Stony Brook University

Exploring the ERP and Facial Response to Cocaine and Emotional Cues

A Thesis Presented

by

Hojin Choi

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Computer Science

Stony Brook University

August 2014

Stony Brook University

The Graduate School

Hojin Choi

We, the thesis committee for the above candidate for the

Master of Science degree, hereby recommend

acceptance of this thesis.

Dimitris Samaras
Associate Professor, Department of Computer Science

Gregory Zelinsky
Associate Professor, Department of Psychology

Samir R. Das
Professor, Department of Computer Science

This thesis is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Thesis

Exploring the ERP and Facial Response to Cocaine and Emotional Cues

by

Hojin Choi

Master of Science

in

Computer Science

Stony Brook University

2014

With the advancement of affective computing technologies, spontaneous and complex emotions that were difficult or even impossible to recognize in the past are rapidly becoming the focus of many research. One intriguing example of these non-basic emotions is drug craving, a sophisticated concept which neuroscientists have studies for a long time. Nevertheless, it is not yet known whether craving is accompanied by a characteristic expressive behavior when it occurs, and only little work has been done on exploring the relationship between craving and facial expressions.

In this thesis, we assume that drug-related stimuli can induce craving-related emotional states in drug users, and study the neurobiological and expressive behavior during these states by means of event-related potential (ERP) analysis and facial expression recognition. We design a passive image viewing task where cocaine users and control subjects watch neutral, drug-related, and non-drug emotional images. We record both electroencephalography (EEG) signals and the frontal face in real time to acquire a dataset that contains the subjects' response to multiple emotional cues.

We first aim to establish a baseline by showing that there is a meaningful difference in ERP values between the emotional categories. Individuals with cocaine use disorder (CUD) are expected to display higher magnitude of ERP for cocaine-related cues than for neutral cues. We show that, although to a weaker degree than previous studies, cocaine users react somewhat differently from healthy individuals when exposed to drug stimuli. After that, we use person-specific tracking algorithm to register various facial points of interest. We extract geometric and appearance features from the tracked points and shape to build feature vectors that can capture

the possible occurrence of craving-related behavior. We use several methods of classification and try to distinguish drug cue-induced facial response from other facial expressions.

Finally, we discuss the limitations of the current task design and the video dataset. Our results on ERP and facial activity show that our search for the possible expressive behavior correlated with craving is still inconclusive. We propose how the experiment could be improved in order to set a starting point for future research and make the task of facial expression recognition easier.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Recent years have witnessed significant advancements in the area of Affective Computing (AC). Human emotions play a huge role in how we react to our surroundings and what kind of decisions we make. The desire to understand this mechanism and to establish a standardized and solid model of emotions has attracted researchers from many fields including neuroscience, psychology and computer science. As a result, AC research has found promising applications in many domains, including affective-sensitive devices, mental health, education and gaming.

Since Darwin conducted his research to explore emotions in a scientific way for the first time [1, 2], modeling emotions as a set of behaviors and expressions has become one of the prominent approaches to affection recognition [3]. Early studies have discovered some evidence that there is a small set of prototypical emotional facial expressions that are universal across different cultures [4, 5]. These emotions, which are surprise, happiness, sadness, fear, anger and disgust, are considered to be "basic" emotions in the sense that they are innate and cross-cultural [6]. Reactions such as head and shoulder movements, gaze direction, and especially facial expressions have been used in automatic systems to successfully distinguish between the basic emotions [7]. However, in recent years, significant amount of effort has been put into recognizing more complex emotional states such as fatigue [8], frustration [9], and pain [10]. These emotions are much more difficult to detect than the six basic emotions, since there is usually greater inter-person variability and the reactions are often more subtle.

One of the interesting cases of this non-basic emotional state is drug craving. In the history of addiction research, the term 'craving' has been used to refer to various things including liking, wanting, urges, desires, need, intention or compulsion to use [11], but in general, craving is defined as 'the subjective experience of a desire to use certain substance'. Drug craving has been the center of considerable attention over the past several decades, and it has been shown that drug-related cues can induce meaningful brain responses in substance users [12]. Despite these findings, there are no universal standards for measuring the state of craving as yet [13, 14]. Moreover, only a relatively few number of studies have discussed the relationship between craving and facial activity [15, 16, 17].

The goal of this work is to explore the cue-induced responses for cocaine and other emotional stimuli in cocaine users and control subjects. We will use features acquired from two different modalities: the event-related potential (ERP) measured by electroencephalography (EEG), and facial expressions detected in the frontal face. The role of neuroscience in affective computing has been growing thanks to the advancement of brain imaging technology, and features extracted from EEG have shown promise for classifying different emotional states [18, 19]. It is therefore reasonable to include both EEG and facial features in our experiment.

In chapter 2, we provide an overview of research in affective computing with a focus on facial expression and EEG analysis. Chapter 3 briefly describes the history of craving analysis and presents some of the relevant studies. Experiment design and setup is discussed in chapter 4, and the results are shown in chapter 5. Conclusions and future work are presented in chapter 6.

Chapter 2

An Overview of Affective Computing

Affective computing aims to develop systems that can detect human emotions, express what a human would perceive as an emotion, and ultimately "feel" actual emotions [20]. The increase in computing power and the development of powerful sensors such as high resolution webcam and Kinect have led to significant advancements in the field, especially in affection detection. Since accurate detection of emotions is crucial to building a successful affective system, and because emotions are conceptual quantities with large individual variations, affection detection remains to be a very challenging and interesting problem. In the following subsections, different modalities that are used in affection detection will be discussed.

2.1 Facial Expressions

Human face plays a great role in delivering thoughts, emotions and intentions during communication. We perceive the complex movements of numerous facial muscles as facial expressions, and the ability to quickly intercept and interpret these expressions is one of the most fascinating functionality of our brain. In order to mimic this ability, it is important to understand how facial expressions are formed and design a good way of quantifying different facial behaviors.

The Facial Action Coding System (FACS) developed by Ekman and Friesen [21] has provided researchers with the means to objectively measure facial motion. FACS describes the movement of different facial muscles in terms of "facial actions". Each independent motion of the face is coded as an Action Unit (AU). These action units, 46 in total, are combined with each other to define the characteristics of various facial expressions. Since manually coding a facial image or video is very time consuming, there has been considerable effort to automatize this process [22]. Automated facial expression recognition system has many promising applications: examples include depression detection [23], neuropsychiatric disorder detection [24], and indexing and searching of video contents.

Earlier works in facial recognition field mostly focused on the six basic emotions, and have used datasets consisting of facial expressions that are deliberate, or "posed", under controlled environment. However, it is apparent that natural, spontaneous expressions are quite different from posed expressions, both in terms of facial behavior [25] and neural activities [26]. Spontaneous expressions display more smooth and synchronized movements in the facial

muscles, whereas posed expressions are less synchronized and show more variable dynamics [25**,** 26**,** 27]. Therefore, more effort is being put into identifying spontaneously occurring expressive behavior, such as distinguishing between natural versus deliberate smiles [25] and real versus faked pain [28],

Also, more studies are moving beyond the six basic expressions and tackling more sophisticated everyday life emotions. Ji et al. [8] developed a complex Dynamic Bayesian Network (DBN) structure for detecting driver fatigue. Their model attempts to capture the relationship between physical and mental conditions of the driver, environmental factors that could influence fatigue (light, temperature, humidity, etc.), and sensory observations that result from fatigue. Ashraf et al. [10] explored machine learning approaches for recognizing pain expression from video. Dinges et al. [29] devised a system that discriminates between high and low levels of stress by examining facial expressions of subjects while they completed a set of computerized neurobehavioral tests. Kapoor et al. [9] presented a framework for detecting frustration from learners who are solving the Towers of Hanoi puzzle by combining facial expressions, body movement and skin conductance. Reed et al. [23], showed neutral, comedy and control video clips to individuals suffering from depression disorders, and found that those with current depressive symptomatology were more likely to express smile controls during smiles.

Although coding schemes such as FACS are universal and easy to understand, they require access to large amounts of labeled data. This might not be feasible when we need to recognize naturally occurring behavior in highly uncontrolled settings, since the number of labeled datasets for natural expressions is significantly smaller compared to that of posed datasets. Therefore, some of the studies have tried automatic temporal segmentation or prior clustering of face events. In the early work of Hoey [30], a hierarchical dynamic Bayesian network was used to learn the high-level dynamics of facial expressions in a weakly supervised manner. Bettinger et al. [31] tracked each frame of video sequence with Active Appearance Models (AAM) and clustered facial behavior by segmenting the trajectories of features into sub-sequences. De la Torre et al. [32] devised a framework for segmenting facial events in the mouth region with automated face registration. Zhou et al. [33] approached the problem of unsupervised temporal clustering across different individuals as a versatile energy minimization problem, and used an extension of kernel k-means to solve it.

In the following subsections, we will discuss two important problems that every automatic facial expression recognition algorithm must deal with: representing the facial data in an efficient and effective manner, and designing a classifier that can reliably distinguish one facial expression from another.

