion and steering position for 60 seconds in an alert state (left) and 60 seconds prior to a crash (right). Head motion is the output of the roll dimension of the accelerometer.

3.7 Coupling of eye openness and eyebrow raise.

We observed that for some of the subjects coupling between eye brow raise and eye openness increased in the drowsy state. In other words subjects tried to open their eyes using their eyebrows in an attempt to keep awake. See Figure 3.7.

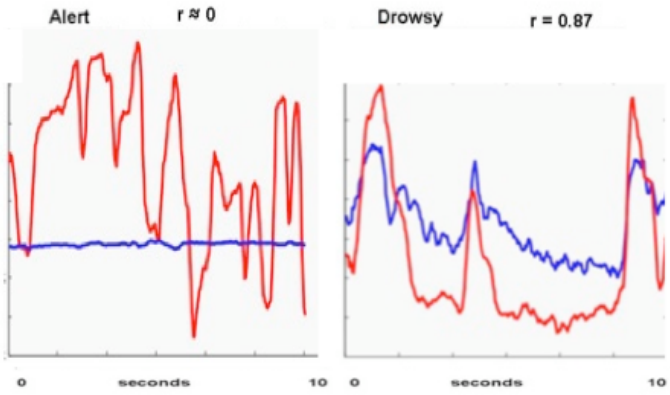

Figure 3.7: Action Unit Intensities for Eye Openness (red/black) and Eye Brow Raises (AU2) (Blue/gray) for 10 seconds in an alert state (left) and 10 seconds prior to a crash (right).

3.8 Conclusion

This chapter presented a system for automatic detection of driver drowsiness from video. A system for automatically measuring facial expressions was employed to explore spontaneous behavior during real drowsiness episodes. This is the first work to our knowledge to reveal significant associations between facial expression and fatigue beyond eyeblinks. The project also revealed a potential association between head roll and driver drowsiness, and the coupling of head roll with steering motion during drowsiness. Of note is that a behavior that is often assumed to be predictive of drowsiness, yawn, was in fact a negative predictor of the 60-second window prior to a crash. It appears that in the moments before falling asleep, drivers yawn less, not more, often. The pilot study reveals that facial expression information is This highlights the importance of using examples of fatigue and drowsiness conditions in which subjects actually fall sleep.

Chapter 4

Study II : Fine Discrimination of Fatigue States

In this chapter we explore the fine discrimination of fatigue states using facial expressions automatically extracted from video. We discriminate moderate drowsiness from acute drowsiness in the UYAN-2 dataset. In particular the degree to which individual facial action units can predict the difference between moderately drowsy to acutely drowsy is studied. Signal processing techniques and machine learning methods are employed on the UYAN-2 dataset to build a person independent acute drowsiness detection system. Temporal dynamics are captured using a bank of temporal filters. How to extract the prominent set of features for individual action units is analyzed. A person independent acute drowsiness detector was built by selecting and combining relevant features of all the action units.

For this study in order to increase the set of subjects for drowsiness and to expand the drowsiness measures a new dataset is collected. In this chapter we describe the "UYAN-2" dataset which consists of 11 subjects using the driving simulator while their faces are captured with a DV Camera and the brain dynamics and upper torso movements are measured through EEG and Motion Capture facilities respectively. Prior to the collection of 11 subjects an additional 6 subject data were collected under different illumination conditions. However these subjects are not included in the dataset as the illumination was not sufficient for the CERT system to operate. In the future low-lighting conditions can be explored using near infra-red (NIR) technology. The UYAN-2 dataset differs from the UYAN-1 dataset in two ways: (1) It is a larger dataset (2) It includes an expanded set of measures including

brain dynamics (EEG), upper torso and head movements through a motion capture. The head movement measures are also captured using a computer vision based head tracking system.

4.1 UYAN-2 Dataset

4.1.1 Experimental Setup

Subjects were asked to drive a virtual car simulator on a Windows machine using a steering wheel and an open source multi-platform video game. Same experimental setup is being used as in UYAN-1 dataset with the following differences: Subjects arrived at the motion capture and brain dynamics lab at 11:00 pm. Preparation of the EEG and motion capture setup took over an hour. Eleven subjects performed the driving task beginning at 12:30 am over a three hour period.

For the "UYAN-2" dataset brainwaves were measured through an EEG facility [1] that records 64 channels of brainwaves with a rate of 512 Hz. Upper torso and head movements are measured through a Motion Capture facility. In order to measure body and head movements subjects wear an led sensor affixed jacket and a head band. The the led sensor positions are captured by a set of 12 infrared cameras [2] and recorded at a rate of 480 Hz. In Figure 4.1 the subject is displayed wearing an EEG cap for the measurement of brain waves and a motion capture jacket and a hat for the analysis of body and head movements.

The motion capture and EEG facilities are run on Windows and Unix servers respectively using different sampling rates. Similarly the driving game is run on a different Windows server recording driving signals at a 30 fps sampling rate. The synchronization of the signals is a crucial step for this experimental setup. For the synchrony of these three signals a hardware solution was installed. The three signals were synchronized with the help of the trigger port of the box that converts optical data to USB2 output (USB2 Box of EEG Biosemi device). The driving simulator and the motion capture facilities sent a trigger signal for each sample that is being recorded. The trigger channel of EEG device was decoded in an offline manner for matching the time steps of the signals.

[1] Biosemi Instrumentation Inc

[2] PhaseSpace Inc

4.1.2 Measures of Drowsiness

For this experimental setup measures of drowsiness include objective measures, such as crash, EEG, and steering signals, as well as subjective measures obtained from human ratings of fatigue. Note that for the scope of the thesis we only used time to crash as a measure of drowsiness.

4.1.3 Subject Variability

11 subjects were selected for the Study II analysis. The set of subjects in the "UYAN-2" dataset is quite different in terms of initial drowsiness levels. For this dataset most of the subjects were very tired while beginning the experiment. This can be mainly attributed to the initial preparation stage of the experiment which lasted longer than the preparation stage of UYAN-1 dataset. Each subject had to go through a process of an hour long process of EEG cap preparation and motion capture calibration stage which were quite exhaustive for the subjects. As a result the subjects are mostly more tired than the UYAN-1 dataset.

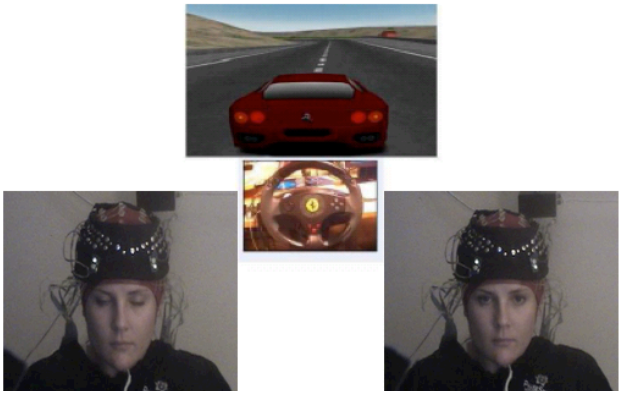

Figure 4.1: In this task samples of real sleep episodes were collected from 11 subjects while they were performing a driving simulator task at midnight for an entire 3 hour session.

4.1.4 Extraction of Facial Expressions

Computer Expression Recognition Toolbox (CERT), which is a user independent fully automatic system for real time recognition of facial actions from

the Facial Action Coding System (FACS), is again used for the analysis of the fully automated extracted facial actions. The system automatically detects frontal faces in the video stream and codes each frame with respect to 36 Action units. For this analysis we work on 22 Action units including a set of promising action units of Study I. While choosing these CERT AUs we also considered the criteria of having enough training samples for reliable CERT outputs. A snapshot of the CERT system running on one of the subjects from UYAN-2 dataset is shown in Figure 4.2.

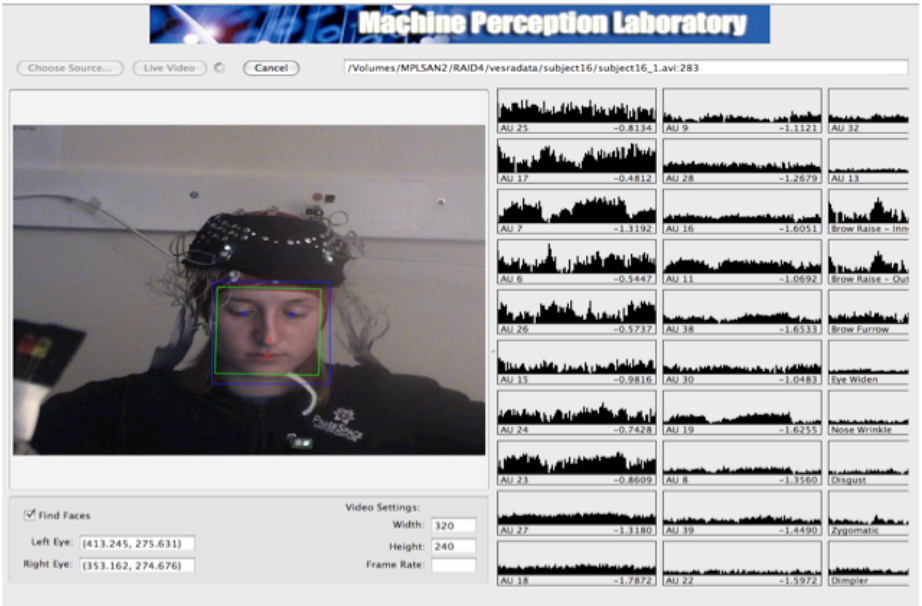

Figure 4.2: Facial expressions are measured automatically using the Computer Expression Recognition Toolbox (CERT). 22 Action Units from the Facial Action Coding System (Ekman & Friesen, 1978) are measured. Head and body motion are measured using the motion capture facility, as well as the steering signal. Measures of alertness include EEG, distance to the road center, and simulator crash. For the context of the thesis simulator crash is being used as a measure of drowsiness..

A list of these 22 action units can be found in the following table. Note that head movement action units such as Head Pitch, Head Roll and Head Yaw are also included as part of the analysis. These measure are output through CERT toolbox. For this analysis CERT outputs a single signal

output for AU 53 (Head Up) and AU54 (Head Down) which are treated part of the head pitch signal. Similarly AU 55 and AU 56 are treated as the head roll signal. Lastly AU 51 (Head Turn Left) or AU 52 (Head Turn Right) are treated as the head yaw signal.

Action Unit	Action Unit Name
AU 1	Inner Brow Raise
AU 2	Outer Brow Raise
AU 4	Brow Lowerer
AU 5	Upper Lid Raise
AU 6	Cheek Raise
AU 7	Lids Tight
AU 9	Nose Wrinkle
AU 10	Upper Lip Raise
AU 12	Lip Corner Puller
AU 14	Dimpler
AU 15	Lip Corner Depressor
AU 17	Chin Raise
AU 18	Lip Puckerer
AU 20	Lip Stretch
AU 23	Lip Tightener
AU 24	Lip Presser
AU 26	Jaw Drop
AU 28	Lip Suck
AU 53 or AU 54	Head Pitch
AU 55 or AU 56	Head Roll
AU 51 or AU 52	Head Yaw
AU45	Eye Closure

Table 4.1: A list of 22 action unit outputs from CERT toolbox that are chosen for the analysis.

As in Study I we use time to crash crash as a measure of drowsiness for this analysis and the goal is to predict acutely drowsy (60 second window before crash) versus moderately drowsy episodes. The Study I analysis revealed that 10 seconds is a satisfactory temporal duration for predicting drowsiness. Therefore in this study we employ temporal analysis over 10 seconds. Contrary to the subjects in the UYAN-1 dataset, some subjects in the UYAN-2

dataset are already quite drowsy at the beginning of the driving task. This is due to the fact that prior to the driving task subjects had to go through a 1 hour period to set the EEG cap and the motion capture system. Most of the subjects for the UYAN-2 dataset crash within 25 minutes after starting the experiment. Hence we call the initial state of the subjects in UYAN-2 dataset "moderately drowsy (MD)" rather than "alert". Table 4.3 displays the mean and standard deviation of the time to the first crash for the alert and moderately drowsy segments of the UYAN-1 and UYAN-2 datasets respectively. Notice that the two datasets have different set of subjects. For this analysis five one minute non-overlapping moderately drowsy episodes were selected per subject from the first 10 minutes of the driving task based on the criteria of being as far away from the crash point as possible. First the one minute episode that is as far away from a crash point is chosen from the first 10 minutes of the driving task and then the next one minute episode that is the second furthest episode from a crash point and not necessarily consecutive to the previous episode is selected and the iteration continues in this fashion. Similar to study I one minute before crash points were taken as drowsy episodes and this state is called "acute drowsiness (AD)". In UYAN-2 dataset the two classes are in close proximity. The A' for some action units such as eye closure also supports this fact. In UYAN-1 across subjects A' for eye closure was .94 whereas here the A' for eye closure action unit as will be shown later in the chapter (See Table4.5) is .83. Table 4.2 displays some statistics for the average duration in minutes to the initial and next crash after an MD and AD episode respectively. 5 subjects crashed within 20 minutes after an MD episode. Two subjects crashed within the first 10 minutes after an MD episode. The duration in minutes to the next crash point after an AD episode is less than 5 minutes for 7 subjects.

	time to first crash for MD		time to first crash for AD	
Subject:	mean	standard deviation	mean	standard deviation
Subject 1	16.35	1.58	1.0	0.0
Subject 2	24.86	1.58	1.0	0.0
Subject 3	56.32	1.58	1.0	0.0
Subject 4	21.43	1.58	1.0	0.0
Subject 5	16.27	1.58	1.0	0.0
Subject 6	48.46	1.58	1.0	0.0
Subject 7	-	-	1.0	0.0
Subject 8	42.53	1.51	-	-
Subject 9	2.87	1.40	1.0	0.0
Subject 10	8.74	1.64	1.0	0.0
Subject 11	13.98	1.61	1.0	0.0

	time to second crash for MD		time to second crash for AD	
Subject:	mean	standard deviation	mean	standard deviation
Subject 1	28.07	1.58	5.23	4.60
Subject 2	31.08	1.58	1.93	4.22
Subject 3	66.32	1.58	19.78	20.67
Subject 4	31.90	1.58	1.84	3.16
Subject 5	41.90	1.58	0.73	2.49
Subject 6	64.60	1.58	4.39	9.81
Subject 7	-	-	26.92	18.50
Subject 8	72.53	1.51	-	-
Subject 9	7.98	1.40	1.24	1.67
Subject 10	15.93	1.64	3.02	3.15
Subject 11	17.02	1.61	1.83	2.32

Table 4.2: The mean and standard deviation of time to crash for one minute segments of moderate drowsiness (MD) and acute drowsiness (AD).

	time to first crash for MD UYAN-2	
Subject:	mean	standard deviation
Subject 1	16.35	1.58
Subject 2	24.86	1.58
Subject 3	56.32	1.58
Subject 4	21.43	1.58
Subject 5	16.27	1.58
Subject 6	48.46	1.58
Subject 7	-	-
Subject 8	42.53	1.51
Subject 9	2.87	1.40
Subject 10	8.74	1.64
Subject 11	13.98	1.61
	time to first crash for Alert UYAN-1	
Subject:	mean	standard deviation
Subject 1	49.6	4.18
Subject 2	71.89	34.83
Subject 3	33.31	4.18
Subject 4	90.35	4.18

Table 4.3: Table displays the mean and standard deviation of the time to the first crash for the alert and moderately drowsy segments of the UYAN-1 and UYAN-2 datasets respectively. Notice that the two datasets have different set of subjects.

The selected episodes are segmented into non-overlapping 10 second video patches. CERT action unit outputs are obtained over each of these patches. The UYAN-2 dataset contains data from 11 subjects. 9 subjects have both crash and alert episodes. One subject does not have any alert segments as the first hour of the driving task is lost accidentally by the subject. One other subject's data for crash episodes is lost therefore the subject's crash data is not included. For some of the 10 second segments the face could not be located due to occlusion or false alarms in CERT face detection module. The 10 second segments that have more than 30 video frames with occlusions or false alarms for the face (assuming the video is 30 fps this corresponds to 1 second) were eliminated. Here false alarms in face detection were detected using the measure of distance between the eyes. Usually the subject's

Subject	#. segments for AD	#. segments for MD
Subject 1	136	30
Subject 2	221	27
Subject 3	18	30
Subject 4	422	30
Subject 5	619	30
Subject 6	37	30
Subject 7	59	0
Subject 8	0	30
Subject 9	213	30
Subject 10	41	30
Subject 11	184	30
Mean:	177.27	27
Std:	192.08	9.0

Table 4.4: The number of 10 second segments for acute drowsiness (AD) and moderate drowsiness (MD) is listed in the table. These segments are obtained by partitioning one minute alert and drowsy episodes into six 10 second patches. Note that Subject 7 and 8 do not not have any MD and AD segments respectively. Temporal dynamics are captured by employing temporal filters over these 10 second CERT action unit signals.

distance between his/her eyes should be within a range of a constant measure and if there is a sudden jump in the distance between the eyes indicates that the face detector started focusing on an object other than the subject's face. For the rest of the clips that have false alarms in the face detection or occlusions of the face where the CERT cannot locate the face the action unit signal is interpolated. Table 4.4 demonstrates the number of moderately drowsy and acutely drowsy 10 second segments for each subject. Subjects have a mean of 27 moderately drowsy 10 second patches. Subjects have a mean of 177 acutely drowsy 10 second patches ranging from a minimum of 18 to a maximum of 619. Different subjects exhibit different behavior, some crash more therefore the number of AD segments between subjects change dramatically depending on the number of crash incidents occurring during the experimental task.