2.1.1 Representation of Facial Data

A good representation of facial structure and behavior is necessary in order for the facial recognition system to work. An ideal feature for face representation should be able to contain enough information to represent the facial regions of interest while not being too computation-intensive. Most of the existing facial expression analysis algorithms use two types of features:

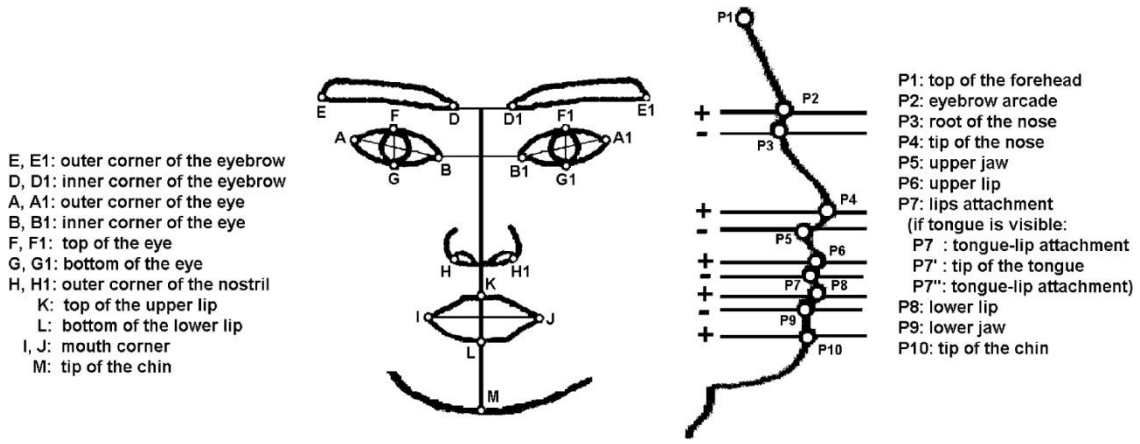

Figure 2.1: Geometry-based feature points used in the work of Pantic and Rothkrantz [34].

2D spatiotemporal features or 3D face models. 2D features can be largely classified into two types: geometric features and appearance features. Geometric features represent the shape and locations of specific facial landmarks such as eyes, brows and mouth corners. Facial feature points are extracted from these landmarks by means such as tracking or model fitting, and the points are used to define a feature vector that describes the behavior of that facial region. In contrast, appearance features aim to capture changes in skin texture of the face. They usually consist of different points of interest and texture descriptors, as well as local histograms with regional weights. One main drawback of 2D facial features is that they are dependent on the conditions of the dataset; a feature that is optimized for frontal face images wouldn't work properly for profile images. To deal with this issue, some studies have used 3D face model to describe the underlying mechanisms of facial deformation.

Geometric features are usually constructed by either detecting the shapes of the facial components (eyes, mouth, etc.) or tracking the locations of certain fiducial points (eye corners, tip of the nose, etc.). Since the displacement of these features is directly related to the facial muscle movement, many researchers have employed geometry-based descriptors as their tool for expression recognition. Pantic and Rothkrantz [34] extracted fiducial points from frontal and profile contours of face components (Figure 2.1). A similar definition of geometric features was used in a later study by Valstar and Pantic [35] where they proposed a fully automatic method for recognizing 22 AUs and modeling their temporal characteristics. Chang et al. [36] defined a face model with 58 feature points along the facial landmark contours and mapped the contour representation from high-dimensional space into a low-dimensional manifold. Kotsia and Pitas [37] proposed a system where a deformable grid is registered with the facial image region and the geometrical displacement of selected nodes is fed into a multiclass Support Vector Machine (SVM) classifier to recognize facial expressions.

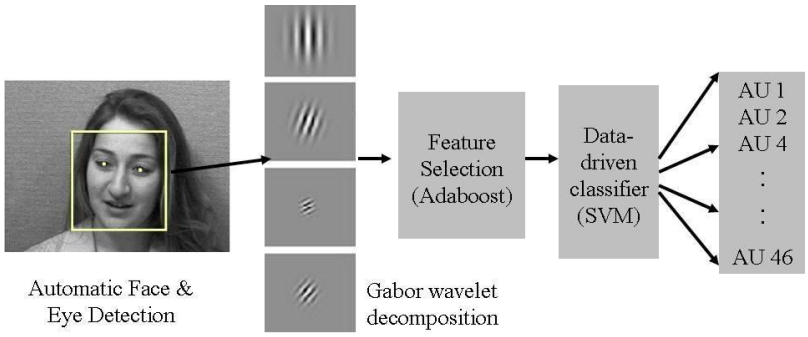

Figure 2.2: Automatic facial expression system by Littlewort et al. [28]

Although geometric features are intuitive and informative, they require robust face registration and tracking algorithms in order to be effective. Although this is not very difficult in a controlled dataset with little pose variation, problems arise in many real-world situations where face direction is not necessarily fixed and factors such as lighting condition could create noise. Therefore, most of the related works either use well-controlled image datasets such as the Cohn-Kanade database [38], or reinforce their geometric features with other modalities.
Appearance features, on the other hand, reflect the model the changes in the texture of the facial skin. For example, appearance-base features can effectively recognize transient facial components like wrinkles and nasolabial folds. Several earlier studies have used optical flow analysis to detect facial movements, such as Yacoob and Davis [39], Essa and Pentland [40], Hoey and Little [41], and Yeasin and Bullot [42]. Some of the appearance feature extraction methods were based on Active Appearance Models (AAM). For example, Lucey et al. [43] defined three facial representations derived from the AAM: similarity normalized shape, similarity normalized appearance, and shape normalized (canonical) appearance.

A widely adapted tool for appearance-based representation of the face is Gabor filter. Gabor filters are Gaussian kernel functions modulated by a sinusoidal wave. By using these filters, one can remove most of the variability in images caused by variation in lighting and contrast. It is known that the frequency and orientation representations of Gabor filter are similar to those of human visual cortical cells [44, 45]. Because of this property, Gabor filters have been applied to many areas including image analysis, face recognition and facial expression analysis [46, 47, 48, 49, 50, 51, 28]. For example, Littlewort et al. [28] designed a Gabor filter-based automatic AU detection system for distinguishing between posed and genuine pain expressions (Figure 2.2).

Local Binary Patterns (LBPs) are appearance features that were originally introduced as a texture classifier [52] and have gained great popularity in facial expression recognition community in recent years [53, 54]. The LBP operator encodes the local intensity structure around each pixel by comparing the center pixel value from its neighbors and then concatenating the binary comparison values into a binary number (Figure 2.3). A histogram of the LBP labels computed over a region is used as a texture descriptor. LBP is a nonparametric method and is simple to compute, compared to Gabor wavelet representation which can be time consuming and memory

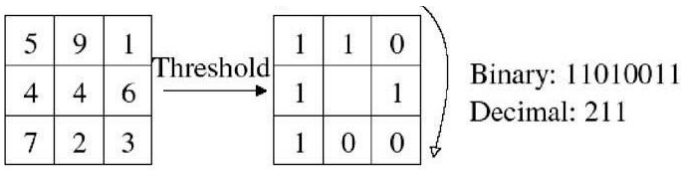

Figure 2.3: The basic LBP operator [55].

intensive. LBP features are also resilient against monotonic illumination changes in face images [55]. Shan et al. [53] examined different machine learning methods, including template matching, SVM, linear discriminant analysis and the linear programming technique, to perform facial expression recognition using LBP features. The authors conclude that, compared to Gabor wavelets, LBP features can be calculated fast and lie in low-dimensional feature space while still retaining discriminative facial information.

3D face models are aimed at tracking facial behavior under large amounts of head or body movements. If the 3D representation is robust and accurate enough, one can achieve view-independency and accurately measure and integrate head pose information into expression recognition architecture. Chang et al. [56] and Yin et al. [57] created a 3D facial expression database for facial expression recognition. Cohn et al. [58] estimated head motion by fitting a cylindrical model to the head region. Dornaika and Davoine [59] used the Candide 3D face model [60] to represent shape units from interpersonal differences and mesh deformations caused by AUs (Figure 2.4). Li et al. [61] presented a 3D face recognition method based on sparse representation and low-level geometric features.

2.1.2 Classifying Facial Expressions

Once the features defined and extracted from the face, a classifier is used on the features to distinguish between different facial expressions. Many types of classifiers have been proposed for facial expression recognition. These include neural network [62], SVM [63], Bayesian network [64] and rule-based classifiers [34, 65]. The expression recognition methods can be further divided into frame-based and sequence-based methods.

Frame-based expression recognition methods use the information of current image, with possible reference to a "neutral" image. Tian et al. [50] employed three-layer neural networks to recognize AU combinations present in images. Pantic and Rothkrantz [34] sampled the contours of facial components and extracted fiducial points from these contours. Based on the points, they recognized 32 AUs occurring alone or in combination by using rule-based classification. Although computationally simple, frame-based methods have the weakness of not being able to handle temporal dynamics of facial actions.

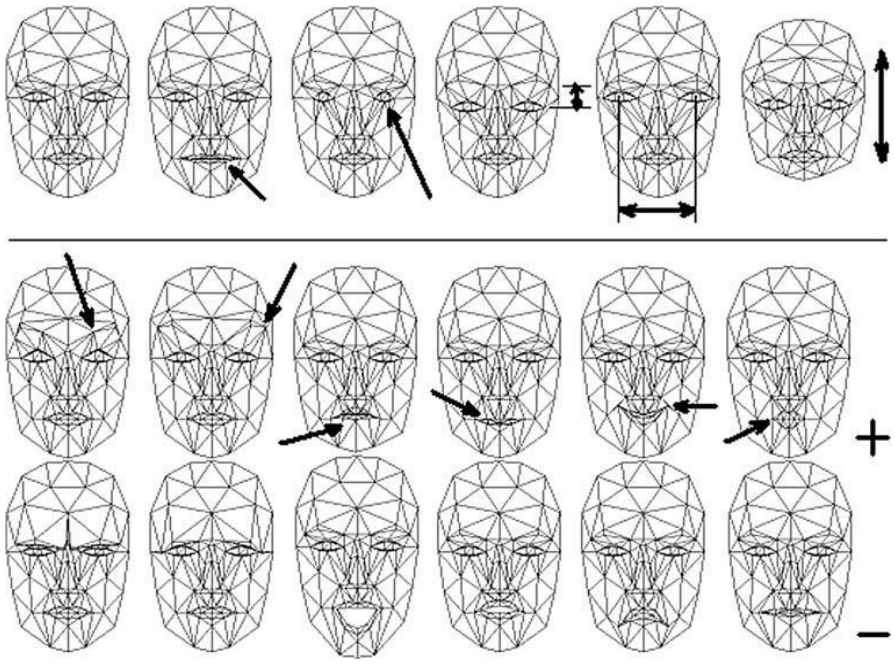

Figure 2.4: The Candid 3D model used in [61]. First row: Facial Shape units. Second and third rows: Positive and negative perturbations of Facial Action Units.

Sequence-based methods use temporal information of the image sequence to recognize the dynamic properties of facial expressions. Techniques such as Hidden Markov Models (HMMs), DBNs and rule-based classifiers were used for sequence-based classification.

HMM is appropriate for modeling temporal behavior of facial expressions. For example, Cohen et al. [64] evaluated automatic facial expression recognition using a multi-level HMM structure, However, HMM is mostly used for representing the temporal dynamics of facial expression on the emotion level. Dependencies between the lower-level features such as AUs are not specified but rather implicitly learned. For example, modeling AU or AU combinations with HMM would usually require training a specific HMM topology for each of those AUs.

In order to explicitly model complex relationship between facial features and action units in a straightforward manner, Dynamic Bayesian Network (DBN) are frequently used. Kaliouby and Robinson [66] implemented a system for inferring mental states from a video stream of facial expressions and head gestures in real-time by using a DBN model for the relationship between head movements, action units and mental states. Zhang and Ji [67] modeled the temporal relationships between the six basic expressions and the AUs with DBN. In their model, the lowest level is the sensory data layer containing actual information variables such as Brows, Lips, Eyelids, Mouth and Wrinkles. Intermediate levels consist of nodes that represent the presence of each AU, and nodes for modeling the conditional probabilities for AU combinations. The topmost level holds the node for expression classes.

2.2 Body Language and Posture

As a result of decreasing cost and increasing reliability of whole-body sensing technology, body expressions are becoming a very important modality in affective computing [68], since emotional expression usually occurs through combinations of multiple channels [69]. Mota and Picard [70] have used Tekscan's Body Pressure Measurement System (BPMS) to infer the affective states of a user in a learning environment. This work was extended by D'Mello and Graesser [71], who developed a system to detect boredom, confusion, frustration and delight from gross body movements during a learning task. Several studies have explored the relationship between face and body movements in spontaneous affective behaviors [72, 73]. Although posture-based affect recognition hasn't received as much attention as those based on facial expressions in the past, it is expected that the importance of body languages in emotion recognition will continue to grow.

2.3 Speech

Speech analysis has been a huge part of emotion research and human-computer interaction studies for a long time. As a result, meaningful relationships between emotional states and certain speech attributes have been discovered [74, 75]. For example, the pitch of the voice can act as an index into arousal. Most of the previous efforts were directed towards recognizing a subset of basic emotions from speech signals. However, recent studies have begun to focus on application-dependent affective states that are more complex [7], such as certainty [76], stress [77], trouble [78], and empathy [79]. Nonlinguistic vocalizations such as laughter, coughs cries have also begun to gather interest, due to the fact that listeners can detect some non-basic emotions like distress, anxiety and boredom rather accurately from these vocalizations [80].

An increasing number of studies are attempting to fuse visual and audio cues for better affect recognition [7, 81]. Zeng et al. [82] presented the framework of Multi-stream Fused HMM (MFHMM) to couple audio and visual streams and detect 11 cognitive/emotive states. Sebes et al. [83] designed a Bayesian network topology for combining audio and visual modalities in a probabilistic manner and used it to classify person-dependent emotions. Petridis and Pantic [84] proposed an audiovisual approach to discriminating laughter from speech while comparing between feature and decision level fusion.

2.4 Brain Imaging and EEG

The field of affective neuroscience has been proposing new techniques and methods to understand human emotional process over the last several decades. Affective neuroscience aims

to understand the underlying neural circuitry of emotion and mood, and to offer better perspectives on conceptualizing emotions [85]. The methods used in affective neuroscience include electrophysiology and imaging techniques such as fMRI. Since these methods can capture activities in the brain which might not be apparent in visual or audio channels, they are being used more frequently in affect recognition studies. It has been reported that picture stimuli with different valence (unpleasant-to-pleasant) and arousal (low-to-high) ratings can influence event-related potential (ERP) amplitudes at several processing stages [18]. Based on this finding, researchers have been able to show emotional and neutral image cues to individuals and distinguish the ERP features between the two [86].

Some studies in affective computing have tried using machine learning techniques and EEG signals to automatically classify emotions. AlZoubi et al. [19] calculated the power spectral density (PSD) from EEG of three subjects and evaluated the performance of several classifiers (NaiveBayes, K-Nearest Neighbor and Support Vector Machines). Sohaib et al. [87] conducted a similar experiment but added regression tree and artificial neural networks in the evaluation. These studies show that there is some promise for constructing EEG-based automatic recognition systems.

Soleymani et al. [88] presented an emotion recognition system that combines EEG with information from the eyes (pupil diameter, gaze and blinking). Using these features, the authors were able to classify different levels of valence and arousal in video clips with SVM.

Chapter 3

An Overview of Craving Analysis

Researchers in the field of neuroscience and addiction have tried for a long time to accurately define the concept of craving. Although it is widely believed that craving has a close relationship to addiction, no single "ultimate" theory has been able to define the nature of craving, nor is there a universal agreement on the most appropriate ways to measure it [13, 14]. An increasing number of studies from a variety of biological and psychological perspectives have been trying to find answers to these open questions. In this section, we focus on past works on cue-elicited craving and attempts to measure craving by analyzing expressive behavior such as facial expressions.

3.1 Existing Works on Cue-Elicited Craving Detection

It is known that drug-related cues can act as a causal factor in drug use and relapse to drug use following treatment. Many works have investigated the relationship between drug-related cues and the EEG activities in human participants with history of chronic substance use, such as alcohol, cocaine, or tobacco [12, 89]. For example, Van de Laar et al. [90] observed that, when cocaine-abstinent patients and nondependent controls were exposed to neutral and cocaine-related pictures and were asked to rate the valence of the pictures, patients exhibited larger N300, late slow positive wave (LSPW) and sustained slow positive wave (SSPW) amplitudes following the cocaine-related pictures. When exposed to drug cues, cocaine-addicted individuals [91] and alcohol-dependent patients [92] displayed increased cortical activation. Nicotine users, when given cigarette-related cues, have displayed an increase in theta and beta spectral power [93]. Also, while performing a psychological test such as Stroop task or dot-probe task, current smokers showed more biased attention toward smoking-related cues than non-smokers or former smokers do [94, 95]. Shadel et al. [96] compared the effect of exposure to in vivo cues and video cues, and concluded that video is a viable cue delivery channel for manipulating craving in smoking research. Tong et al. [97] reported that smokers who are exposed to smoking videos produced greater craving reactions in self-reported measures, as well as elevated skin conductance and skin temperature.

3.2 Facial Expression and Craving

Since there is no single accepted definition of craving, researchers have used different measuring methods that are optimal for a particular research or clinical application [13, 14]. These measures can be largely divided into two groups: self-report measures and non-verbal measures. Self-report measures are popular in the community because they display a high degree of face validity and can be easily constructed and collected. However, it might be incorrect to regard self-reports as a direct readout of one's actual craving state. Moreover, there is the debate about whether an instrument for these measures should include the term "craving" or any other items that directly refers to some form of desire for a substance [14].

Non-verbal measures are based on behavioral or neurobiological responses to stimuli. This includes drug reinforcement proxies, psychophysiological and neurobiological responding, cognitive processing, and expressive behavior [13]. Although non-verbal measures can provide us with a more accurate involuntary response from the subjects, those responses can be interpreted in different ways depending on what one's theory of craving actually is.

Facial expression could be employed as a non-verbal measurement of craving. If craving happens to be affective in nature, there is a possibility that craving will trigger some kind of characteristic facial behavior like other emotions do. To verify if this is indeed the case, we can use the established feature extraction and classification methods from the field of facial expression recognition. However, only few of the past works looked into the relationship between drug-cue response and facial expressions. This may have been due to extremely time-consuming nature of manually labeling the facial expressions and the difficulty of extracting very subtle face deformations from subject's responses. A notable work is that of Sayette and Hufford [98], where subjects were exposed to smoking and control cues while their faces were videotaped. Their goal was to create a strong urge, associated first with positive affect (during initial exposure to smoking cues) and then negative affect (when told to extinguish the cigarette) to find out if affective valence of the urge will change as a function of drug availability. Subjects were divided into nicotine-deprived and nondeprived groups. For smoking cue, they were told to light a cigarette and hold it for a fixed period without placing it in their mouths. Then they held a roll of tape which is similar in size and weight to a cigarette. Subjects were more likely to express positive AUs (combinations of AU 1, AU 6 and AU 12) during initial exposure to smoking cues, although the correlations between self-reported urge and the appearance of negative AU combinations (AU 4, 1 + 4, 4 + 17, 12 + 15 and 14 + 15) were not significant. In a later study [16], the authors assumed that subjects exposed to smoking cue will display ambivalence about smoking. Ambivalence is characterized by competing inclinations to approach and avoid drug use, and it has been recognized as a central feature of drug addiction. The authors tried to detect specific AU combinations that are regarded as positive affect-related (AU 12, 6 + 12 along with AU 1 + 2, 25, 26) and negative affect-related (AU 9, 10, 14, 15, 20, 1 + 4). Ambivalence was defined as the simultaneous occurrence of both a positive AU and a negative AU. They compared the results with self-report measures about smoking ambivalence completed by the subjects. The results showed that there was a strong correlation between ambivalent detected by AUs and self-report measures.