4.2 Discriminating Acute versus Moderate Drowsiness Using Raw Action Unit Output

The goal in this analysis is to explore the predictive power of the individual facial action units. Averages of raw action unit outputs are computed over 10 second patches of individual CERT action unit outputs. Separability of the averaged raw action unit outputs is analyzed. We perform a leave-one-out cross validation training procedure. A subject was left out for testing and the model was trained with the rest of the subjects. MLR classifier is trained with 10 training subjects at a time and tested with a novel test subject. For training averaged 10 second patches of an individual action unit are passed to an MLR classifier. The training data has a single feature which is the average of the action unit for a 10 second patch. The data points are equal to the number of 10 second patches of 10 subjects. Similarly the test data has a single feature which is the 10 second average of an action unit and the data points are equal to the number of 10 second patches of the test subject. Since this is a single feature model, the only parameter that can change the A' value for the test subject is the sign assigned by the MLR model to the AU under consideration. For example based on 10 training subjects the MLR model may assign a positive weight to a particular AU. This means that most of the training subjects have increasing intensity value for this particular AU as the subject gets more drowsy under the assumption that the acute drowsiness state is assigned to a positive class. For each test subject MLR weight obtained from training was multiplied with the test feature to estimate the A' over the test outputs. While for training an A' below 0.5 is not possible, for an actual system that is required to generalize to new people the test subject A's may take values less than 0.5. For testing 9 test subjects were tested as two subjects did not have either drowsy or alert episodes. Individual action unit discriminability measure is estimated by averaging 9 test subject's A's. Note that the number of crashes across subjects changed dramatically and this way of calculation helps the performance measure such that it is not biased for subjects having a large number of drowsy patches. Using this method we can highlight the action units that are informative for a person independent system. This analysis is repeated for all the action units. Table 4.5 displays the individual action unit mean A' estimates over all subjects. The 5 most informative units were (1) Eye Closure (AU45) with an A' of 0.83 (2) Lip Puckerer (AU18) with an A' of 0.82 (3) Head Roll with

an A' of 0.77 (4) Lid Tightener (AU7) with an ROC of 0.71 (5) Nose Wrinkle (AU9) with an A' of 0.69.

In this study we found that brow raise(AU2) increases in drowsy conditions in some subjects. As these subjects get more drowsy they raise their eye brows. However as displayed in Figure 4.9 for one subject (subject 3) AU2 decreased dramatically in acute drowsiness and increased (the subject raised his brows) in the moderately drowsy condition. As this subject contradicts with the rest of the subjects for this action unit, the A' value is less than 0.5 and closer to 0. This results in AU2 to be a non-informative action unit for a subject independent system due to the cancelling effect of the decrease and increase and neutral state of the intensity values. In the UYAN-1 dataset all of the 4 subjects have increasing AU2 intensity values as they are getting more drowsy. With UYAN-2 dataset we are seeing variabilities among subjects for AU2.

Action Unit	A'	Standard Error
AU 45	0.8346	0.0587
AU 18	0.8247	0.0367
Head Roll	0.7761	0.0723
AU 7	0.7175	0.0884
AU 9	0.6951	0.0702
AU 14	0.6402	0.0943
AU 23	0.6315	0.0857
AU 26	0.6233	0.0728
AU 12	0.5920	0.0971
Head Pitch	0.5665	0.1033
AU 1	0.5632	0.0973
AU 28	0.5368	0.0755
AU 4	0.5035	0.0735
AU 2	0.4758	0.0938
AU 15	0.4684	0.1346
AU 20	0.4629	0.0880
Head Yaw	0.4543	0.0830
AU 17	0.4393	0.0918
AU 6	0.3457	0.0765
AU 24	0.3150	0.0494
AU 10	0.2787	0.0777
AU 5	0.2514	0.0602

Table 4.5: ROC performance results for the output of the raw action unit outputs over individual action units.

To illustrate how the subjects differ from each other for certain action units we display the histograms of action unit output values for individual subjects summed over 10 second segments. Figure 4.3 displays the histograms over sums of eye closure action unit for 10 second segments of the acute drowsy and moderate drowsy cases. The red histogram corresponds to the acute drowsy samples and the blue corresponds to moderately drowsy samples. Here 9 subjects are plotted as 2 subjects do not have AD or MD samples. The A' here is computed using the samples of the subject without multiplying with the training weight. Eye closure (AU45) is discriminative for 7 of the test subjects. Similarly 4.4 displays the histograms for the head

roll action unit. This action unit is discriminative for all the subjects. Except for one subject the head roll average intensity value increases as the subject gets more drowsy. Figure 4.5 displays the histogram for the lip pucker (AU18) action unit. Except for one subject this action unit is discriminative for all subjects in the increasing direction as the subject gets more drowsy. Figure 4.6 displays the histograms for the lid tighten (AU7) action unit. For 7 subjects this action unit looks discriminative and for 6 subjects it is increasing as the subject gets more drowsy. We also analyzed subjectwise discriminability of some of the action units that do not perform well for a subject independent test. Figure 4.8 displays the histograms for the upper lip raiser (AU10) action unit. This action unit is discriminative for 2 subjects in the increasing direction as the subject gets more drowsy and for 3 subjects in the decreasing direction as the subject gets more drowsy. For the rest of the subjects this action unit is not discriminative. Similarly Figure 4.9 displays the histograms for the action unit eye brow raise (AU2). For 4 of the subjects this action is discriminative. For 3 subjects this action unit increases as the subject gets more drowsy and for one subject the intensity value decreases as the subject gets more drowsy.

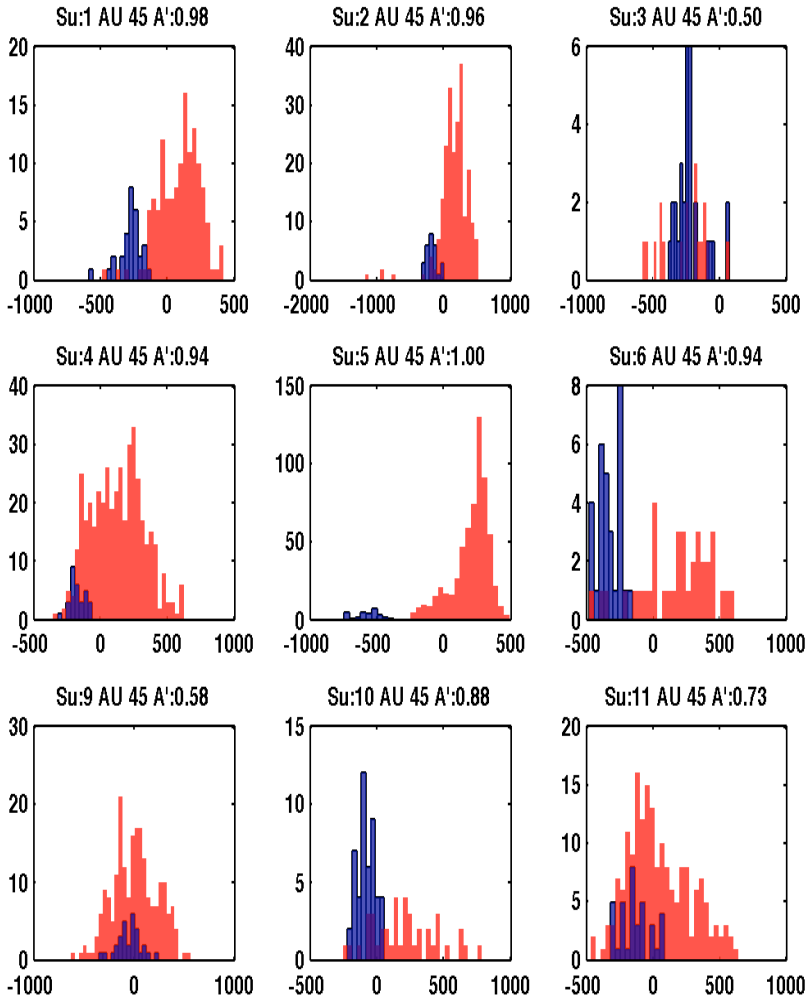

Figure 4.3: Figure displays the histograms of eye closure (AU45) signal for individual subjects summed over 10 second segments of acute drowsy and moderate drowsy samples. The red histogram corresponds to the acute drowsy samples and the blue histogram corresponds to moderately drowsy samples. Here 9 subjects are plotted as 2 subjects do not have either AD or MD samples. The A' here is computed using the samples of the subject without multiplying with training weight.

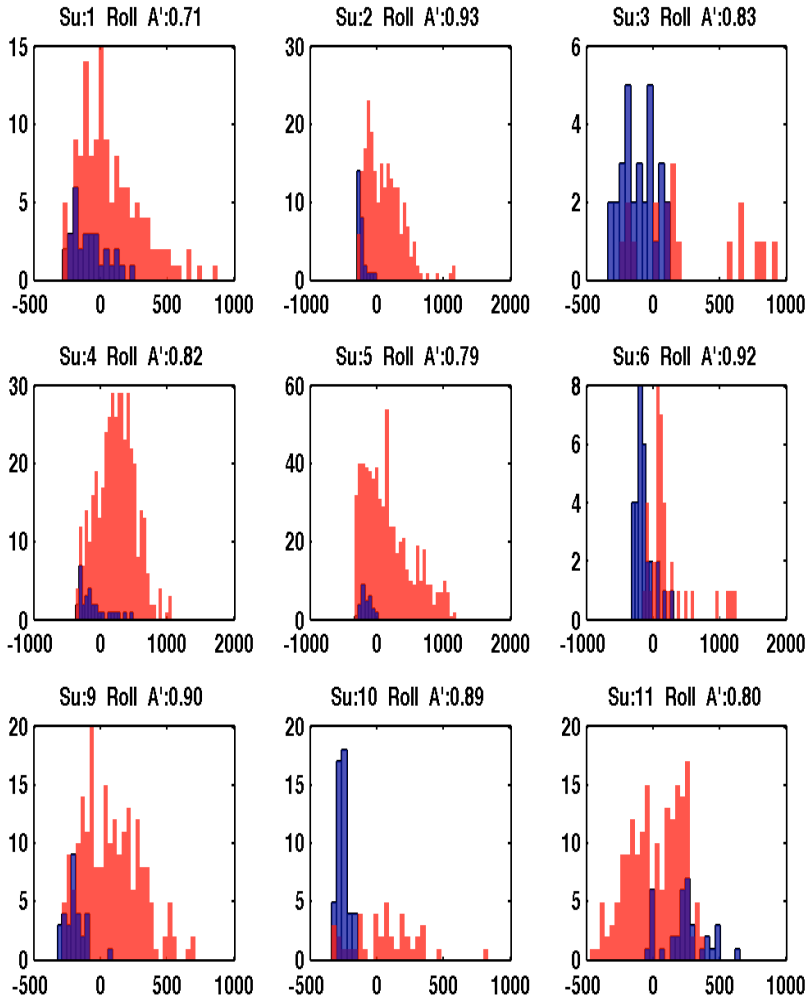

Figure 4.4: Figure displays the histograms of head roll signal for individual subjects summed over 10 second segments of acute drowsy and moderate drowsy samples. The red histogram corresponds to the acute drowsy samples and the blue histogram corresponds to moderately drowsy samples. Here 9 subjects are plotted as 2 subjects do not have either AD or MD samples. The A' here is computed using the samples of the subject without multiplying with training weight.

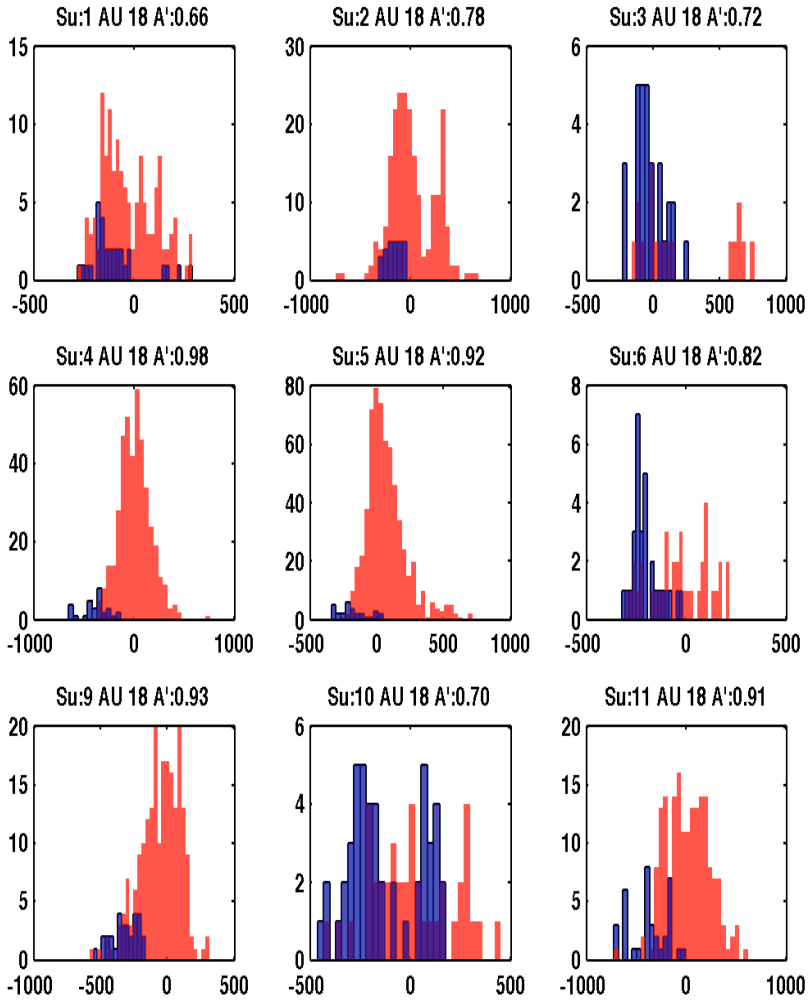

Figure 4.5: Figure displays the histograms of lip pucker (AU18) signal for individual subjects summed over 10 second segments for acute drowsy and moderate drowsy samples. The red histogram corresponds to the acute drowsy samples and the blue histogram corresponds to moderately drowsy samples. Here 9 subjects are plotted as 2 subjects do not have either AD or MD samples. The A' here is computed using the samples of the subject and without using a training weight.

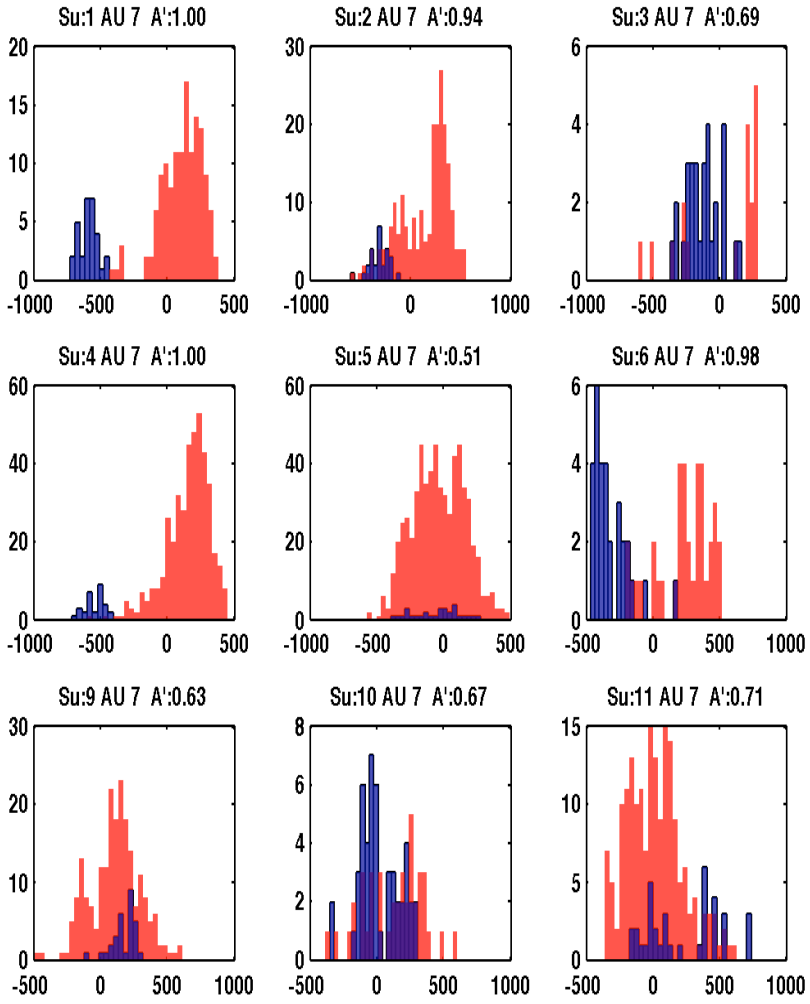

Figure 4.6: Figure displays the histograms of summed lid tighten (AU7) signal for individual subjects summed over 10 second segments of acute drowsy and moderate drowsy samples. The red histogram corresponds to the acute drowsy samples and the blue histogram corresponds to moderately drowsy samples. Here 9 subjects are plotted as 2 subjects do not have either AD or MD samples. The A' here is computed using the samples of the subject without multiplying with a training weight.

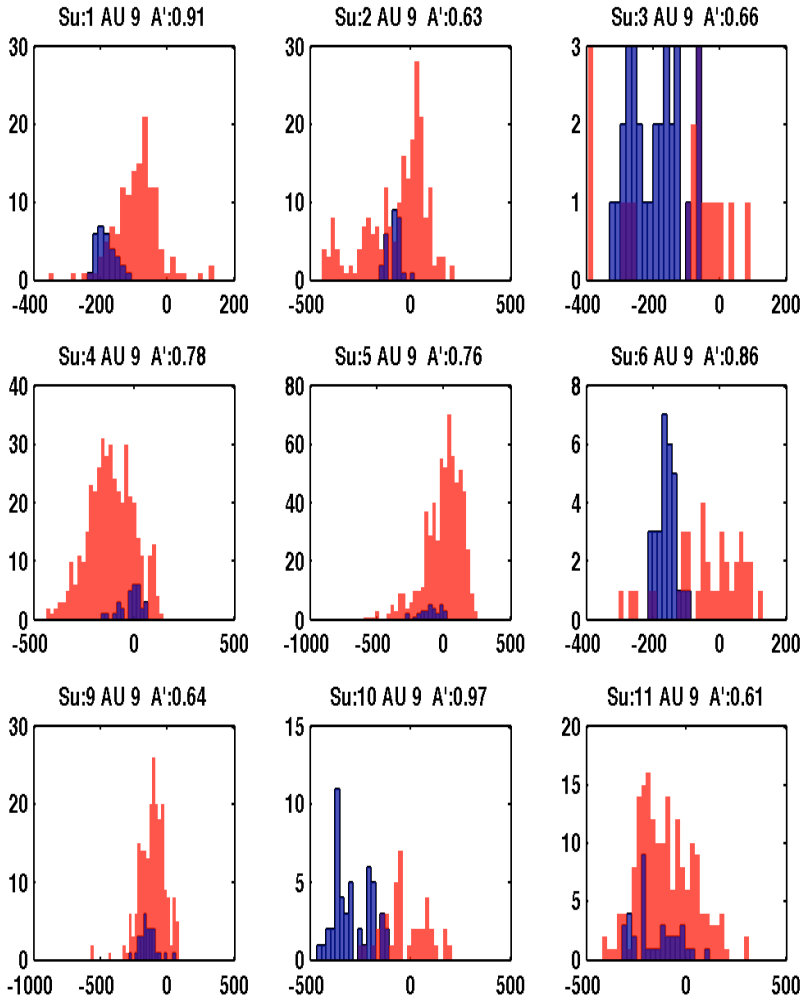

Figure 4.7: Figure displays the histograms of summed nose wrinkle (AU9) signal for individual subjects summed over 10 second segments of acute drowsy and moderate drowsy samples. The red histogram corresponds to the acute drowsy samples and the blue histogram corresponds to moderately drowsy samples. Here 9 subjects are plotted as 2 subjects do not have either AD or MD samples. The A' here is computed using the samples of the subject without multiplying with a training weight.