Chapter 4

Experiment Design

4.1 The Goal of the Experiment

The purpose of this study is to understand the previous accomplishments in facial expression recognition and craving detection, and to explore the possible existence of facial behavior in drug cue-induced states which might be related to craving. In order to acquire a meaningful facial difference between the neutral and drug-induced state, it is necessary to first establish a neurobiological baseline so that we can distinguish the two states with some certainty. As explained in section 3.1, it is possible to expose substance users to drug-related and neutral cues and then observe the difference in EEG response. Assuming that cue-induced craving elicits a certain emotional state in drug users, we designed our experiment so that we can provide drug-related and non-drug stimuli to subjects and measure the activities in the brain and the face. We first try to validate our dataset by showing that the brain signals for drug and non-drug stimuli exhibit statistical difference. Then we examine the face and search for possible behaviors that only appear during drug-related cues.

Based on this assumption, we designed our experiment to follow two steps: the EEG analysis and the facial expression analysis. The following subjections discuss the participants and data collection environment, stimuli design, and the feature extraction and classification methods for both EEG and facial analysis steps.

4.2 Data collection

For this study, data from individuals with cocaine use disorders (CUD) and healthy control subject were used. CUD subjects are further divided into those who tested positive for recent cocaine use (CUD+) and those who tested negative (CUD-). The data was collected by the medical department of Brookhaven National Laboratory. In total, data from 19 CUD+ subjects, 41 CUD- subjects and 47 control subjects were used.

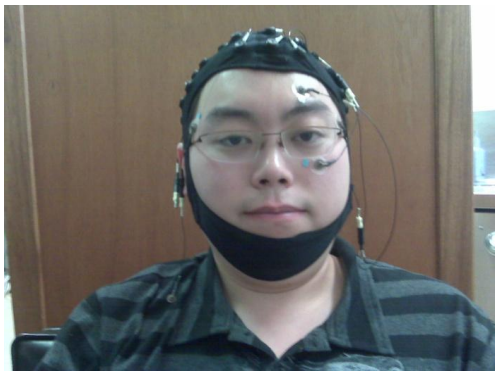

Figure 4.1: EEG cap and electrodes attached to the head.

Figure 4.2: Monitor and camcorder configurations

Data collection was conducted in the ERP room of Brookhaven National Laboratory medical department, which is a long rectangular room about 10 feet wide with fluorescent lighting. On one end of the room is a wall-mounted monitor connected to a working computer, and a single couch placed in front of the monitor. Subjects are required to sit on the couch and watch the monitor while the experiment is in progress.

For EEG capturing purposes, subjects wore a cap which is connected to electrodes that are placed on the head (Figure 4.1). EEG activity is recorded in real-time to a computer connected to the electrodes. To record facial expressions, a high resolution camcorder is mounted on top of the monitor, pointing to the subject's face from a near-frontal angle (Figure 4.2). The camcorder captured 1080p videos in 60 frames per second during the task. The pitch and heading of the camcorder, as well as the zoom level, was adjusted for each subject in order to compensate for inter-subject differences such as sitting height.

Figure 4.3: Example images from each category. From left to right: pleasant, unpleasant, drug-related and neutral.

4.3 Stimuli and Procedure

The stimuli for this study follows the design explained in [86]. Ninety pictures were selected from the International Affective Picture System (IAPS) [99] for the task. IAPS is a database of pictures that are intended to elicit a range of emotions. These pictures include a wide spectrum of contents, ranging from everyday objects such as furniture and silverware to highly disturbing scenes such as mutilation and robbery. Each picture has been given a valence and arousal score based on a normative rating procedure. Among the ninety pictures that were selected, 30 were labeled as pleasant (e.g. smiling faces, family photos, nudes), 30 as unpleasant (e.g. violent or depressing images), and 30 as neutral (e.g. expressionless faces, household objects). The categories were arranged such that pleasant and unpleasant pictures would be more arousing than neutral pictures, and that each category would differ in their respective valence scores.

On top of the three IAPS categories, 30 pictures of cocaine and individuals preparing or using cocaine were included as a fourth category. The images were acquired from freely available online sources and from a cocaine video used in a previous study [100]. Cocaine pictures were adjusted in overall size and human to non-human content ratio in order to match the pictures from the IAPS. Figure 4.3 shows an example image for each category.
The 120 images were arranged in random order across all four categories to create a sequence. Each sequence is divided into four blocks of 30 pictures. Each picture was displayed on the screen for 2000ms, followed by a 500ms inter-trial interval.

After EEG sensors were attached to the head, participants were given instructions about the task. They were told that they would be viewing a series of pictures depicting a wide range of contents that belong to pleasant, unpleasant, neutral, or drug categories. Participants were asked to concentrate on the screen and simply watch all of the pictures as they were displayed. After the sequence has ended, they rated each picture on valence (“rate how pleasant or unpleasant you felt about this picture”), arousal (“rate how strong of an emotional response you had to this picture”), like cocaine (“rate how much you like (or do not like) cocaine in response to this picture”), and want cocaine (“rate how much you want (or do not want) cocaine in response to this picture”). The rating process was done by a computerized version of the Self-Assessment Manikin (SAM) [101].

4.4 EEG Analysis

The EEG data is collected over 62 separate electrodes and digitized at a rate of 500 Hz. The dimension of the EEG data for each subject is 62 channels × 1101 timeframes × 120 trials. Because some trials were filtered during the preprocessing of the EEG signals, the actual number of trials for most subjects is lower than 120.

4.4.1 Baseline for EEG Analysis

Dunning et al. [86] has proposed a method to construct event-related potentials (ERPs) from EEG data by separately averaging trials based on picture type. They hypothesized that the two ERPs, the early posterior negativity (EPN) and the late positive potential (LPP), will be larger for both pleasant and unpleasant cues compared to neutral ones. Also, they expected that the LPPs elicited by drug cues will be larger than neutral cues in CUD subjects, but not in control subjects. After analyzing the results, they were able to verify that early and late LPP between the 400 – 2000ms timeframe showed statistical difference between drug and neutral cues in CUD subjects but not in control subjects. The experiment design of our study, as well as some of the EEG data, is directly derived from their work. Among the 19 CUD+, 41 CUD- and 47 control subjects, 5 CUD+, 17 CUD- and 19 controls have the same EEG data as used in [86].

4.4.2 ERP Features

We begin the construction of the ERP by first averaging trials based on the image categories (pleasant, unpleasant, neutral, and drug) on each subject. For each ERP averaged waveform, the average activity in the 200ms window prior to picture onset is used as the baseline. After adjusting the inter-subject variability, a 15Hz low pass filter was applied to reduce the noise.

Since LPP was able to capture the difference between drug and other categories, we concentrate on early and late LPP features. In [86], early LPP is defined as the average activity in an early window (400 – 1000ms) after picture onset, and late LPP as the average in the late window (1000 – 2000ms). We further divide the time window into eight segments that span across the entire LPP range (400 – 2000ms). The individual length of each window is 200ms. Then, instead of averaging across early and late windows, we calculate the average for eight separate windows and use these values as ERP features.

Scalp topographies from [86] show that drug-specific LPP modulation is maximal at fronto-central recording sites (Figure 4.4). Therefore, we use the signals from the Cz, FCz, FC1, FC2, and Fz electrodes which correspond to this region.

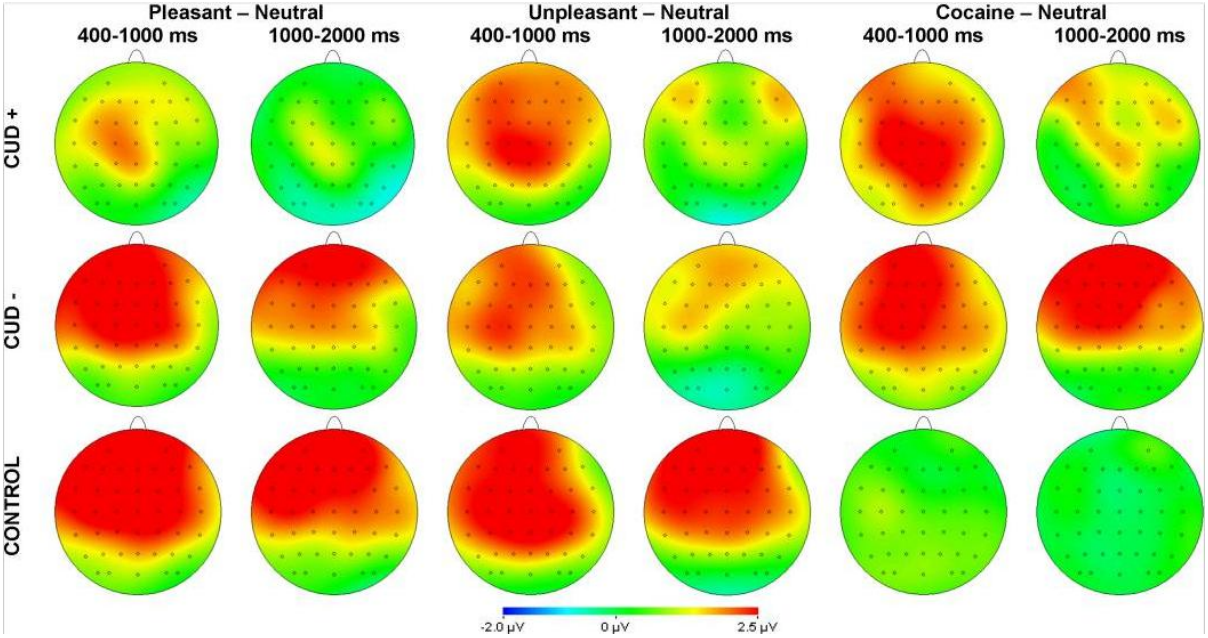

Figure 4.4: Scalp topography of pleasant minus neutral (left columns), unpleasant minus neutral (middle columns), and cocaine minus neutral (right columns) differences from [86].

Signal differences between the four groups were then calculated for each time window. This is done by conducting independent *t*-tests between the groups.

4.5 Facial Expression Analysis

Frontal face video was recorded during the picture viewing task for 10 subjects, two of which are control. Each video is approximately 6 to 10 minutes long, and contains audio signals of four different frequencies (one for each category) that are played at the onset of each picture cue. The video dimension is 1920 × 1080p and 60 frames per second. In the following subjections, we describe the means of tracking the face across the video, and what kind of features to extract from the tracked video.

4.5.1 Face Tracking

Due to the controlled environment of the experiment, the recorded facial videos display ideal qualities for tracking. The face is near-frontal for the most of the duration, with little or no

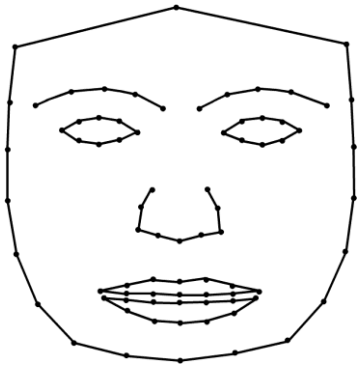

Figure 4.5: 77 landmark points that are used for

movement from the head or the body. The lighting condition is also consistent throughout the task.

To maximize the stability and the accuracy of tracking, we decided to use person-specific Active Appearance Model (AAM). The main reasoning behind using the person-specific AAM is that, due to the small number of subjects in the dataset, it would be extremely difficult to train a reliable universal model for all subjects. Also, it has been shown that person-specific AAM performs substantially better than generic AAMs [102].

We used the AAM-API implementation [103] for our purposes. For training, we selected 10 to 12 still images from each subject's video and manually labeled 77 landmark points on the brows, eyes, nose, mouth, and contour of the face (Figure 4.5). The training images were selected such that they contain all extents of the facial movement that are present in the video, in order to ensure that there are no unexpected jitters or registration failures in the final tracking result. After the labeling is complete, the images are then used to build an AAM model for each subject. Once the models are built, we ran the AAM tracking code on the entire video to locate the face and register the landmark points.

4.5.2 Feature Extraction

AAM provides the means to extract both geometric features (from the position of each landmark point) and appearance features (from the face texture of the warped shape). We first compensate for any head movement by aligning the tracked frames to the center-of-mass point of the contour points. Then, for each frame, we compute the difference of the landmark positions between the current frame and a "neutral" frame. The neutral frame is selected from the timeframe where the subject is watching a fixation screen before the experiment and no apparent expression appears on the face. The x- and y-displacement, moving velocity and direction of the points are calculated and concatenated as a feature vector with a dimension of $(2 + 1 + 1) \times 77 = 308$.

To extract the appearance features, the face in each frame is normalized using the trained model to match the mean shape, and the face texture is warped accordingly. We calculate two types of appearance features. First, we follow the method proposed in [28] and pass the face images through a bank of Gabor filters of eight orientation and 9 spatial frequencies (2 – 32 pixels per cycle at 1/2 octave steps). Then the output magnitudes are saved as features for classifiers. We also compute the Local Binary Pattern (LBP) histograms of the face image as described in [54]. The face is divided into 6 × 6 cells, and rotation-invariant uniform LBPs were calculated in each cell with parameters P = 8 and R = 1. This results in local histogram of size 10, and the final feature dimension is 10 × 36 cells = 360.

Chapter 5

Results

In this section, we present the results from ERP and facial expression analysis.

5.1 ERP Analysis

In [86], the authors defined the LPP as the average activity in an early (400-1000 ms) and late (1000-2000 ms) time windows after picture onset. We follow this method by first calculating the grand average of LPP for each picture type (Figure 5.1). One of the CUD+ subjects was showing some unusually high values for neutral trials due to a possible error in the artifact rejection algorithm, which created the noisy spike in the averaged signal for CUD+. However, this occurs before the early LPP window (400ms after stimuli onset) and therefore does not directly affect the analysis. Regardless, we include both results, with and without this subject. The averaged signal of each category for this subject is shown in Figure 5.2, and Figure 5.3 shows the grand average without this subject.

We first explore the early LPP window (400-1000 ms). In CUD+ subjects, cocaine-related compared to neutral LPPs was larger for both the entire group ($t(18) = -2.34$, $p < 0.05$) and without the noisy subject ($t(17) = -2.16$, $p < 0.05$). This was not the case for CUD- subjects ($t(40) = 1.06$, $p > 0.25$) and control subjects ($t(46) = 0.34$, $p > 0.5$). Also, cocaine LPPs in CUD+ did not differ from pleasant (all CUD+: $t(18) = 0.11$, $p > 0.9$; CUD+ without noisy subject: $t(17) = 0.1563$, $p > 0.8$) but was significantly smaller in CUD- ($t(40) = 2.23$, $p < 0.05$). Cocaine LPPs were not different from unpleasant in all CUD (all CUD+: $t(18) = -0.21$, $p > 0.8$; CUD+ without noisy subject: $t(17) = -0.23$, $p > 0.8$; CUD-: $t(40) = 0.75$, $p > 0.4$). In contrast, cocaine LPPs in controls were notably smaller than pleasant ($t(46) = 2.48$, $p < 0.05$) but not as much than unpleasant ($t(46) = 1.43$, $p > 0.15$) LPPs. Also, in control subjects, pleasant and unpleasant LPPs were not greater than neutral LPPs in a meaningful way ($t(46) = -1.52$, $p > 0.1$ and $t(46) = -0.5$, $p > 0.6$, respectively), and did not differ from each other ($t(46) = 0.89$, $p > 0.3$). In CUD+, pleasant and unpleasant were both larger than neutral for all CUD+ ($t(18) = -2.82$, $p < 0.01$ and $t(18) = -2.39$, $p < 0.05$, respectively) and CUD+ without the noisy subject ($t(17) = -2.65$, $p < 0.015$ and $t(17) = -2.18$, $p < 0.05$, respectively). Pleasant and unpleasant was not different from each other (all CUD+: $t(18) = 0.37$, $p > 0.7$; CUD+ without noisy subject: $t(17) = 0.44$, $p > 0.6$). As for CUD-, pleasant and unpleasant LPPs did not differ from neutral ($t(40) = -1.3$, $p > 0.1$ and $t(40) = 0.23$, $p > 0.8$, respectively) and could not be differentiated from each other ($t(40) = 1.37$, $p > 0.15$).

In summary, the magnitude of LPPs elicited by cocaine, pleasant, and unpleasant pictures was greater than that elicited by neutral pictures in CUD+, but not in CUD-. Cocaine was not distinguishable from either pleasant or unpleasant in CUD+, and only distinguishable from pleasant in CUD-. In control subjects, LPPs elicited by pleasant and unpleasant pictures were not significantly different from LPPs elicited by cocaine and neutral pictures, except that pleasant was larger than cocaine. This result loosely follows the findings in [86] for CUD+, but not for CUD- and controls. See Figure 5.4 for the grand average of LPP in [86].

In the late window (1000-2000 ms), cocaine LPPs was not significantly different from neutral LPPs for CUD+ (all CUD+: $t(18) = -1.86$, $p > 0.05$; CUD+ without noisy subject: $t(17) = -1.69$, $p > 0.1$), CUD- ($t(40) = 0.34$,
$p > 0.7$) and control ($t(46) = 0.83$, $p > 0.4$). Cocaine LPPs in CUD+ did not differ from pleasant (all CUD+: $t(18) = 0.63$, $p > 0.5$; CUD+ without noisy subject: $t(17) = 0.67$, $p > 0.5$) but was again smaller in CUD- ($t(40) = 2.21$, $p < 0.05$). Cocaine LPPs were not distinguishable from unpleasant in all CUD (all CUD+: $t(18) = 0.09$, $p > 0.9$; CUD+ without noisy subject: $t(17) = 0.1$, $p > 0.9$; CUD-: $t(40) = 0.32$, $p > 0.7$).

In control subjects, cocaine LPPs was significantly smaller than pleasant ($t(46) = 3.13$, $p < 0.005$) and unpleasant ($t(46) = 2.18$, $p < 0.05$) LPPs. Pleasant LPPs in control was greater than neutral ($t(46) = -2.7$, $p < 0.01$), but unpleasant LPPs was not ($t(46) = -1.59$, $p > 0.1$). Pleasant and unpleasant did not differ from each other ($t(46) = 0.88$, $p > 0.3$). For all CUD, pleasant was larger than neutral (all CUD+: $t(18) = -2.86$, $p < 0.01$; CUD+ without noisy subject: $t(17) = -2.69$, $p < 0.015$; CUD-: $t(40) = -2.53$, $p < 0.05$) but unpleasant was not (all CUD+: $t(18) = -1.98$, $p > 0.05$; CUD+ without noisy subject: $t(17) = -1.81$, $p > 0.05$; CUD-: $t(40) = -0.66$, $p > 0.5$). Pleasant and unpleasant was not different from each other in CUD (all CUD+: $t(18) = 0.53$, $p > 0.5$; CUD+ without noisy subject: $t(17) = 0.56$, $p > 0.5$; CUD-: $t(40) = 1.84$, $p > 0.05$).

In summary, pleasant LPPs was greater than neutral LPPs in all subjects, but unpleasant and cocaine LPPs could not be differentiated from neutral. Cocaine was clearly different from both pleasant and unpleasant in control, only different from pleasant in CUD-, and not different from any of them in CUD+, and only distinguishable from pleasant in CUD-. Again, this result follows some trends shown in [86], but also deviates in other places. Considering that 66 out of 107 subjects are not used in [86] and are newly introduced in this work, we could conclude that, after adding a significant number of new data, the results still display some trends between LPPs of different categories that were observed in the previous study, although to a much weaker degree. Table 5.1 shows the results of *t*-tests for each of the smaller sub-windows (200ms each).