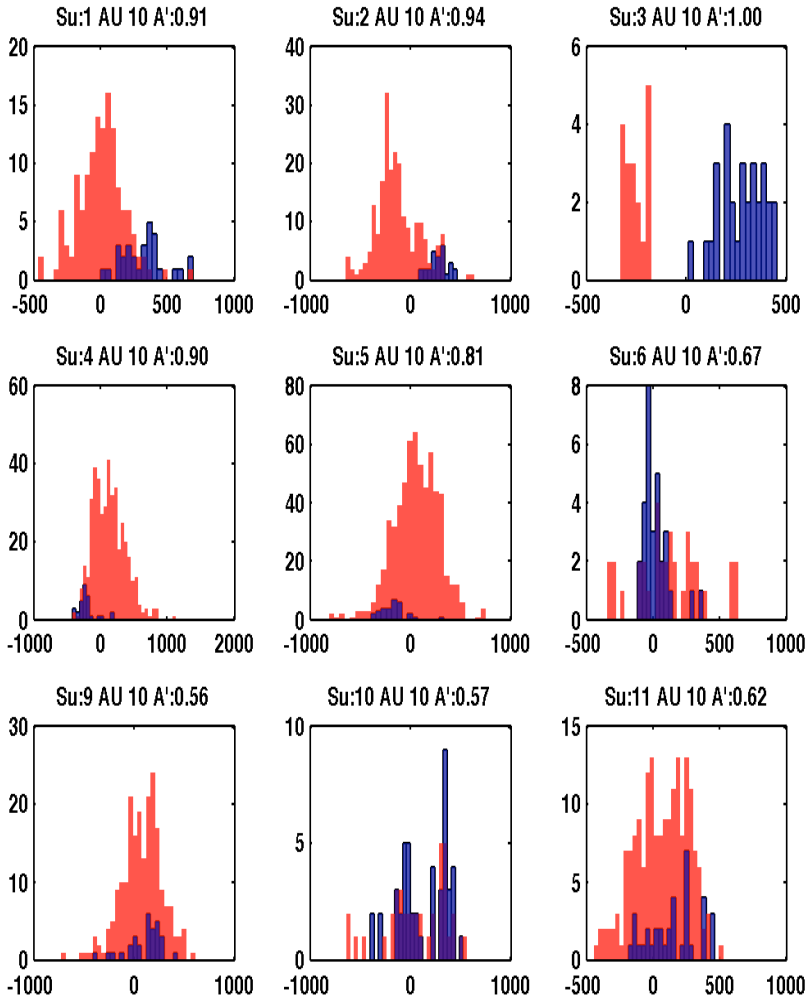

Figure 4.8: Figure displays the histograms of upper lid raiser (AU10) signal for individual subjects summed over 10 second segments for acute drowsy and moderate drowsy samples. The red histogram corresponds to the acute drowsy samples and the blue histogram corresponds to moderately drowsy samples. Here 9 subjects are plotted as 2 subjects do not have either AD or MD samples. The A' here is computed using the samples of the subject without multiplying with training weight.

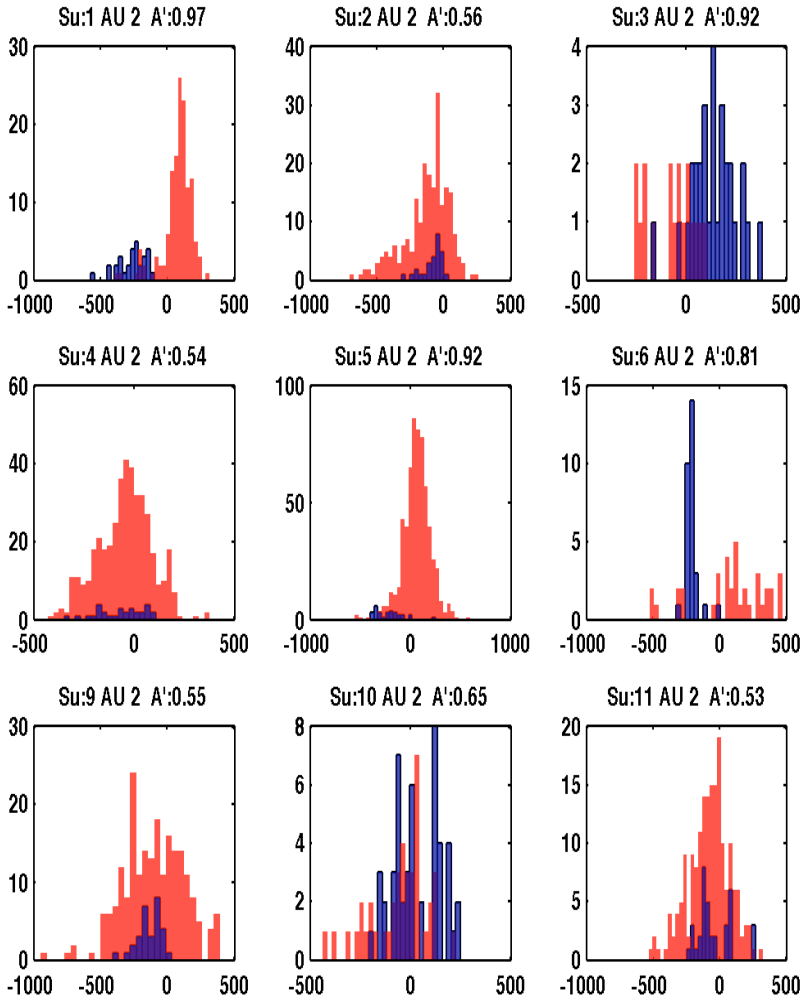

Figure 4.9: Figure displays the histograms of eye brow raise (AU2) signal for individual subjects summed over 10 second segments of acute drowsy and moderate drowsy samples. The red histogram corresponds to the acute drowsy samples and the blue histogram corresponds to moderately drowsy samples. Here 9 subjects are plotted as 2 subjects do not have either AD or MD samples. The A' here is computed using the samples of the subject without multiplying with training weight.

Finally MLR model is trained with the combined 5 most informative action units by performing leave-one-out cross validation. At each training iteration one subjects was left out. The model was trained with different regularization constants. The MLR combined model achieved 0.92 performance. Figure 4.10 displays the A' performances in the vertical axis and the regularization constant in the horizontal axis for the combined model.

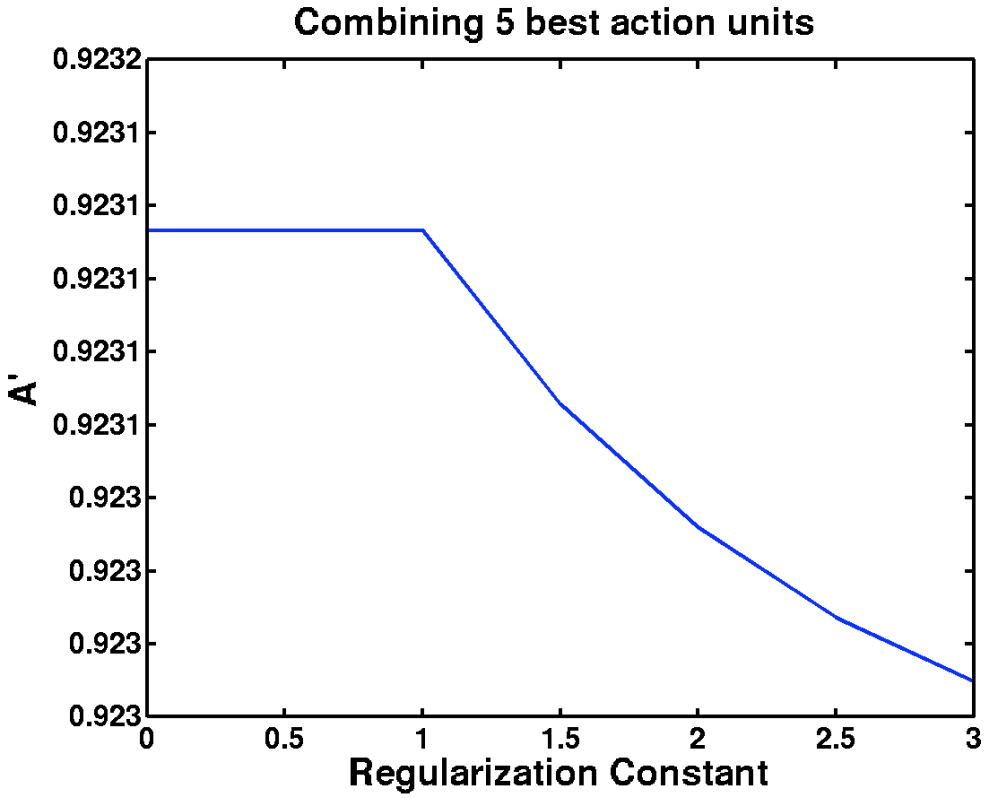

Figure 4.10: MLR model performances for the combined 5 most informative action units by performing leave-one-out cross validation

4.3 Discriminating Acute versus Moderate Drowsiness Using Temporal Gabor Filter Output

Analysis of some video clips from the UYAN-2 dataset revealed the fact that averaging the AU outputs over 10 second segments may lose important infor-

mation about dynamics. Consider for example, the data displayed in Figure 4.11. The figure displays the output of AU 45 over two different 10 second episodes, one from the MD condition and the other one from the AD condition. These two figures have the same mean action unit output value over 10 seconds. Thus the mean filter approach would not be able to differentiate these two episodes. However the temporal analysis of these segments can bring additional information for discriminating the two episodes.

The approach we have used to capture temporal dynamics is to process the AU outputs using a bank of temporal Gabor Filters. Gabor filters are sine wave gratings modulated by a Gaussian and they model the simple cells in the mammalian visual system. Daugman showed that for 2D banks of Gabor filters at multiple scales and orientations model the response properties of primary visual cortical cells in primates (Daugman 1988) . 2D Gabor filters have been used in the literature to extract the spatial information in images. Here we use Gabor filters on uni-dimensional action unit signal to extract temporal information. For this analysis Gabor filters are applied to 10 second patches of unsmoothed CERT action unit outputs to get periodicity information.

For capturing the temporal dynamics we applied Gabor energy (Magnitude Gabor), Gabor sine carrier (Real Gabor) and Gabor cosine (Imaginary Gabor) carrier filters to CERT action unit outputs. Figure 4.12 displays a sinusoidal sample signal convolved with these three types of filters. The real and imaginary components of a complex Gabor filter are phase sensitive, i.e., as a consequence their response to a sinusoid is another sinusoid. By getting the magnitude of the output (square root of the sum of squared real and imaginary outputs) we can get a response that is phase insensitive and thus unmodulated positive response to a target sinusoid input. Note that the real and imaginary Gabor filters are linear filters whereas the magnitude Gabor filter is a nonlinear filter. For all of the filters the output response magnitudes are half the magnitude of the input signal. The real and imaginary filter input with a sinusoid input outputs a sinusoidal output with half magnitude. However the magnitude filter basically outputs the location of the sinusoid but does not keep the sinusoidal properties of the signal. For this study we found that both Gabor energy and Gabor sine and cosine carrier filters were informative for discriminating subject state. Therefore we used all three outputs for our feature analysis.

For this analysis different bandwidth and frequency temporal Gabor filters have been applied to ten second segments of an action unit signal for

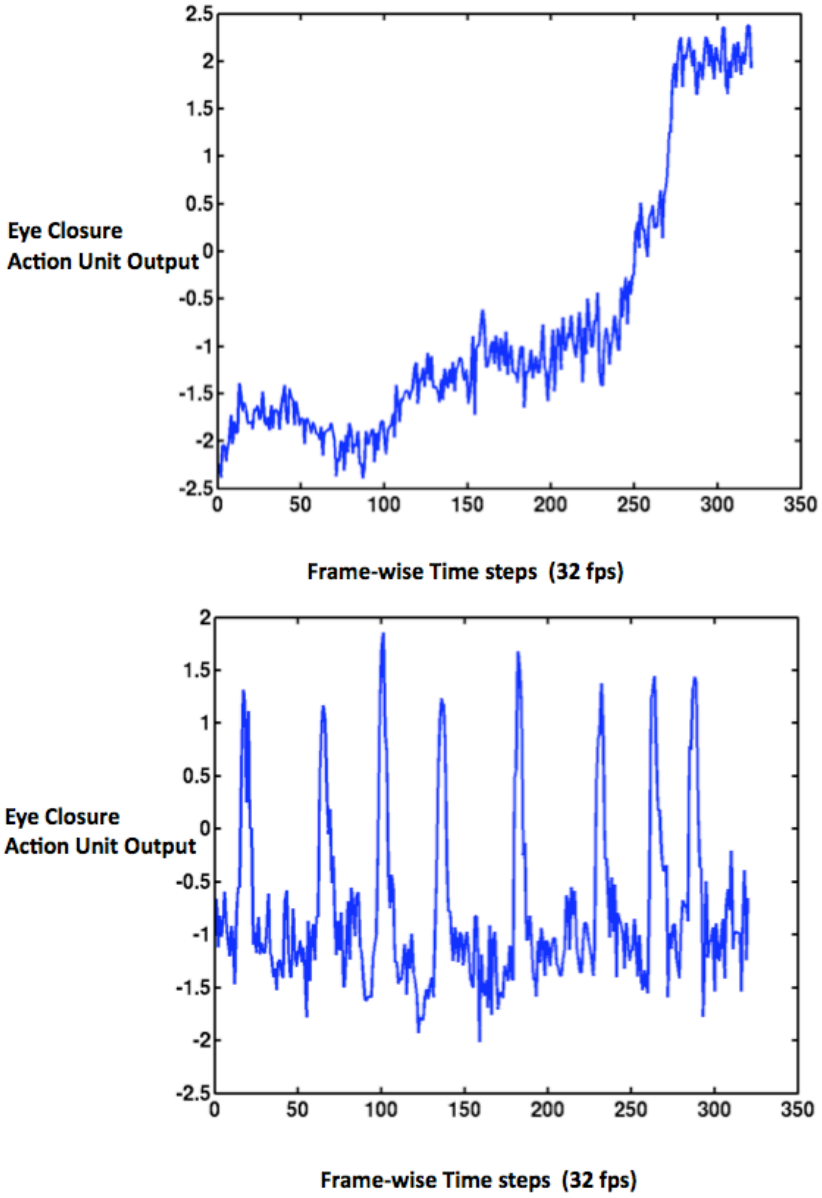

Figure 4.11: This figure displays a case where temporal dynamics plays an important role in discriminating two cases. The first case (figure on the top) corresponds to a AD. The subject's eyes are open all the time except towards the end of the clip. The second case (figure on the bottom) demonstrates an moderately drowsy (MD) clip from another subject. These two eye closure signals have approximately the same mean. The output would not be able to tell apart which of these two clips belongs to the AD or MD episode

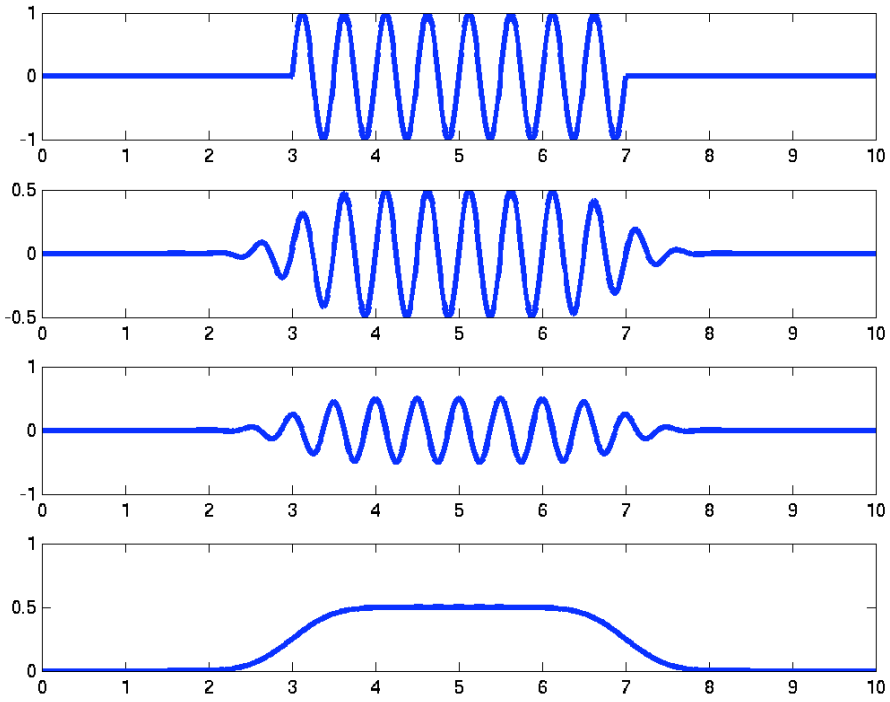

Figure 4.12: Top: An input signal. Second: Output of Gabor filter (cosine carrier). Third: Output of Gabor Filter in quadrature (sine carrier); Fourth: Output of Gabor Energy Filter [43]

individual subjects. The Gabor representations of temporal action unit signals were obtained by convolving the temporal signal with a bank of 306 Gabor filters consisting of 18 frequency components ranging from 0 to 8 Hz and 17 bandwidths ranging from 0 to 8 Hz. The Gabor filter frequencies used for the analysis have the following values : 8.0, 7.0, 6.0, 5.0, 4.0, 3.0, 2.25, 1.6875, 1.2656, 0.9492, 0.7119, 0.5339, 0.4005, 0.3003, 0.2253, 0.1689, 0.01 0. The bandwidths have the same values excluding zero frequency. Figure 4.13 displays filtered version of the signals in Figure 4.11 where the applied filter is a magnitude Gabor Filter with frequency 1.26 and bandwidth 1.26. The AD signal has a mean of 0.11 and the MD signal has a mean of 0.36. The magnitude filter can discriminate the two signals whereas the mean filter approach was not able to discriminate the two signals.

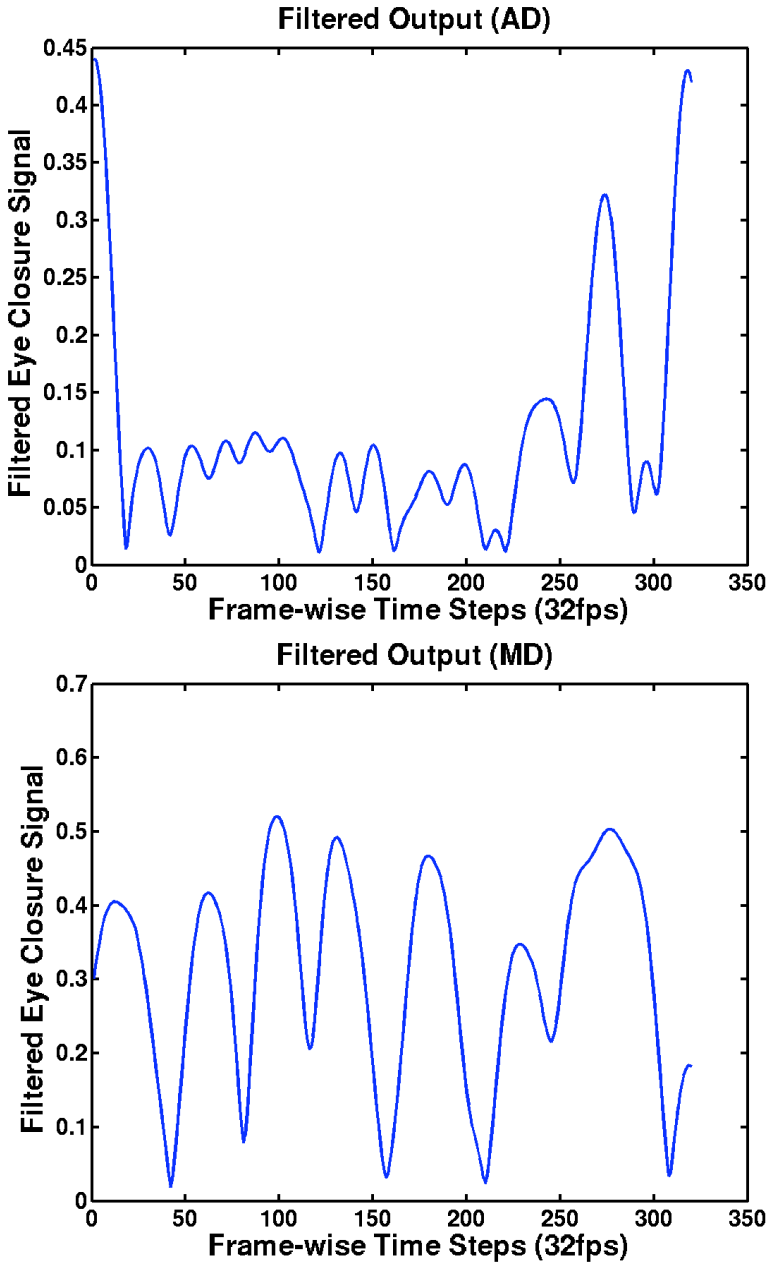

Figure 4.13: Filtered version of the signals in Figure 4.11 where the applied filter is a magnitude Gabor Filter with frequency 1.26 and bandwidth 1.26. The AD signal has a mean of 0.11 and the MD signal has a mean of 0.36.

4.4 Predictive Power of Individual Gabor Filters

The goal in this analysis is to understand which Gabor filters have predictive power at discriminating moderately drowsy from acutely drowsy segments for a person independent system. For individual action units Gabor filter outputs were obtained by convolving 10 second moderately drowsy and acutely drowsy patches with each of the 918 Filters where these filters were Gabor energy (Absolute Gabor -306 filters 18 frequencyx 17 bandwidth), Gabor sine carrier (Real Gabor- 306 filters 18 frequencyx17 bandwidth) and Gabor cosine carrier filters (Imaginary Gabor- 306 filters- 18 frequencyx17 bandwidth). In order to estimate each individual Gabor filter's predictive power for an action unit, averages of individual Gabor filter outputs are used as an input to a single feature MLR model by leaving one subject out at a time. For the magnitude filter output averaging operation outputs the average energy over the temporal signal. For linear filters such as real and imaginary Gabor filters averaging is basically the operation of compounding two filters: the Gabor filter and the averaging filter (lowpass). Here we could have also used other statistics such as median, percentiles, range or other filters for the linear filter output that might keep more information about the linear filter output. As a future work we are planning to explore more of these filters instead of using the simple averaging. The training data has single feature which is the average of the Gabor filter output for 10 second patches and the number of data points is equal to the total number 10 second AD and MD patches of 10 training subjects. Similarly the test data have a single feature which is again the 10 second average of the Gabor filter output for an action unit and the number of data points is equal to the number of 10 second AD and MD patches of the test subject. For each test subject, MLR weight obtained from training was multiplied with test data points. From these weighted outputs an A' was estimated for each test subject. Individual action unit discriminability for each Gabor Filter is estimated by averaging 9 test subject's A's. Since this is a single feature model training has no effect other than changing the sign of the A' value. Figure 4.14 displays the A' for eye closure action unit (AU45) of the Real filter for a set of 18 frequencies and 17 bandwidths. The horizontal axis displays the frequency and the vertical axis displays the bandwidth whereas the color denotes the A' value. Notice that low frequencies with all the bandwidths are discriminative of acute

drowsiness for this action unit. As the frequency increases low bandwidths become less discriminative. One reason for this could be due to not capturing the negative frequencies with smaller bandwidths.

In general the Real filter for the eye closure action unit (AU45) is very discriminative of acute drowsiness. Similarly Figure 4.15 displays the A' of individual Gabor Filters for all the action units. Each of the 66 (22x3) boxes above represent the A' performances for a specific action unit for either absolute real or imaginary filter sets. For each box the horizontal axis represents the frequency (0-8Hz), vertical axis represents the bandwidth (0-8Hz) and the color denotes the A' value. Note that the A' values are represented between 0 and 1 for this figure. Here values less than 0.5 indicate that this is not a prominent filter for subject independent drowsiness. This shows that there are variabilities in the way the Gabor filter outputs increase or decrease with acute drowsiness. If all the subjects have an increasing value for a Gabor filter output the A' will have a value higher than 0.5. If all the Gabor filter outputs for an action unit have a decreasing value then the A' will be again higher than 0.5. However if it is less than 0.5 and a value closer to 0 than the the feature might be a prominent feature for a subject dependent system. This can happen if the direction of increase or decrease of the filter output for the test subject is different from the increase or decrease direction of train subjects. Notice that here head roll (AU55, AU56), eye closure (AU45) and lip puckerer (AU18), (nose wrinkle) AU9 and (lid tightener) AU 7 are some of the most important action units in predicting acute drowsiness. The real filter looks the most prominent for these action units. One reason for this might be due to the shape of the real filter. While convolving filters we align the middle point of the filter with the time point t of the action unit signal. The real filter's shape in some frequencies and bandwidths look similar to the onset apex and offset transition of an action unit signal. In these cases the real filter looks like an averaging filter whereas the imaginary filter is similar to a difference filter. Imaginary and magnitude filters were also prominent with certain frequencies and bandwidths for the eye closure (AU45) action unit.

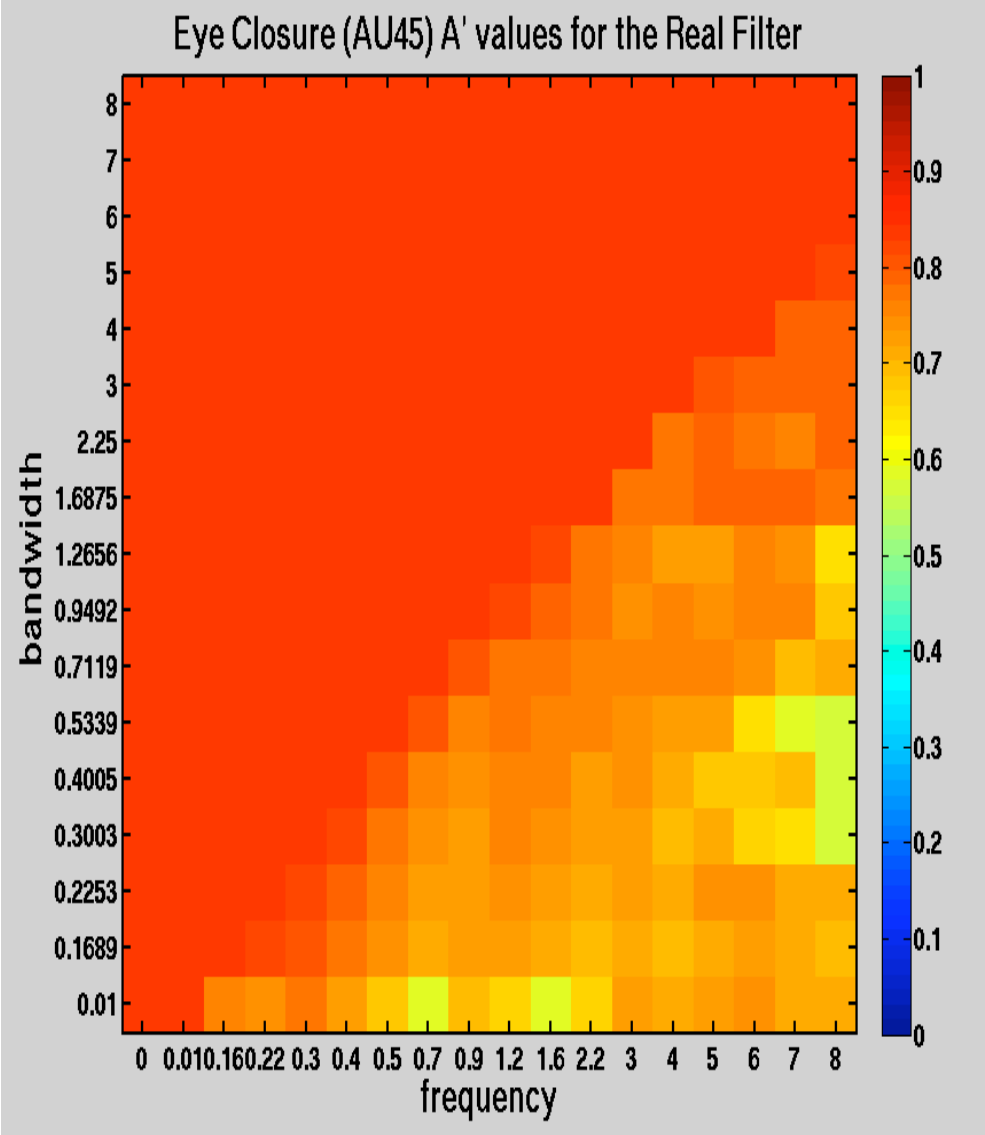

Figure 4.14: A' performances of Real Gabor Filters for the Eye Closure (AU45) action unit. The horizontal axis represents the frequency (0-8Hz), vertical axis represents the bandwidth (0-8Hz) and the color denotes the A' value. Note that the A' values are represented between 0 and 1 for this figure. Here values more than 0.5 closer to 1 indicate prominent filters of a subject independent system. A' values that are less than 0.5 and closer to 0 may indicate prominent filters that are subject dependent.

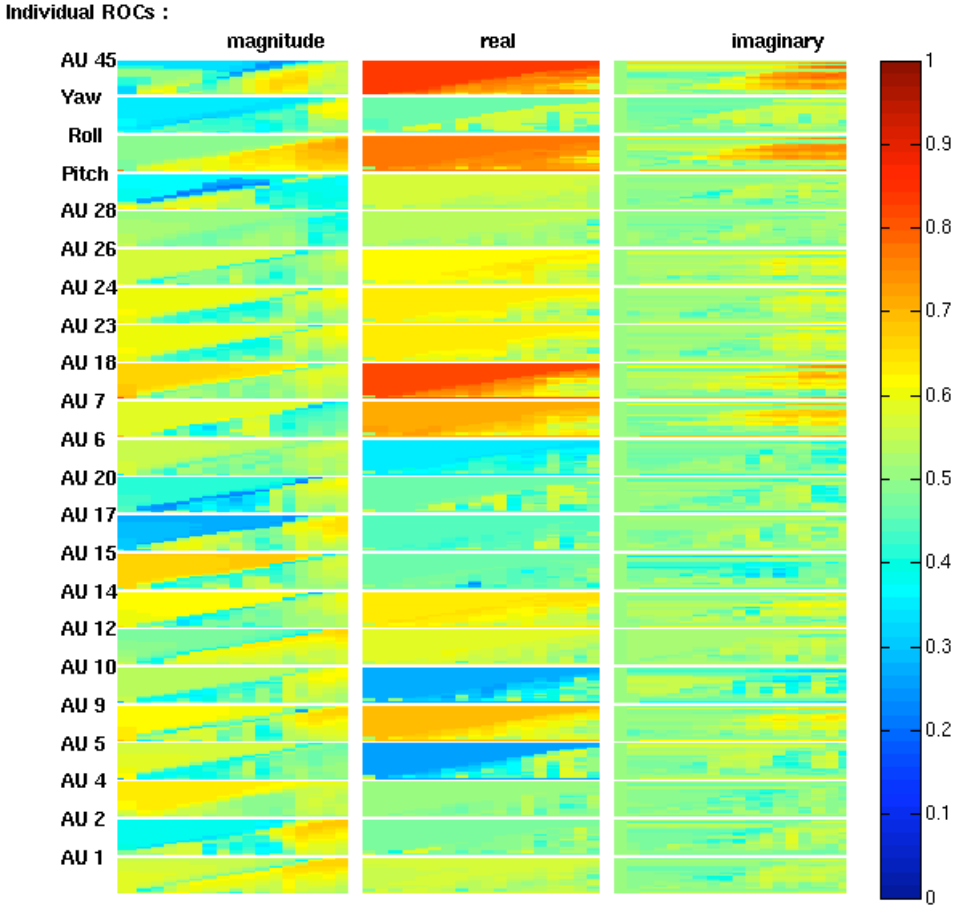

Figure 4.15: A' performances of individual Gabor Filters for all the action units. Each of the 66 (22x3) boxes above represent the A' performances for a specific action unit for either magnitude real or imaginary filter sets. For each box the horizontal axis represents the frequency (0-8Hz), vertical axis represents the bandwidth (0-8Hz) and the color denotes the A' value. Note that the A' values are represented between 0 and 1 for this figure. Here values more than 0.5 and closer to 1 indicate prominent filters a subject independent system. A' values that are less than 0.5 and closer to 0 may indicate prominent filters that are subject dependent.

4.5 Feature Selection

The goal for this analysis is to select relevant features from the 918 Gabor filter possibilities for each of the 22 action units. The number of data points for training a classifier for each subject is equal to the number of 10 second AD and MD segments and for nearly for all of the subjects there are more feature points than data points (See Table 4.4). The following scheme is followed: First the best feature was selected that has the best A' performance and then the next feature that achieves the best performance combined with the previous feature is chosen and the iteration continues in this fashion and chooses the next feature performing best combined with the previously selected features. All possible feature sets were trained with MLR by leaving one subject out at a time and using cross validation with generalization to novel test subjects. The obtained weights are tested on the test subject to estimate the predictive power of this features set on the test subject. An average of the test subject A's determined the dicriminability power of this feature set. At each iteration of the feature selection the feature that has the highest discriminability power combined with the previous features is included as a prominent feature. We tried 1 to 10 features with different regularization constants.

4.5.1 Eye Closure (AU45)

The results obtained with the eye closure action unit (AU45) are displayed in Figure 4.16. As the performance saturates with 10 features we stopped after picking 10 features for all of the action units. In the figure the horizontal axis represents the regularization constant for an L2 MLR model. The vertical axis displays the average A' over all test subjects for a specific feature count and regularization constant. The highest discriminability for eye closure was obtained with regularization constant 0 and using 10 features. Notice that the average filter was able to obtain an average A' of 0.83 for eye closure (AU45) whereas here the performance is higher approximately 0.90 with temporal Gabors using regularization 0 and 10 features.