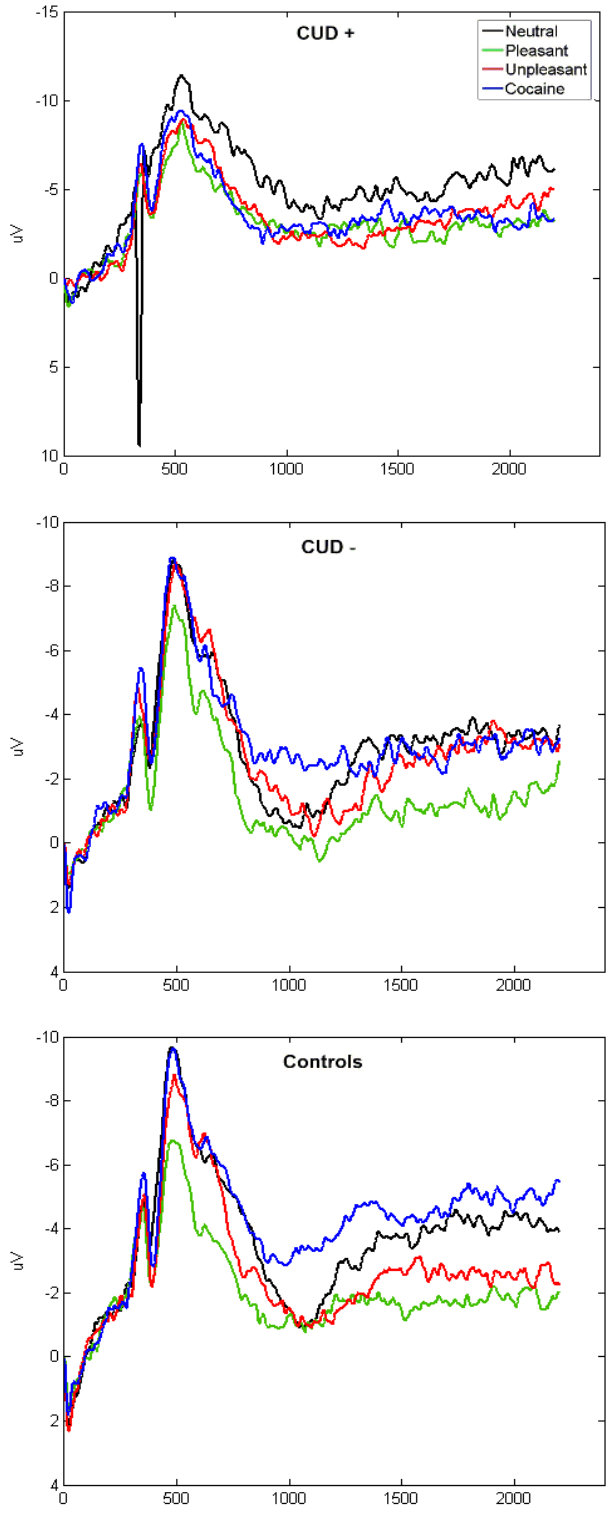

Figure 5.1: Grand average late positive potentials of all subjects (at the average of sites Cz, FCz, FC1, FC2, and Fz) elicited by neutral, pleasant, unpleasant, and cocaine-related pictures. Stimulus onset occurs at 0 ms.

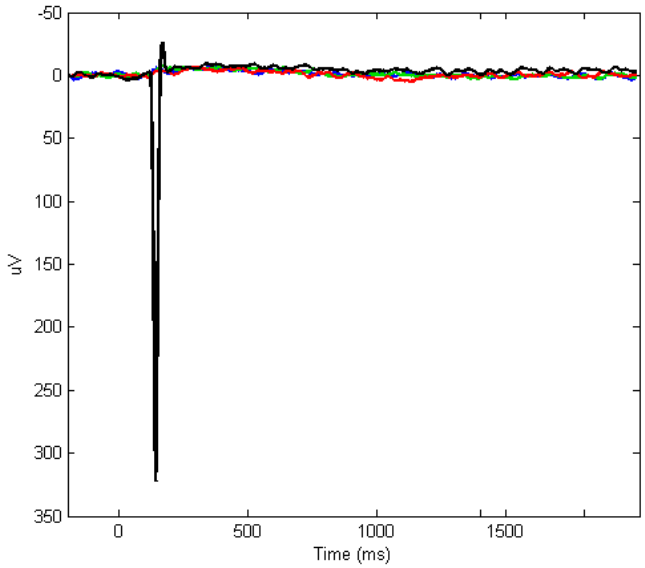

Figure 5.2: CUD+ subject with noisy data in the early window

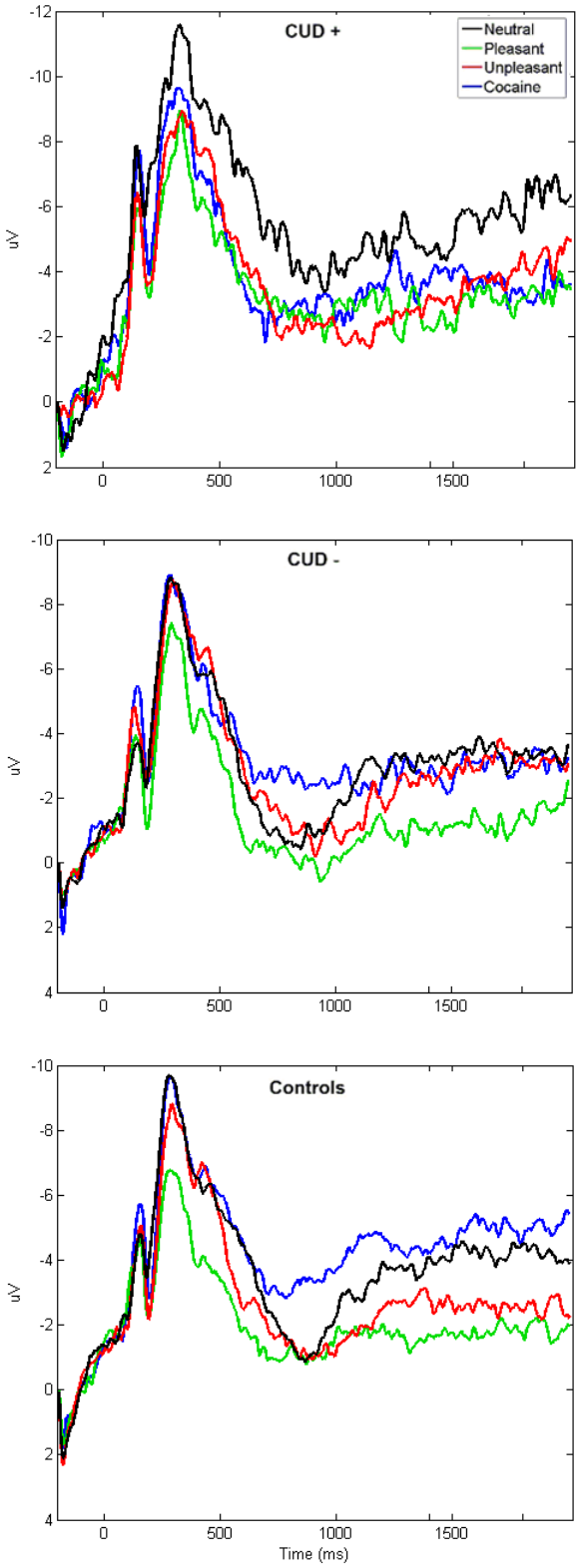

Figure 5.3: Grand averaged late positive potentials, after removing the noisy subject.

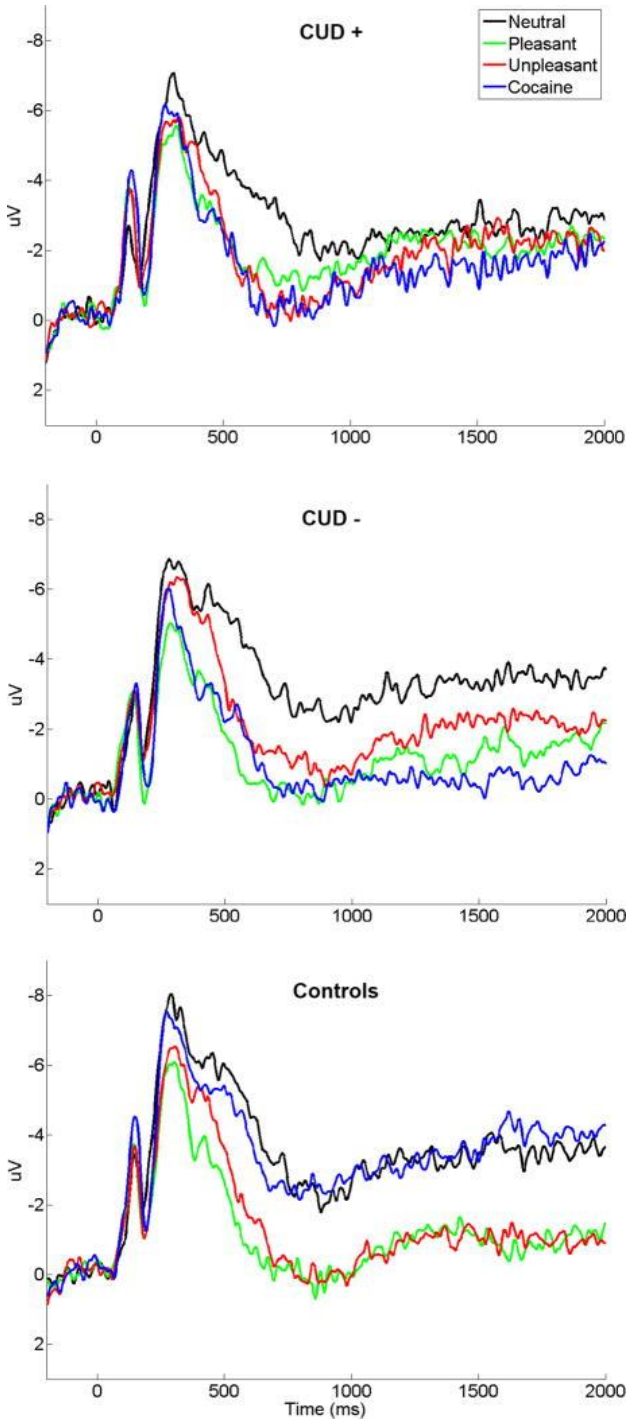

Figure 5.4: Grand averaged late positive potentials from [86]