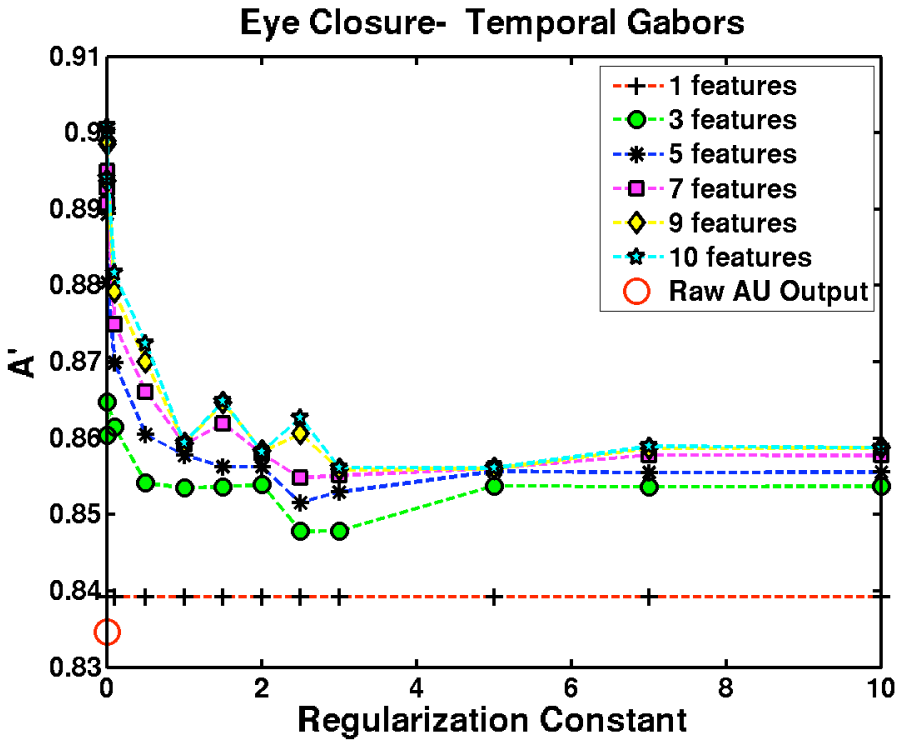

Figure 4.16: A' performance for Action Unit 45 (eye closure) versus regularization constant for different number of features selected with an iterative feature selection policy. The vertical axis displays the A' and the horizontal axis displays the regularization constant. Each colored graph displays different number of best features selected with iterative feature selection. Best A' is obtained with regularization constant zero and 10 features.

The features selected by the best model is displayed in Figure 4.17.

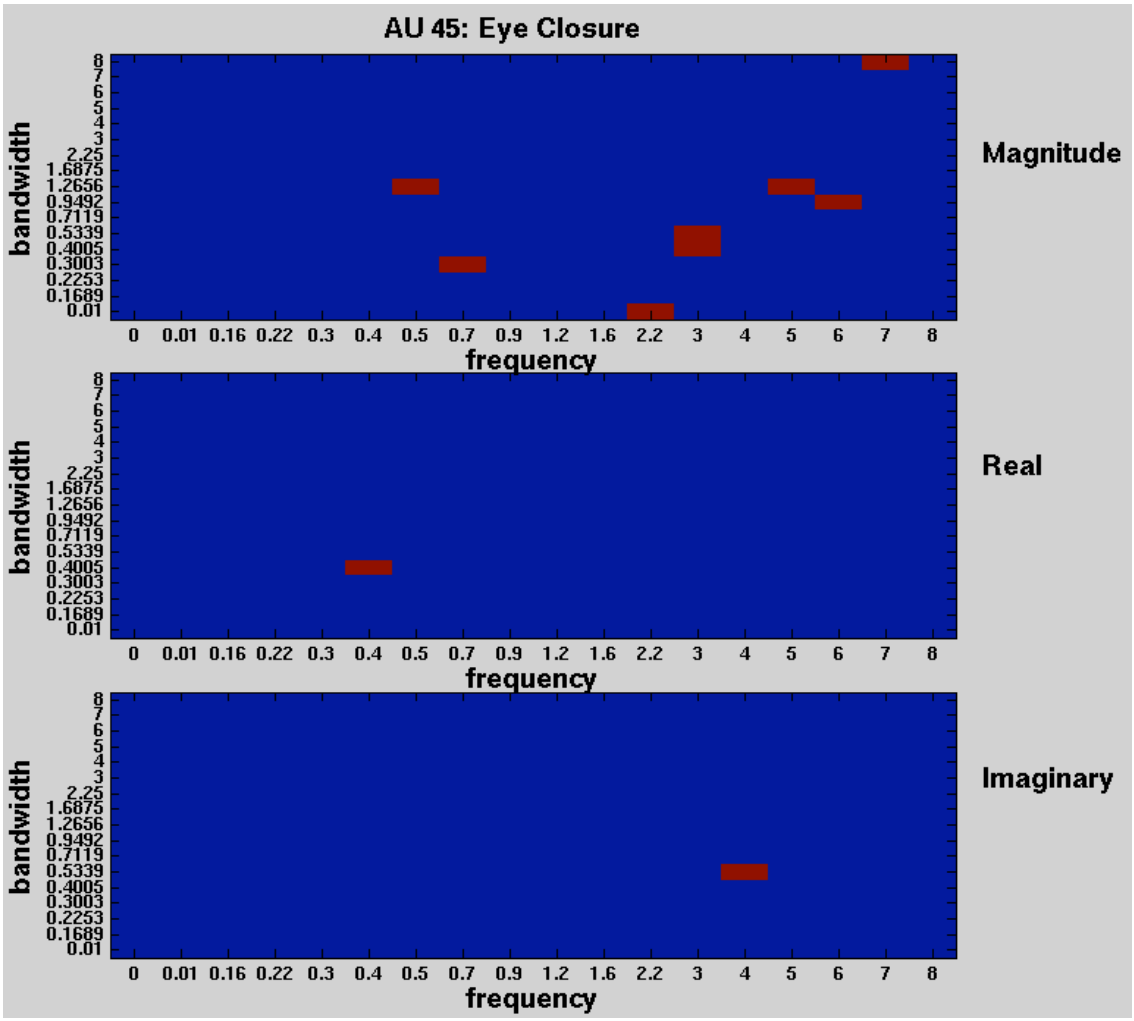

Figure 4.17: Features selected for the best model for eye closure action unit (AU45)

Figure 4.18 displays the best average A' (for the optimal regularization parameter) for eye closure as a function of the number of features. Each point (red dot) on the blue line displays the average A' over test subjects with the best performing regularization constant for a certain number of features. The green dots represent the standard error over the test subjects. The standard error is computed by dividing the standard deviation of subject performances

s with the square root of number of subjects n as displayed below.

$$SE = \frac{s}{\sqrt{n}} \tag{4.1}$$

As displayed in the figure the performance saturates towards 10 features. Therefore for this analysis we stopped the iteration for feature search at the tenth round.

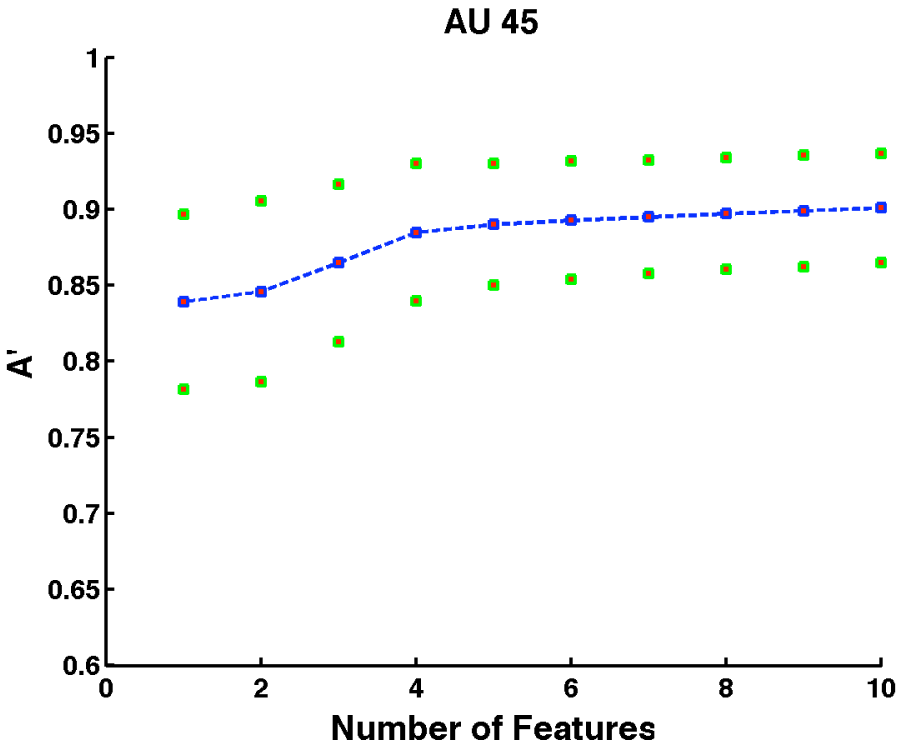

Figure 4.18: The blue line represent the best average A' among test subjects achieved for different number of features. Each point (red dot) on the blue line displays the average A' over test subjects with the best performing regularization constant for a certain number of features. The green dots represent the standard error over the test subjects.

4.5.2 Lip Pucker (AU18)

We next report the results for the second best action unit Lip Pucker (AU18). The A' performance averaged over test subjects for the Lip Pucker action

unit (AU18) is displayed in Figure 4.19. In the figure the horizontal axis represents the regularization constant for an L2 MLR model. The vertical axis displays the average A' over all test subjects for a specific feature count and regularization constant. The highest A' obtained for lip pucker (AU18) is 0.84 with regularization constant 0.1 and using 10 best features. Notice that the average filter is able to obtain an average A' of 0.82 for lip pucker (AU18) whereas here the performance is approximately 0.84 with regularization 0 and 10 features. The features selected by the best model is displayed in Figure 4.20.

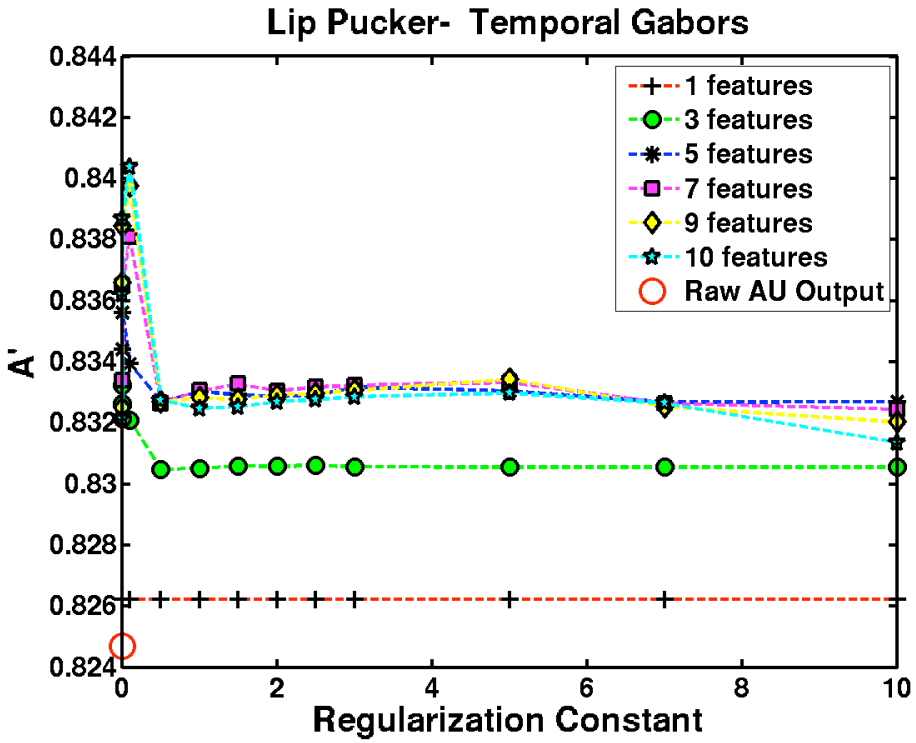

Figure 4.19: A' performance for Action Unit 18 (Lip Pucker) versus regularization constant for different number of features selected with an iterative feature selection policy. The vertical axis displays the A' and the horizontal axis displays the regularization constant. Each colored graph shows different number of best features selected with iterative feature selection. Best A' is obtained with regularization constant 0.1 and 10 features.

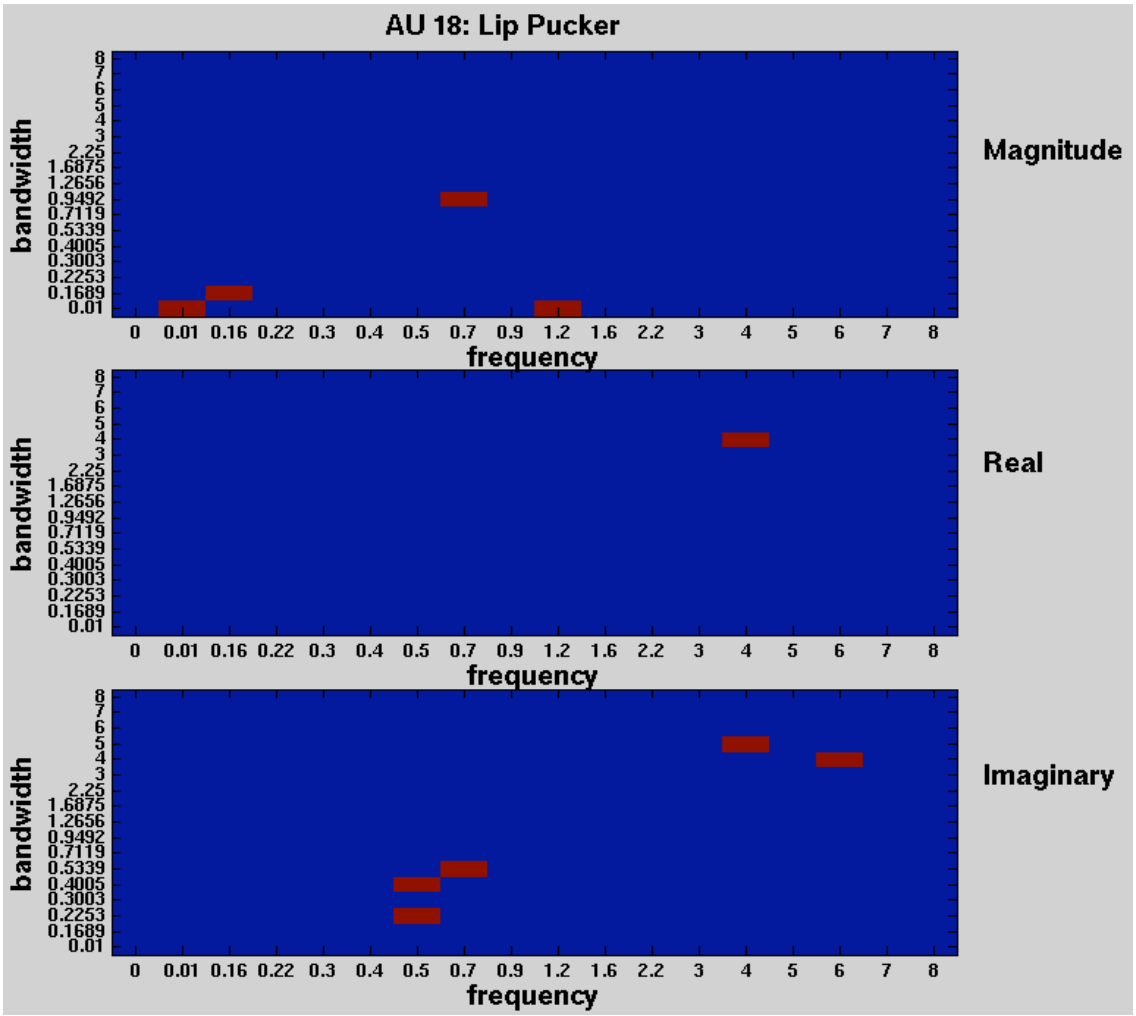

Figure 4.20: Set of features selected for the best model of lip pucker action unit (AU18). For the best model the regularization constant is 0.1.

Figure 4.21 displays the best average A' (for the optimal regularization parameter) for the lip pucker action unit (AU18) as a function of the number of features.

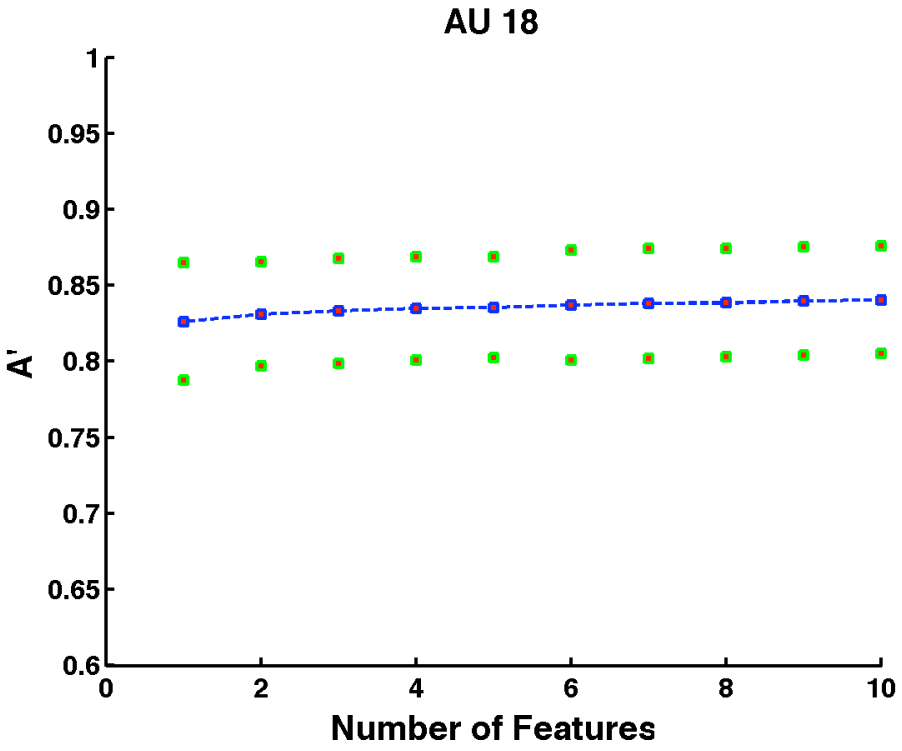

Figure 4.21: Best A' achieved as a function of different number of features for Lip Pucker action unit (AU18). The blue line represent the best average A' among test subjects achieved for different number of features. Each point (red dot) on the blue line shows the average A' over test subjects with the best performing regularization constant for a certain number of features. The green lines represent the standard error over the test subjects.