			Time Window (ms)							
			400-600	600-800	800-1000	1000-1200	1200-1400	1400-1600	1600-1800	1800-2000
Control	Neutral vs. Pleasant	t-statistics	-2.3252	-1.4846	-0.2984	-1.5568	-2.329	-2.7078	-2.8179	-3.0099
		p	0.0223	0.1411	0.7661	0.123	0.022	0.0081	0.0059	0.0034
	Neutral vs. Unpleasant	t-statistics	-0.5513	-0.5579	-0.2638	-1.3501	-0.9302	-1.6779	-1.8211	-1.7734
		p	0.5828	0.5783	0.7925	0.1803	0.3547	0.0968	0.0718	0.0795
	Neutral vs. Drug	t-statistics	0.2619	0.6127	2.1179	1.4581	0.5886	0.4521	0.7023	0.8032
		p	0.794	0.5416	0.0369	0.1482	0.5576	0.6523	0.4843	0.4239
CUD+	Neutral vs. Pleasant	t-statistics	-3.3023	-2.6051	-1.5187	-1.3779	-2.1539	-2.653	-3.2276	-3.318
		p	0.0022	0.0133	0.1376	0.1767	0.038	0.0118	0.0027	0.0021
	Neutral vs. Unpleasant	t-statistics	-2.1273	-2.8307	-1.5152	-2.0565	-1.7057	-1.5737	-1.5543	-1.8518
		p	0.0403	0.0076	0.1385	0.047	0.0967	0.1243	0.1289	0.0723
	Neutral vs. Drug	t-statistics	-2.4207	-2.8096	-1.0065	-1.2124	-1.1569	-1.2016	-2.1641	-2.5564
		p	0.0207	0.008	0.3209	0.2333	0.2549	0.2374	0.0372	0.0149
CUD-	Neutral vs. Pleasant	t-statistics	-1.7203	-0.8113	-1.0522	-2.2295	-2.444	-2.7163	-2.4643	-1.9069
		p	0.0892	0.4196	0.2959	0.0286	0.0167	0.0081	0.0159	0.0601
	Neutral vs. Unpleasant	t-statistics	0.1343	0.5452	-0.0504	-1.3365	-0.7321	-0.5721	-0.0916	-0.3143
		p	0.8935	0.5871	0.96	0.1852	0.4662	0.5689	0.9273	0.7541
	Neutral vs. Drug	t-statistics	-0.0769	1.4962	1.6184	-0.1674	-0.2283	-0.6598	-0.3922	-0.1945
		p	0.9389	0.1385	0.1095	0.8675	0.82	0.5113	0.696	0.8462
CUD+ without noisy subject	Neutral vs. Pleasant	t-statistics	-3.1749	-2.4658	-1.3487	-1.2414	-1.9849	-2.4773	-3.0857	-3.1607
		p	0.0032	0.0189	0.1864	0.2229	0.0553	0.0184	0.004	0.0033
	Neutral vs. Unpleasant	t-statistics	-1.9733	-2.59	-1.329	-1.8204	-1.5522	-1.4529	-1.4499	-1.7055
		p	0.0566	0.014	0.1927	0.0775	0.1299	0.1554	0.1562	0.0972
	Neutral vs. Drug	t-statistics	-2.2589	-2.6234	-0.8583	-1.0781	-0.9949	-1.0463	-2.0128	-2.3839
		p	0.0304	0.0129	0.3967	0.2886	0.3268	0.3028	0.0521	0.0229

Table 5.1: t-test results between neutral and other categories for different time windows.

5.2 Facial Expression Analysis

We extracted geometric and appearance facial features from video frames that correspond to each of the picture type (pleasant, unpleasant, neutral, and cocaine). Since each trial lasts for about 2 seconds, we acquired roughly 120 video frames for each image stimulus. We cut off the first 24 frames (400ms) and used the remaining 96 frames per trial for classification. Also, we ignored the video for all subjects that were rejected due to noisy EEG signals in section 5.1.

We first tried an automatic AU detection scheme based on FACS. If there were specific facial behaviors appearing when certain types of images are shown, a robust AU detection system would be able to recognize it. We followed the framework design of [28] to implement the system. For each training face image that contains a certain AU, the Gabor filter output is calculated and the magnitudes were passed to the classifiers. For each action, a linear SVM was trained in a "one versus all" manner. Two databases were used to train the system: the Cohn-Kanade dataset [38] and the MMI web-based dataset [104]. When tested on these databases by leave-one-subject-out cross validation, the system was able to reasonably detect several AUs that are considered important in facial expression analysis. Table 5.2 shows the classification performance for each of the trained AUs. However, when tested on the face video data, the system failed to detect any action units in most of the frames except only a few. By closely examining the face videos, it is apparent that the subjects are not moving their face or body during most of the task. Since the AU detection system was trained on databases with strong posed expressions, it is reasonable to think that it would not work well with spontaneous and very subtle expressions.

AU	Name	N	Hit	FA
1	Inner brow raiser	281	82	11
2	Outer brow raiser	205	80	15
4	Brow lowerer	262	69	8
5	Upper lid raiser	179	76	9
6	Cheek raiser	160	74	5
7	Lid tightener	276	84	10
10	Upper lip raiser	26	15	3
12	Lip corner puller	113	84	12
15	Lip corner depressor	68	75	9
16	Lower lip depressor	23	13	4
20	Lip stretcher	50	40	8

Table 5.2: Performance of the automatic AU detection system on Cohn-Kanade and MMI databases. N: Total number of positive examples. Hit: Hit rate. FA: False alarm rate.

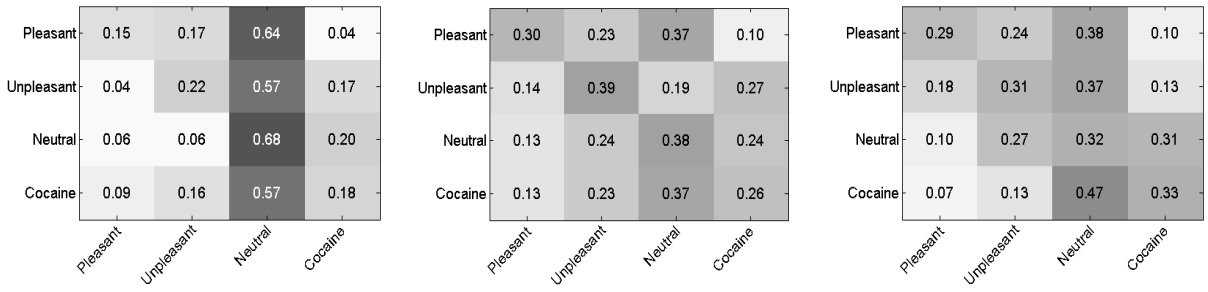

Figure 5.5: Confusion matrix for SVM classification with grid-based LBP features. From left to right: CUD+, CUD-, and controls.

Classification with LBP histograms yielded similar results. When the grids of the whole face region were used for feature extraction, classification rate between pleasant, unpleasant, neutral and drug video frames did not produce meaningful outcomes (Figure 5.5). Most of the frames were labeled as neutral in CUD+ subjects, and control subjects showed no significant improvement above chance. It is notable that unpleasant category is recognized slightly better in CUD- subjects than others.

Although grid-based LBP histograms are known to perform extremely well for many kinds of posed and spontaneous facial expressions, it usually cannot deal with the case where facial activity is extremely weak across all regions. It would be logical to say that the poor classifier performance is resulting from the difficulty of face videos themselves and not necessarily from the choice of facial features or classification methods.

Lastly, we tried a sequence-based classification approach with geometric features rather than trying to classify each frames. In [34], the authors calculate the movement direction and magnitude of facial points and store them as feature vectors. Based on the definition of action units (for example, outer corner of the brow moving upward corresponds to AU2), they measure the movement of certain points and detects an action unit when it exceeds the pre-defined threshold. Although this method is less flexible and prone to issues of inter-subject variance, it is simple to compute and could prove useful when only miniscule facial movements are present in the dataset. We implement a similar method by first processing through the video frames of the entire single trial, cumulating the magnitude of facial point displacement over time, and then deciding whether an action has occurred within the duration of the current trial. Because the subjects seldom move in the video and the person-specific AAM provides us with a very reliable tracking, we can use a very low threshold value to detect small changes in facial landmark points. For each direction (upwards, downwards, left and right), we define the "displacement strength" for landmark points as follows:

Displacement strength = (number of pixels moved / length of the face in pixels × 2) × (2 / time elapsed in seconds since motion onset).