4.5.3 Head Roll (AU55-AU56)

We next report the results for the third best action unit Head Roll (AU55-AU56). The results obtained with the Head Roll action unit (AU55-AU56) for a person independent action unit is reported in Figure 4.22. In the figure the horizontal axis represents the regularization constant for an L2 MLR model. The vertical axis shows the average A' over all test subjects for a specific feature count and regularization constant. The highest A' performance for head roll was 0.81 obtained with a regularization constant term of 0.5 and 8 features. The figure displays the results with 1 to 10 features. Notice that

the average filter for Head Roll was able to obtain an average A' of 0.77 whereas here the performance is approximately 0.81 with regularization 0.5 and 8 features. The selected 8 best features is displayed in Figure 4.23.

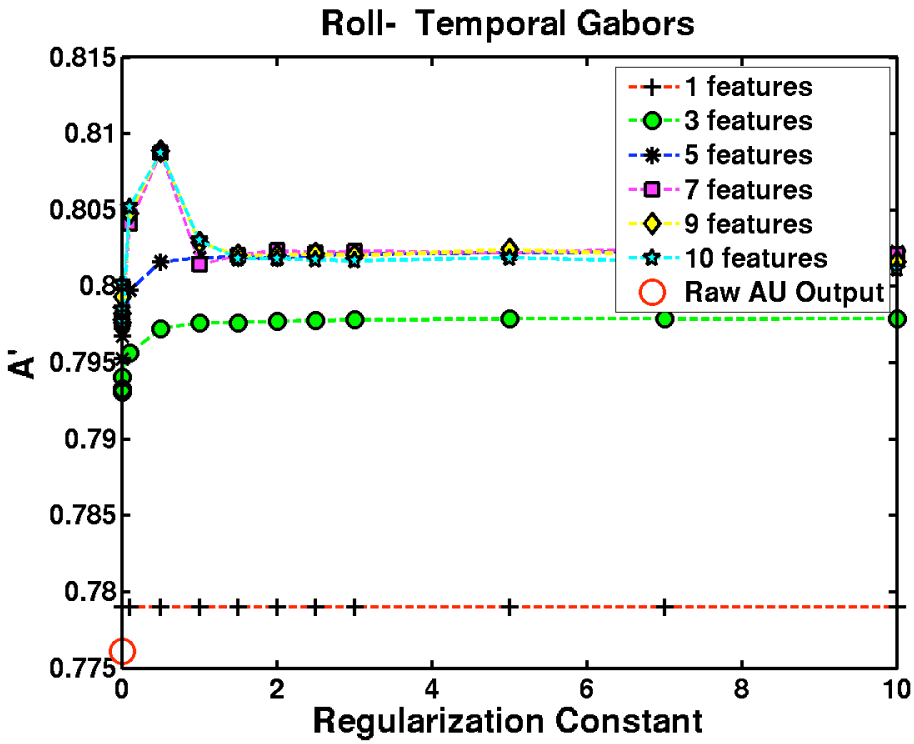

Figure 4.22: A' performance for Head Roll versus regularization constant for different number of features selected with an iterative feature selection policy. The vertical axis displays the A' and the horizontal axis displays the regularization constant. Each colored graph shows different number of best features selected with iterative feature selection. Best A' performance of 0.81 is obtained with regularization constant 0.5 and 8 features.

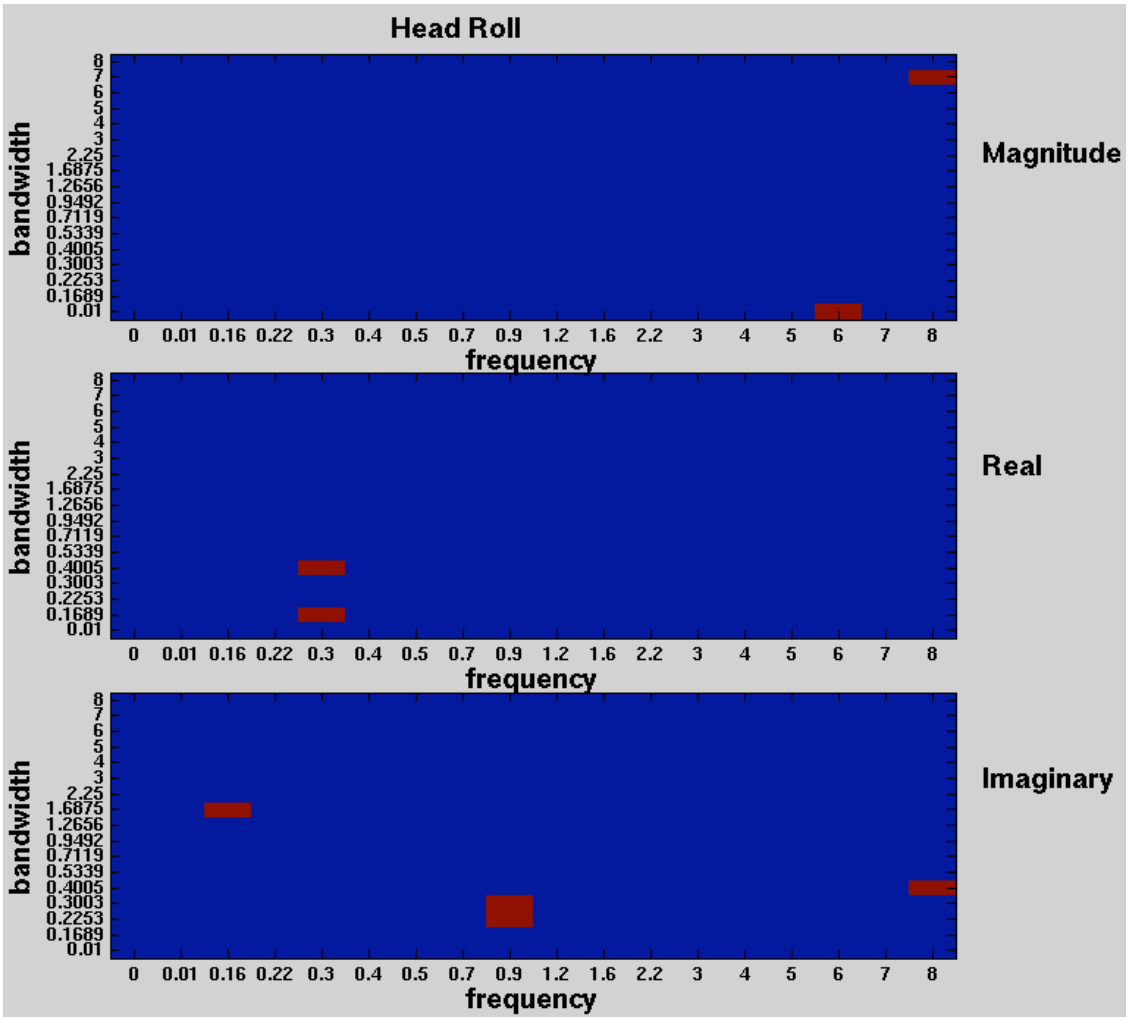

Figure 4.23: Selected features for the best model for Head Roll (AU55-AU56) action unit.

Figure 4.24 displays the best average A' (for the optimal regularization parameter) for head roll as a function of the number of features.

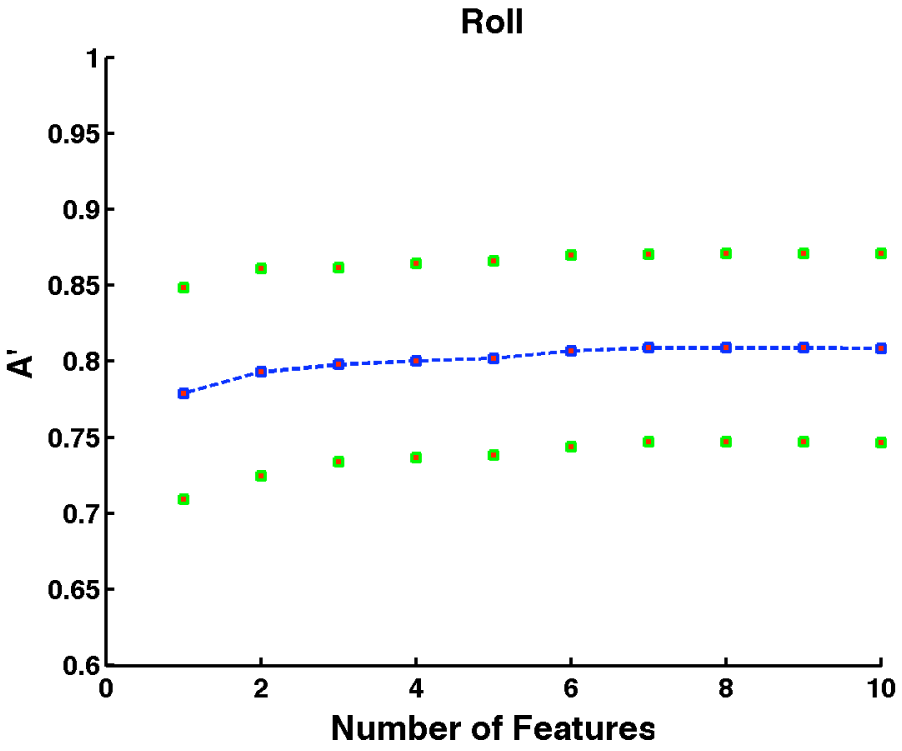

Figure 4.24: Best A' achieved with different number of features for Head Roll. The blue line represent the best average A' among test subjects achieved for different number of features. Each point (red dot) on the blue line shows the average A' over test subjects with the best performing regularization constant for a certain number of features. The green lines represent the standard error over the test subjects.

4.5.4 Lid Tighten (AU7)

We next report the results for the fourth best action unit Lid Tighten (AU 7) for a person independent action unit. The A' performance results averaged over test subjects for the Lid Tighten (AU7) is displayed in Figure 4.25. In the figure the horizontal axis represents the regularization constant for an L2 MLR model. The vertical axis shows the average A' over all test subjects for a specific feature count and regularization constant. The highest A' performance for lid tighten was 0.74 with temporal Gabors using regularization constant 2 and 10 features. Notice that the average filter for Lid Tighten was

able to obtain an average A' of 0.71 whereas here the performance is approximately 0.75 with regularization constant 2 and 10 features. The selected features for the best model is displayed in Figure 4.26.

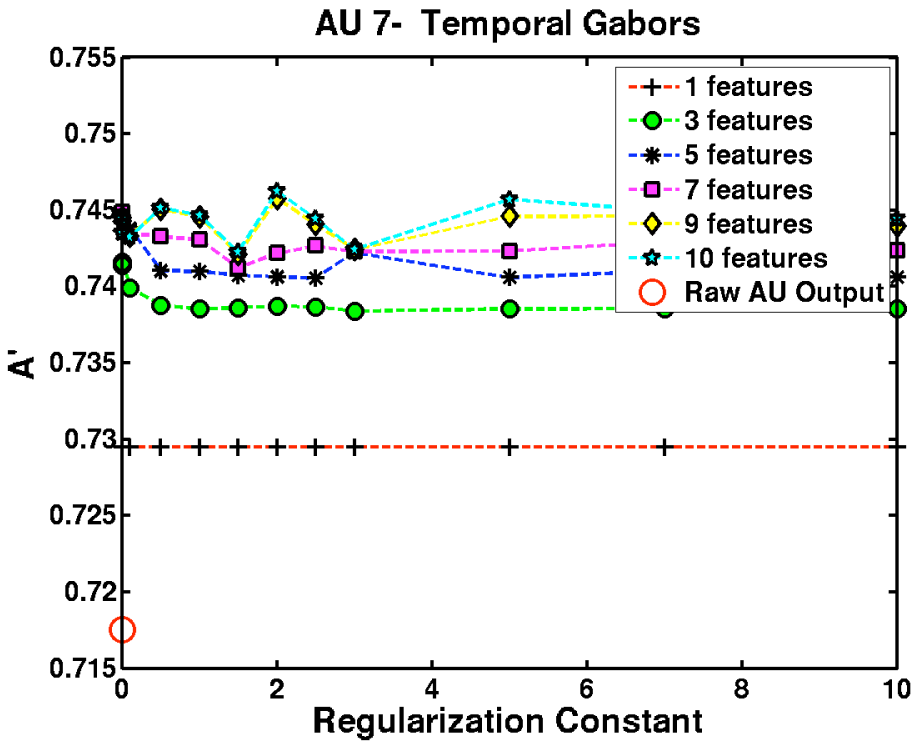

Figure 4.25: A' performance for Action Unit 7 (Lid Tighten) versus regularization constant for different number of features selected with an iterative feature selection policy. The y axis shows the A' and the x axis shows the regularization constant. Each colored graph shows different number of best features selected with iterative feature selection. Best A' of 0.74 is obtained with regularization constant 2 and 10 features.

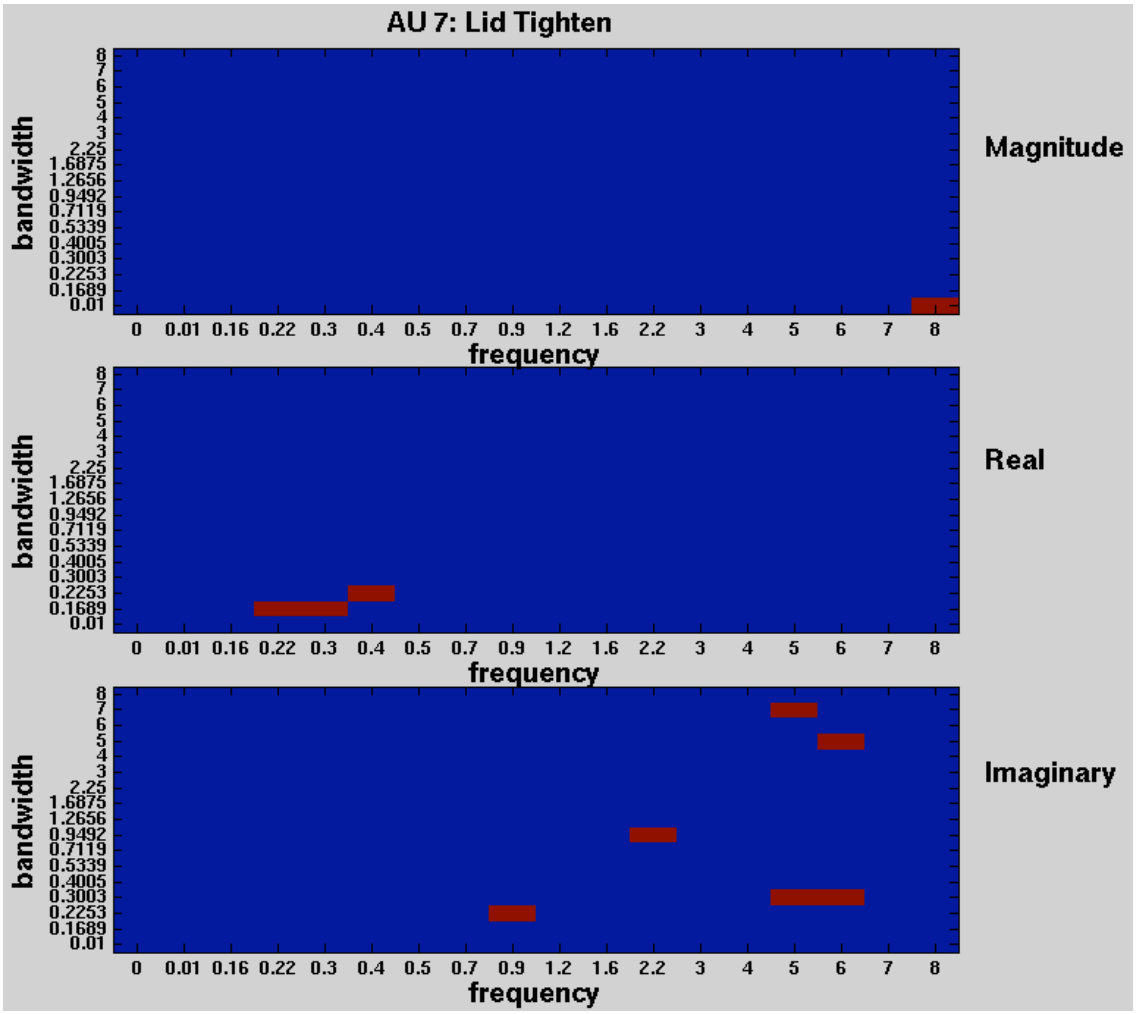

Figure 4.26: Best set of features selected for Lid Tighten (AU7) with regularization constant 2

Figure 4.27 displays the best average A' (for the optimal regularization parameter) for the lid tighten action unit (AU7) as a function of the number of features.

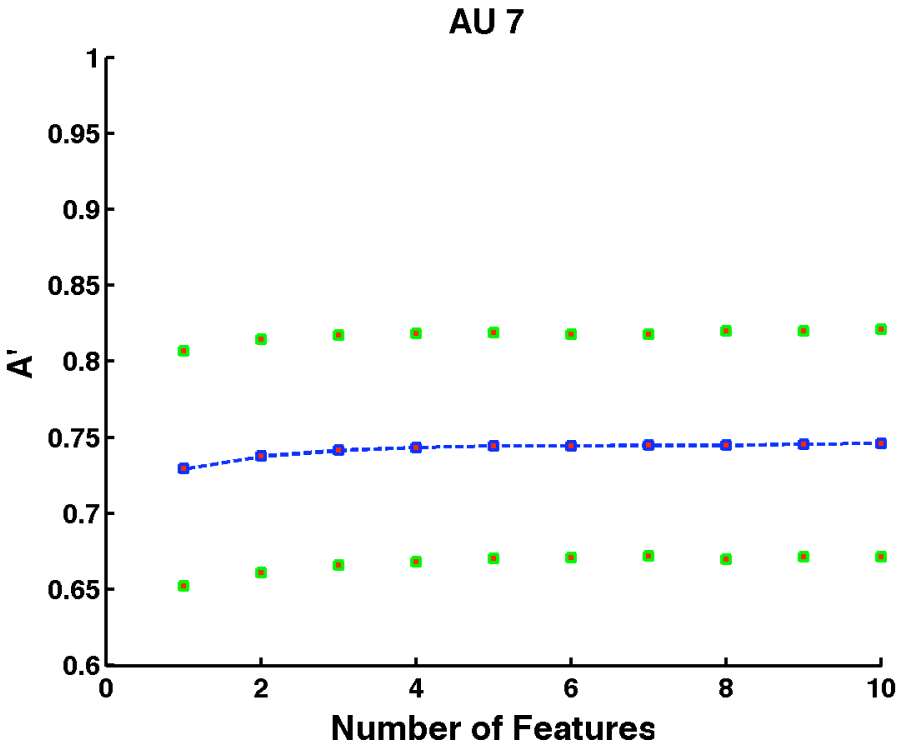

Figure 4.27: Best average A' (for the optimal regularization parameter) as a function of the number of features for Lid Tighten action unit (AU7). The blue line represent the best average A' among test subjects achieved for different number of features. Each point (red dot) on the blue line shows the average A' over test subjects with the best performing regularization constant for a certain number of features. The green lines represent the standard error over the test subjects.

4.5.5 Nose Wrinkle (AU9)

We next report the results for the fifth best action unit Nose Wrinkle (AU9). The A' performance results averaged over test subjects for the Nose Wrinkle is reported in Figure 4.28. In the figure the horizontal axis represents the regularization constant for an L2 MLR model. The vertical axis shows the average A' over all test subjects for a specific feature count and regularization constant. The highest A' performance for nose wrinkle (AU 9) was 0.80 obtained with a regularization constant term of 0.001 and 10 features. Notice

that the average filter for Head Roll was able to obtain an average A' of 0.69 whereas here the performance is approximately 0.81 with temporal Gabors using regularization 0.001 and 10 features. The features selected for the best model is displayed in Figure 4.29.

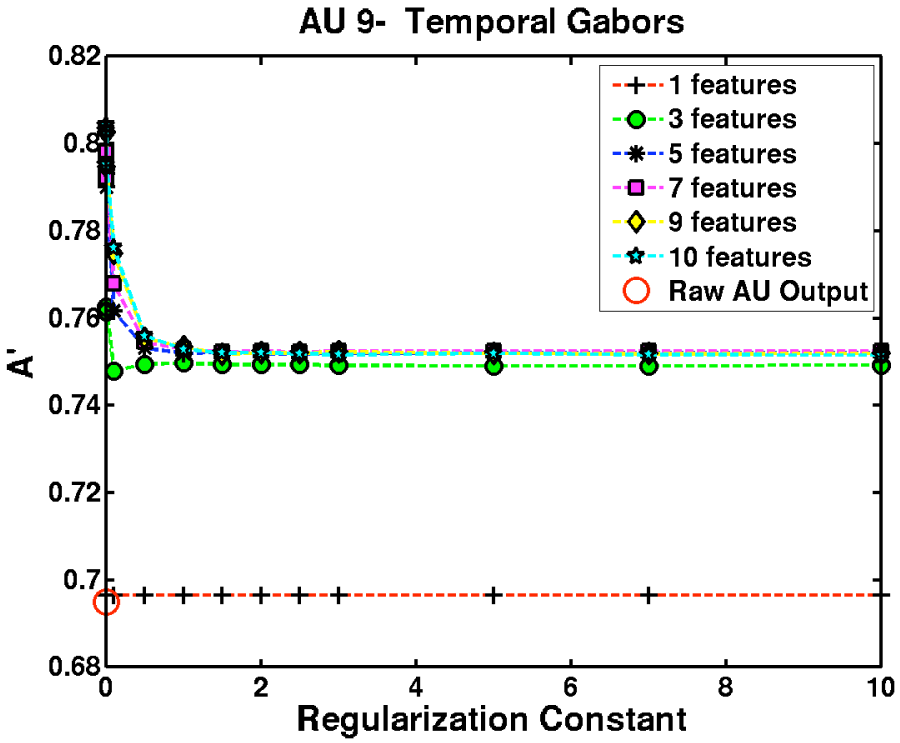

Figure 4.28: A' performance for Action Unit 9 (Nose Wrinkle) versus regularization constant for different number of features selected with an iterative feature selection policy. The vertical axis shows the A' and the horizontal axis shows the regularization constant. Each colored graph shows different number of best features selected with iterative feature selection. Best A' is obtained with regularization constant 0.001 and 10 features.

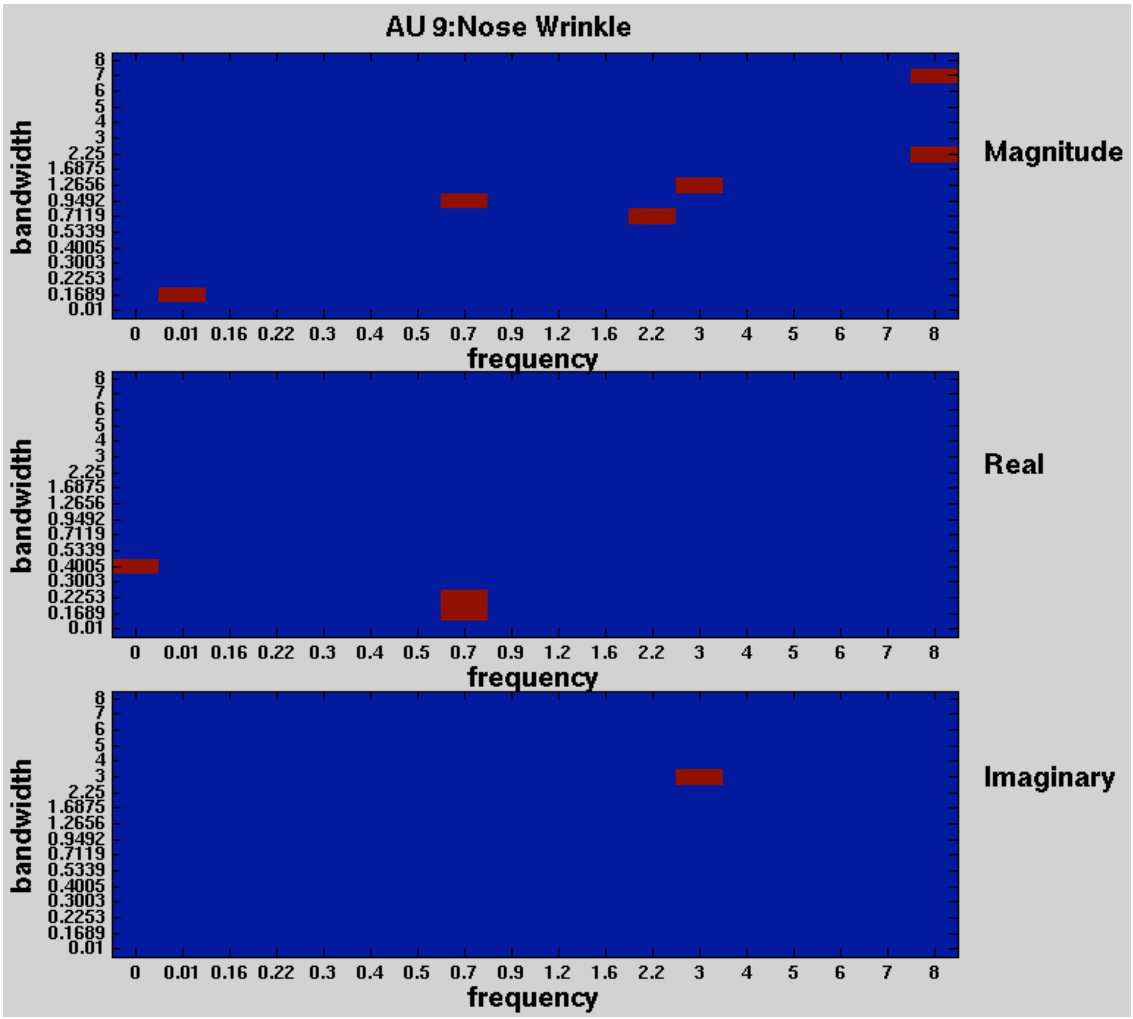

Figure 4.29: Features selected for the best model for Nose Wrinkle (AU9) action unit.

Figure 4.30 displays the best average A' (for the optimal regularization parameter) as a function of the number of features for the nose wrinkle action unit (AU9).

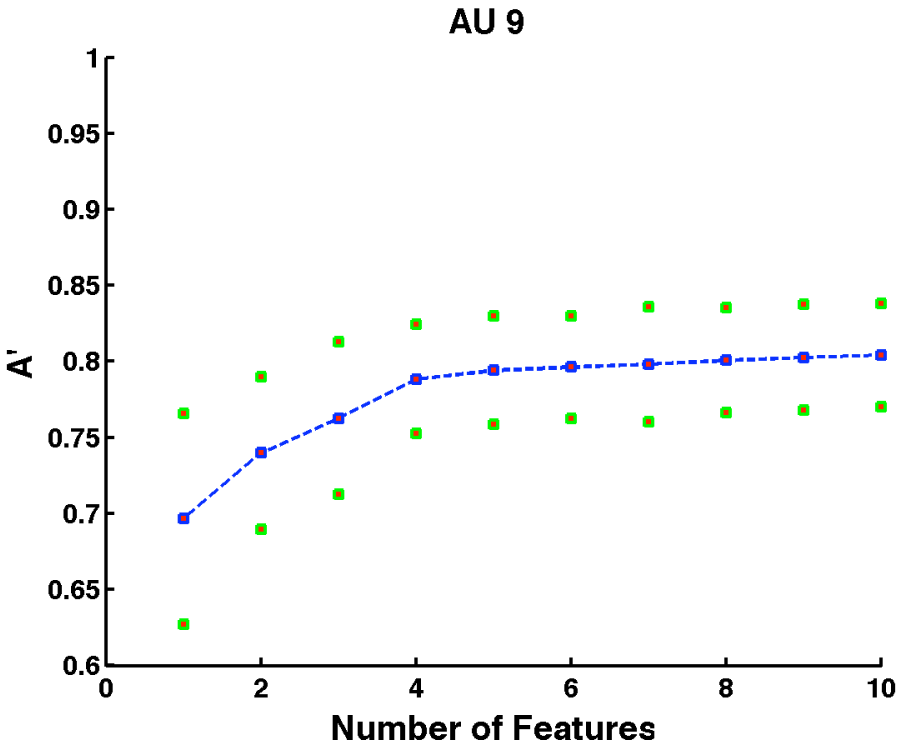

Figure 4.30: Best A' achieved with different number of features for Nose Wrinkle (AU9). The blue line represent the best average A' among test subjects achieved for different number of features. Each point (red dot) on the blue line shows the average A' over test subjects with the best performing regularization constant for a certain number of features. The green lines represent the standard error over the test subjects.

4.6 Combining Multiple Action Units

For this analysis the features of best performing five action units were combined to build a person independent drowsiness detector. These action units were as follows : Eye Closure (AU45), Lip Pucker (AU18), Head Roll (AU55-AU56), Lid Tighten (AU7), Nose Wrinkle (AU9). The best performing features for these action units obtained in the single AU models were combined to train a classifier. Except Head Roll achieving the best performance with 8 features all the other four action units had a set of 10 best features resulting in a total of 48 features. Iterative feature selection was performed over these

48 features. The following procedure was followed: First the best feature was selected that has the best A' performance and then the next feature that achieved the best performance combined with the previous feature was chosen and the next iteration continued in this fashion and chose the next feature performing best combined with the previously selected features. Up to 10 best features for 8 regularization constant ranging from 0 to 3 were explored with an L2 MLR model. All possible feature sets were trained with MLR by leaving one subject out at a time and using cross validation with generalization to novel test subjects. The obtained MLR weights for a train set are tested on the test subject to estimate the predictive power of this feature set on the test subject. An average of the test subject A"s determined the dicriminability power of this feature set. At each iteration of the iterative feature selection the feature that has the highest discriminability power combined with the previous features is included as a prominent feature. We tried feature counts from 1 to 10 with different regularization constants. The results for the selected best feature sets from a set of 48 possible features is displayed in Figure 4.31. As the performance mostly begins saturating with 10 features we stopped after picking 10 features for all of the action units. In the figure the horizontal axis represents the regularization constant for an L2 MLR model. The vertical axis shows the average A' over all test subjects for a specific feature count and regularization constant. The highest discriminability A' performance of .96 was obtained for the combined action units with temporal Gabors using regularization constant 0.01 and using 10 features. Note that the highest A' performance of .96 cuts the error in half when compared with the highest A' performance of .90 for eye closure action unit (AU 45). Hence combining other action units helps to build a more accurate person independent drowsiness detector.

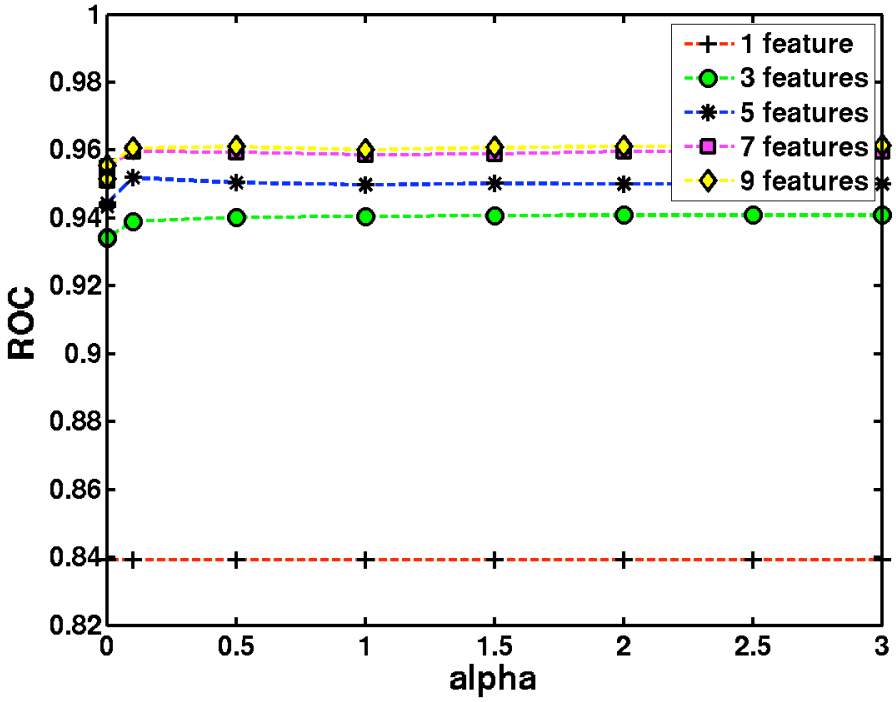

Figure 4.31: A' performance for 5 best action units combined versus regularization constant for different number of features selected with an iterative feature selection policy. The vertical axis shows the A' and the horizontal axis shows the regularization constant. Each colored graph shows different number of best features selected with iterative feature selection. Best A' of 0.96 is achieved with regularization constant 0.01 and 10 features.

4.7 Conclusions

In this study we notice that the markers for different levels of drowsiness change. When differentiating acute drowsiness from moderate drowsiness the most discriminative action units are: Eye Closure (AU45), Head Roll (AU55-AU56), Lip Pucker (AU18), Lid Tightener (AU7) and Nose Wrinkle (AU9). We found that some features are good at discriminating both alert from acutely drowsy and moderately drowsy from acutely drowsy. For example eye closure is one of these action units. However some features are good at discriminating alert from acutely drowsy but are not good at discriminating moderately drowsy from acutely drowsy. Yawning (AU26) and Smile (AU12) are examples of these action units. For this study head roll (AU55-AU56) was a good predictor for drowsiness supporting the preliminary study results in Study I. The subjects move their heads in the roll dimension as they get more drowsy. In this study consistent with Study I brow raise (AU2) increases in drowsy conditions in some subjects. However for this study one subject lowered his eyebrows dramatically (AU2 intensity decreased) in acute drowsiness state and raised his eyebrows in the moderate drowsiness condition. This resulted in AU2 to be non-informative for a person independent system. With UYAN-2 dataset we saw that there exists variabilities among subjects for the eye brow raise (AU2) signal. Using temporal properties of the signal and employing a Gabor representation increased the performance. Figure 4.32 displays a bar graph for the performance gain of Gabor filter outputs in comparison with raw action unit outputs for the 5 best performing action units . In this figure the red bars indicate the Gabor Filter output performances. For all the action units the performance with temporal Gabors was higher in comparsion with the raw action unit outputs. For example eye closure action unit performance A' increased from .83 to .90 after employing temporal Gabors. Finally by combining other action units the error could be cut in half in comparison to using only single action unit, eye closure (AU45), even for a harder task of discrimination of fine drowsiness. The performance increased from .90 using temporal Gabors (Eye closure) to .96 by employing all 5 action units using temporal Gabors. Temporal Gabor model achieved .96 A' by employing 5 action units whereas the raw action unit model achieved .92 A' with the same 5 action units. Thus the temporal Gabors increased the overall performance.

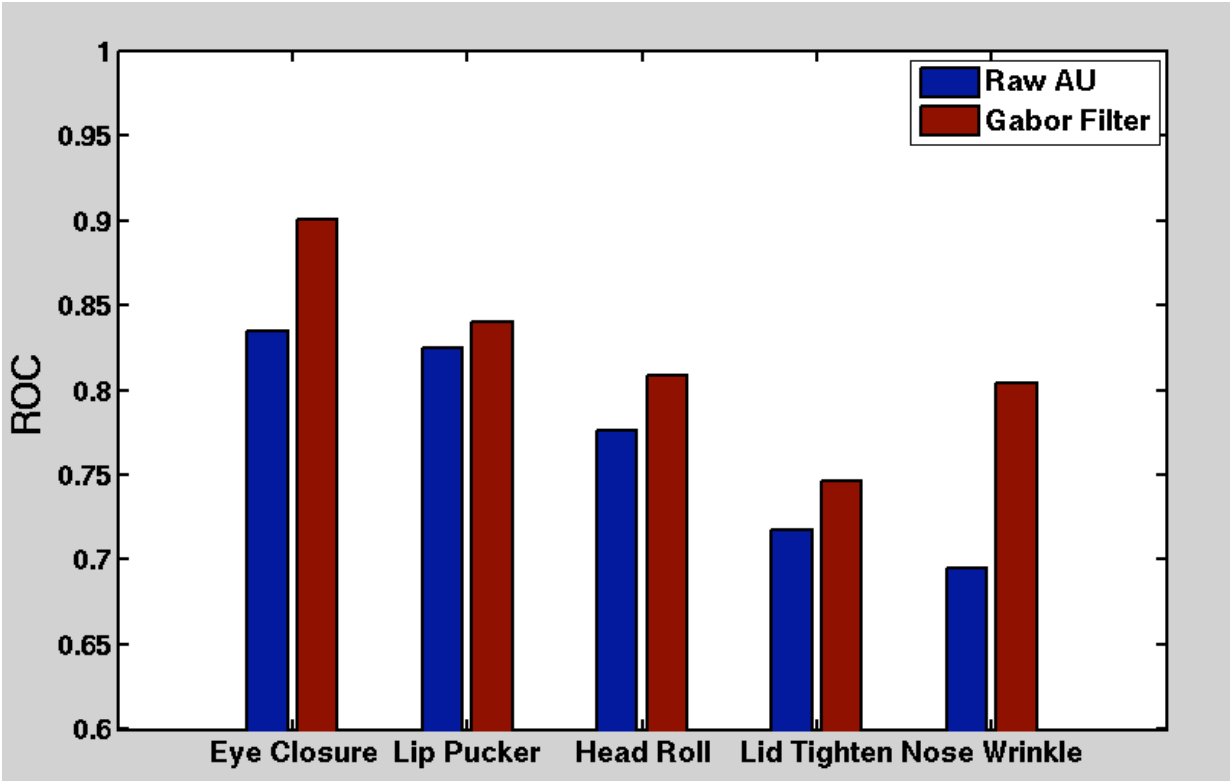

Figure 4.32: Bar graph displaying the performances for 5 best performing action units with the raw action unit output and the best model of Gabor Filter outputs.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This is the first study that uses a large set of spontaneous facial expressions for the detection of drowsiness. Previous approaches to drowsiness detection primarily make pre-assumptions about the relevant behavior, focusing on blink rate, eye closure, and yawning. Here we employ machine learning methods to datamine actual human behavior during drowsiness episodes. Spontaneous expressions have a different brain substrate than posed expressions. They also typically differ in morphology and dynamics. This study reveals that facial expressions are very reliable indicators of driver drowsiness and facial expressions can be used to do fine discrimination in the different levels of drowsiness and reliably predict the time to crash. In laboratory conditions computer vision expression recognition systems can be used to reliably detect drowsiness and predict crash with high reliability. Field studies are needed to evaluate the performance of these systems in actual driving environments. Spontaneous facial expressions under drowsiness are very different from posed expressions of drowsiness. We confirmed that some facial behavior previously reported in the literature (e.g., eye closure) is a reliable indicator of fatigue. However there are other expressions that are as reliable as eye closure and can be combined to improve detection. For discrimination of alertness vs drowsiness we found out that the Nose Wrinkle (AU 9), Eyebrow Raise (AU2), Eye Closure (AU 45), Chin Raise (AU 17), Yawn (AU 26), Head Roll were some of the most discriminative action units. When combined these facial expressions can discriminate alert from drowsiness with

0.98 area under the ROC curve performance for a temporal window of 12 seconds and saturates to 0.99 for 30 seconds or above. Yawning, actually occurred *less* often in the critical 60 seconds prior to a crash. We found that yawning does not tend to occur in the final moments before falling asleep. Finer discrimination, differentiating moderately drowsy from acutely drowsy, is a much more challenging task than differentiation of alert from drowsy. We found that the markers for different levels of drowsiness change : Notice that some of these markers may also be related to subject-wise differences as the two studies use different set of subjects. When differentiating acutely drowsy from moderately drowsy the most discriminative action units are :Eye Closure, Head Roll, Lip Puckerer, Lid Tightener, Nose Wrinkle. Some features are good at discrimination of both alert from acutely drowsy and moderately drowsy from acutely drowsy. For example eye closure is one of these action units. However some features are good at discriminating alert from acutely drowsy but are not good at discriminating moderately drowsy from acutely drowsy. Smile and yawning are examples of these action units. There are also subject-wise differences. In discriminating acutely drowsy from moderately drowsy brow raise(AU2) increases in acutely drowsy conditions in some subjects decreases in one subject and is neutral in others. With UYAN-2 dataset we saw that there are variabilities among subjects for the AU2. In the first study preliminary results indicated that Head Roll is an important measure for some subjects in discriminating alert from drowsy. For Study II head roll was a good predictor of drowsiness consistent with the finding that most subjects move their heads in the roll dimension as they get more drowsy. For Study II using temporal properties of the signal and using Gabor representation increased the performance. By combining other action units in addition to eye closure (AU45) the error can be cut in half even for a harder task of discrimination of fine drowsiness. Combining other action units also provides extra information for the detection of drowsiness in the case of occlusions of the eye such as sun-glasses.

5.2 Future Work

In this study we built reasonable classifiers using feature extraction methods such as Gabor features and iterative feature selection approaches. Here Gabor filters are used to analyze temporal structure but other methods such as ICA or power spectrum could also be used instead. In addition other feature

selection methods such as PCA could have been used instead of iterative feature selection method. In the future these feature extraction and selection methods are planned to be further explored.

This study indicates that some of the action units like AU2 and AU10 is discriminative in increasing or decreasing intensity value directions for some subjects and non-discriminative for others. This information is planned to be used in the future for a person dependent system. We do not report the results of our study for adaptive drowsiness detection however person dependent approach was explored in a preliminary study. Individual differences in the way drowsiness is expressed are investigated and how to automatically adapt detection of drowsiness to individual drivers is researched partially. In the future the analysis will be further extended to understand the pros and cons of a person dependent, adaptive and a person independent drowsiness detection system.

We reported some preliminary results for coupling between eye closure and eye brow raise and head movement and steering signals for Study I. Coupling of behaviours is not studied for Study II. This is planned be explored in a future study.

In study I we classified segments into two classes such as acutely drowsy versus alert. In the future classes in Study II and Study I can be expanded into three classes as alert, moderately drowsy, and acutely drowsy and thus the two studies can be further compared. In addition for both studies the problem can be further extended and handled as a regression problem. Thus we plan to predict continuous measure such as time to crash using facial and movement measures.

The results obtained for both studies in this thesis might be subject or dataset dependent. In the future both studies need to be repeated with an expanded set of subjects to verify these results.

Although it is not reported in this thesis we also worked on brain waves and we could predict fatigue based on raw EEG signal. However we discovered that we are actually using the motion artifacts to predict drowsiness. Motion artifacts are a big unsolved problem in EEG. In the future new techniques can be explored to eliminate motion artifacts from EEG.

For the UYAN-2 upper torso movement measures are collected as a separate measure. This measure is planned to be analyzed in detail as a future work.

For the UYAN-2 dataset subjective measures of drowsiness is collected from two human labelers labeling the video in a continuous manner. The

video clips are labeled by two labelers by moving the sliding bar using arrow key while watching the video at full speed. The labels were discrete and ranged from -5 (very alert) to 5 (very drowsy). In the future, how humans perceive fatigue and how the human labels correlate with the the action units is planned to be investigated.

Bibliography

[1] Driver fatigue is an important cause of road crashes. *http://www.smartmotorist.com/traffic-and-safety-guideline/driver-fatigue-is-an-important-cause-of-road-crashes.html*.

[2] Driver State Sensor developed by seeingmachines Inc. *http://www.seeingmachines.com/product/DSS*.

[3] Electroencephalogram (eeg). *http://weirddreams.net/electroencephalogram-eeg/*.

[4] Regulatory impact and small business analysis for hours of service options. Federal Motor Carrier Safety Administration. Retrieved on 2008-02-22.

[5] Statistics related to drowsy driver crashes. *http://www.americanindian.net/sleepstats.html*.

[6] Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology. *Circulation*, 93(5):1043–1065, March 1996.

[7] Planque S. Lavergne C. Cara H. de Lepine P. Tarriere C. Artaud, P. An on-board system for detecting lapses of alertness in car driving. In *14th E.S.V. conference, session 2 - intelligent vehicle highway system and human factors Vol 1, Munich, Germany*, 1994.

[8] M.S. Bartlett, G. Donato, J.C. Hager, P. Ekman, and T.J. Sejnowski. Face image analysis for expression measurement and detection of deceit. In *Proceedings of the 6th joint symposium on Neural Computation*, pages 8–15, 1999.

[9] M.S. Bartlett, G. Littlewort, B. Braathen, T.J. Sejnowski, and J.R. Movellan. A prototype for automatic recognition of spontaneous facial actions. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15, pages 1271–1278, Cambridge, MA, 2003. MIT Press.

[10] M.S. Bartlett, G.C. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J.R. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia.*, 1(6) p. 22-35.

[11] G. Belenky, T.J. Balkin, D.P. Redmond, H.C. Sing, M.L. Thomas, D.R. Thorne, and N.J. Wesensten. *Sustained performance during continuous operations, The US army's sleep management system.* Elsevier Science Ltd In Hartley, L.R. (Ed.) Managing Fatigue in Transportation. Proceedings of the Third International Conference on Fatigue and Transportation, Fremantle, Western Australia., Oxford, UK, 1998.

[12] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152. ACM Press, 1992.

[13] A. E. Burgess. Comparison of receiver operating characteristic and forced choice observer performance measurement methods. *Medical physics*, 22(5):643–655, May 1995.

[14] Mi-Kyeong Byeon, Sang-Whi Han, Hong-Ki Min, Yo-Seop Wo, Young-Bae Park, and Woong Huh. A study of hrv analysis to detect drowsiness states of drivers. In *BioMed'06: Proceedings of the 24th IASTED international conference on Biomedical engineering*, pages 153–155, Anaheim, CA, USA, 2006. ACTA Press.

[15] W.A. Cobb. Recommendations for the practice of clinical neurophysiology. *Elsevier*, 1983.

[16] C. Cudalbu, B. Anastasiu, R. Radu, R. Cruceanu, E. Schmidt, and E. Barth. Driver monitoring with a single high-speed camera and IR illumination. *Signals, Circuits and Systems, 2005. ISSCS 2005. International Symposium on*, 1:219–222 Vol. 1, July 2005.

[17] J. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20(10):847–856, 1980.

[18] D F Dinges and M M Mallis. Managing fatigue by drowsiness detection: Can technological promises be realised. In *Proceedings of the Third International Conference on Fatigue and Transportation, Fremantle, Western Australia. Elsevier Science Ltd*, pages 15–17. Elsevier, 1998.

[19] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974–989, 1999.

[20] D.Royal. National survey of distracted and drowsy driving attitudes and behavior: 2002 Volume I Findings Report. The Gallup Organization, 2003.

[21] P. Ekman. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. W.W. Norton, New York, 3rd edition, 2001.

[22] P. Ekman and Rosenberg E.L. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System*. Oxford University Press, New York, 2005.

[23] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA, 1978.

[24] S. Elsenbruch, M. J. Harnish, and W. C. Orr. Heart rate variability during waking and sleep in healthy males and females. *Sleep*, 22:1067–1071, Dec 1999.

[25] J. C. Falmagne. *Elements of psychophysical theory.* Oxford: Oxford University Press,, 1985.

[26] Movellan J.R. Fasel I., Fortenberry B. A generative framework for real-time object detection and classification. *Computer Vision and Image Understanding.*, 98, 2005.

[27] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28:2000, 1998.

[28] R. Grace, V.E. Byrne, D.M. Bierman, J.-M. Legrand, D. Gricourt, B.K. Davis, J.J. Staszewski, and B. Carnahan. A drowsy driver detection system for heavy vehicles. *Digital Avionics Systems Conference, 1998. Proceedings., 17th DASC. The AIAA/IEEE/SAE*, 2:I36/1–I36/8 vol.2, Oct-7 Nov 1998.

[29] David Marvin Green and John A. Swets. *Signal detection theory and psychophysics [by] David M. Green [and] John A. Swets.* Wiley New York,, 1966.

[30] Haisong Gu and Qiang Ji. An automated face reader for fatigue detection. In *FGR*, pages 111–116, 2004.

[31] Haisong Gu, Yongmian Zhang, and Qiang Ji. Task oriented facial behavior recognition with selective sensing. *Comput. Vis. Image Underst.*, 100(3):385–415, 2005.

[32] Riad I. Hammoud and Harry Zhang. Alertometer: Detecting and mitigating driver drowsiness and fatigue using an integrated human factors and computer vision approach. In *Passive Eye Monitoring*, Signals and Communication Technology, pages 301–321. Springer Berlin Heidelberg, 2008.

[33] L. Hartley, T. Horberry, N. Mabbott, and G P. Krueger. Review of fatigue detection and prediction technologies. Technical report, National Road Transport Commission, 2000.

[34] Kim Hong, Chung. Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *International Journal of Industrial Ergonomics.*, Volume 35, Issue 4, April 2005, Pages 307-320.

[35] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the fourth IEEE International conference on automatic face and gesture recognition (FG'00)*, pages 46–53, Grenoble, France, 2000.

[36] Wang J.S. Knipling, R.R. Revised estimates of the US drowsy driver crash problem based on general estimates system case reviews. In *Proceedings of the 39th Annual Association for the Advancement of Automative Medicine*, pages 451–466, Chicago, IL, 1995.

[37] C. Lavergne, P. De Lepine, P. Artaud, S. Planque, A. Domont, C. Tarriere, C. Arsonneau, X. Yu, A. Nauwink, C. Laurgeau, J.M. Alloua, R.Y. Bourdet, J.M. Noyer, S. Ribouchon, and C. Confer. Results of the feasibility study of a system for warning of drowsiness at the steering wheel based on analysis of driver eyelid movements. In *Proceedings of the Fifteenth International Technical Conference on the Enhanced Safety of Vehicles*, Melbourne, Australia, 1996.

[38] Chin-Teng Lin, Ruei-Cheng Wu, Tzyy-Ping Jung, Sheng-Fu Liang, and Teng-Yi Huang. Estimating driving performance based on EEG spectrum analysis. *EURASIP J. Appl. Signal Process.*, 2005:3165–3174, 2005.

[39] G. Littlewort, M.S. Bartlett, I. Fasel, J. Susskind, and J.R. Movellan. Dynamics of facial expression extracted automatically from video. In *IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Face Processing in Video*, 2004.

[40] Gwen C. Littlewort, Marian Stewart Bartlett, and Kang Lee. Faces of pain: automated measurement of spontaneousallfacial expressions of genuine and posed pain. In *ICMI '07: Proceedings of the 9th international conference on Multimodal interfaces*, pages 15–21, New York, NY, USA, 2007. ACM.

[41] G. Littlewort-Ford, M.S. Bartlett, and J.R. Movellan. Are your eyes smiling? detecting genuine smiles with support vector machines and gabor wavelets. In *Proceedings of the 8th Joint Symposium on Neural Computation*, 2001.

[42] Stilwell-Morecraft KS Louie JL, Herrick JL. Cortical innervation of the facial nucleus in the non-human primate: a new interpretation of the effects of stroke and related subtotal brain trauma on the muscles of facial expression. *Brain*, 124(1):176–208, January 2001.

[43] Javier R. Movellan. Tutorial on gabor filters. *http://mplab.ucsd.edu/tutorials/gabor.pdf* .

[44] Javier R. Movellan. Tutorial on multinomial logistic regression. *http://mplab.ucsd.edu/tutorials/MultivariateLogisticRegression.pdf* .

[45] J.F. O'Hanlon, B.J. Oborne, and J.A. Levis (Eds). *Critical Tracking Task (CTT) Sensitivity to Fatigue in Truck Drivers*. Academic Press Inc., London, MA, 1981.

[46] Karl F. Van Orden, Tzyy-Ping Jung, and Scott Makeig. Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology*, 2000 Apr;52(3):221-40.

[47] W. E. Rinn. The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, 95(1):52–77, 1984.

[48] A.; Senaratne-R.; Halgamuge S. Saeed, I.; Wang. Using the active appearance model to detect driver fatigue. In *Information and Automation for Sustainability, 2007. ICIAFS 2007. Third International Conference*, pages 124–128, 4-6 Dec. 2007.

[49] Saeid Sanei and J. A. Chambers. *EEG Signal Processing*. Wiley-Interscience, September 2007.

[50] Rajinda Senaratne, David Hardy, Bill Vanderaa, and Saman Halgamuge. Driver fatigue detection by fusing multiple cues. In *Advances in Neural Networks ISNN 2007*, volume 4492 of *Lecture Notes in Computer Science*, pages 801–809. Springer Berlin / Heidelberg, 2007.

[51] Y. Takei, Y. Furukawa. Estimate of driver's fatigue through steering motion. In *Man and Cybernetics, 2005 IEEE International Conference*, Volume: 2, On page(s): 1765- 1770 Vol. 2.

[52] Johannes van den Berg and Ulf Landstrom. Symptoms of sleepiness while driving and their relationship to prior sleep, work and individual characteristics. *Transportation Research Part F: Traffic Psychology and Behaviour*, 9(3):207 – 226, 2006.

[53] Eric Wahlstrom, Osama Masoud, Nikos Papanikolopoulos, and Senior Member. Vision-based methods for driver monitoring. In *IEEE Intelligent Transportation Systems Conf*, pages 903–908, 2003.

[54] Jacob Whitehill, Gwen Littlewort, Ian Fasel, Marian Bartlett, and Javier Movellan. Toward practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(2), 5555.

[55] D.Mayhew W.Vanlaar, H.Simpson and R. Robertson. Fatigued and drowsy driving: Attitudes, concern and practices of Ontario Drivers. Technical Report. Traffic Injury Research Foundation, 2007.

[56] Iizuka H. Yanagishima-T. Kataoka Y. Seno T. Yabuta, K. The development of drowsiness warning devices. In *Proceedings 10th International Technical Conference on Experimental Safety Vehicles, Washington, USA.*, 1985.

[57] Xun Yu. Real-time nonintrusive detection of driver drowsiness. Technical Report CTS 09-15, Intelligent Transportation Systems Institute, 2009.

[58] Zutao Zhang and Jia shu Zhang. Driver fatigue detection based intelligent vehicle control. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 1262–1265, Washington, DC, USA, 2006. IEEE Computer Society.