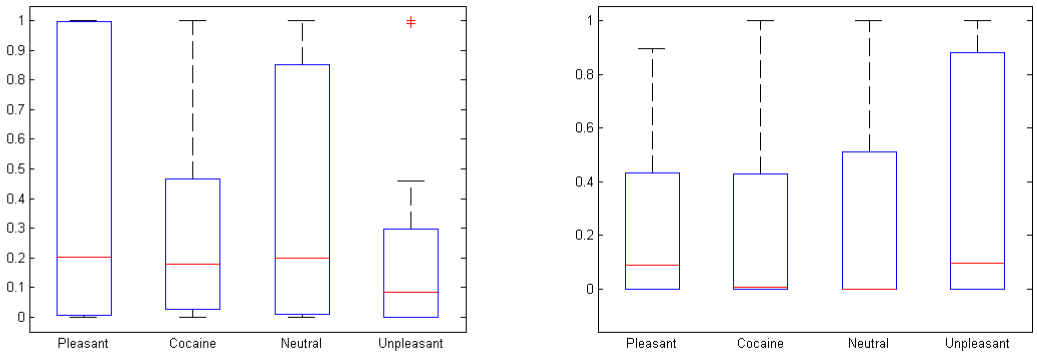

Figure 5.6: Landmark points displacement strength in upwards brow motion (left) and downwards brow motion (right)

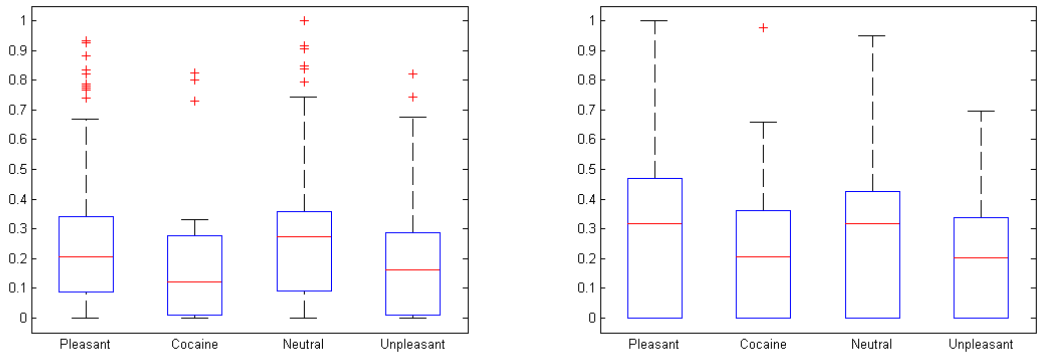

Figure 5.7: Landmark points displacement strength in downwards lower-nose motion (left) and upwards mouth corner motion (right)

This measure becomes large when the magnitude of motion is great and the movement is quicker. We divide the face into several regions of interest (brows, nose, mouth center, mouth corner). Within each region, we calculate the displacement strength of all points over the duration of a trial and then average the strengths to get the final value for that region.

Among the 10 subjects, three CUD- subjects displayed noticeable motions that are somewhat consistent across picture types. Figures 5.6 and 5.7 show some of the motions that are detected from these subjects. The plot suggests that the brow, nose and mouth corner regions show somewhat distinguishable behavior between the picture types but not by a huge degree. If we use these values as features for training a linear SVM classifier, we can see that the result becomes slightly more reasonable but still doesn't clearly separate between the categories (Figure 5.8). It is worth noting that, in all actions shown in Figures 5.6 and 5.7 except for the downwards brow motion, cocaine cues elicit facial response that is similar to those caused by unpleasant cues.

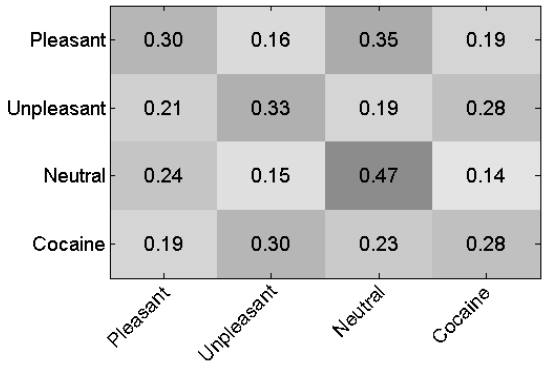

Figure 5.8: Confusion matrix for SVM classification with displacement strength features

Chapter 6

Conclusions and Future Works

In this thesis, we explore the problem of measuring drug cue-induced response by analyzing the neurobiological response from the brain and the expressive response from the face. We hypothesize that, like other human emotions, craving is an affective state and thus could trigger certain non-verbal expressions. Based on the previous findings in the neuroscience community, we design a two-step experiment for validating the relationship between drug cue-induced emotional states and facial expressions: First, we analyze the EEG to prove that cocaine users, when exposed to drug-related stimuli, exhibit different pattern of event-related potentials compared to neutral stimuli. Second, we study the facial behavior while the neural activity is taking place and try to discover what kind of expressions are related to drug-related response. We use several feature extraction and classification methods to distinguish between the facial response of drug stimuli and other cues. We show that the expression recognition techniques that we use do not yield significant results due to the fact that the collected face data mostly lacks of explicit facial expressions. However, some drug-cue induced activities in certain subjects suggest the possibility that there might still be some kind of craving expressions that are subject-dependent in nature.

6.1 Conclusion

The experiment on drug cue-related behavior has produced a mixed result for its two stages. Following a previous study which has shown that cocaine cues elicit neural activities in cocaine users that are not seen in healthy individuals, we attempted to obtain similar results using a part of their dataset with new subjects added. When compared to the baseline research for EEG [86] we found that, even after the addition of new subjects, ERP readings still exhibited some of the traits that were shown in earlier works. In contrast, facial expression analysis task turned out to be much more difficult than initially expected. Visual examination of the recorded face videos reveals that most subjects keep their face and body fixed throughout the experiment and do not display explicit emotional behavior very often. Hence, many of the facial recognition techniques that are popular today cannot work well in this dataset.

The reason why these videos did not contain enough facial information, despite the actual neurological activity that was taking place in the background, could be explained from several perspectives. One reason is that we do not know what the face of craving looks like and therefore

we're searching in the dark. There has not been any past work in neuroscience or computer vision that has discovered a clear and universal definition for facial expression of craving. Basic emotions such as happiness or surprise usually have a set of well-defined action units that are strongly correlated to the occurrence of those emotions. For example, a smile is almost always characterized by the combination of AU6 (Cheek raiser) and AU12 (Lip corner puller), although some attributes such as smoothness and temporal development might vary according to the context. Many modern facial recognition systems take advantage of these characteristics to improve their performance. Those who attempt to detect more sophisticated and subtle emotions such fatigue or anxiety usually include other modalities like gaze and body gestures. Therefore it would be a quite challenging problem to recognize craving just by facial behaviors.

Another reason is that the experiment is designed more with the ERP analysis in mind. In the affective computing community, it is customary to expose the individuals to visual cues for a longer time [88]. 2 seconds of displayed image followed by a 500ms interval is sufficient for many ERP related studies, but the face usually requires some time before it can actually deliver emotion. In some cases it can take several seconds before a facial expression reaches its apex [63]. Therefore, a sequence design with longer trial duration would be beneficial to face research.

The last reason is that, because the subjects were doing a passive viewing task and were asked not to do anything else, they might have involuntarily suppressed some or all of the expressions that might have occurred. During the instruction phase, subjects are told to focus on the screen throughout the experiment, and no specific instruction is given on whether or not to express themselves. Also, it is possible for the cocaine users to feel uncomfortable to express their cravings in response to cocaine stimuli, especially when one or two people are in the room to supervise the task. These factors, plus the fact that each trial only lasts for two seconds, could create stress and therefore cloud the true emotion underneath. These three observations suggest that, in order to gain a better understanding at facial expression of craving, several important design choices will have to be considered in the future.

6.2 Future Work

As stated in section 6.1, the most important step for future research is to design a task that is suited for both ERP and facial expression studies, preferably with longer exposure time. An improved task design will hopefully result in stronger emotions beginning to appear on the face. One possible task is to expose the individual to drug stimuli and instruct him or her to actively suppress craving. Since suppressing an emotion requires conscious effort that is sometimes displayed on the face, it would be interesting to investigate what kind of facial movements occur while craving is being suppressed. Once we are able to detect more consistent facial expressions across subjects, we will be able to study the actual correlation between ERP readings and observed facial behavior.

References

[1] C. Darwin. Expression of the Emotions in Man and Animals. Oxford Univ. Press, Inc., 2002.

[2] C. Darwin. The Expression of the Emotions in Man and Animals. John Murray, 1872.

[3] R. A. Calvo and S. D'Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18-37, 2010.

[4] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 1971.

[5] C. Izard. *The Face of Emotion.* Appleton-Century-Crofts, 1971.

[6] J. Haidt and D. Keltner. Culture and facial expression: open-ended methods find More Expressions and a Gradient of Recognition. *Cognition and Emotion*, 1999.

[7] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39-58, 2009.

[8] Q. Ji, P. Lan, and C. Looney. A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on Systems, Man, and Cybernetics Part A*, vol.36, no. 5, pp. 862-875, 2006.

[9] A. Kapoor, W. Burleson, and R. W. Picard. Automatic prediction of frustration. *International Journal of Human-Computer Studies*, vol. 65, no. 8, pp. 724-736, 2007.

[10] A. B. Ashraf, A. Bilal, S. Lucey, J. F. Cohn, T. Chen, Z. Ambadar, K. M. Prkachin, and P. E. Solomon. The Painful Face – Pain expression recognition using active appearance models. *Image and Vision Computing*, vol. 27, no. 12, pp. 1788-1796, 2009.

[11] D. C. Drummond. Theories of drug craving, ancient and modern. *Addiction*, vol. 96, no. 1, pp. 33-46, 2001.

[12] N. A. Ceballos, L. O. Bauer, and R. J. Houston. Recent EEG and ERP findings in substance abusers. *Clinical EEG and Neuroscience*, vol. 40, no. 2, pp. 122-128, 2009.

[13] M. A. Sayette, S. Shiffman, S. T. Tiffany, R. S. Niaura, C. S. Martin, and W. G. Schadel. The measurement of drug craving. *Addiction*, vol. 95, no. 8s2, pp. 189-210, 2000.

[14] S. T. Tiffay and J. M. Wray. The clinical significance of drug craving. *Annals of the New York Academy of Sciences*, vol. 1248, no. 1, pp. 1-17, 2012.

[15] M. A. Sayette and M. R. Hufford. Urge and affect: A facial coding analysis of smokers. *Experimental and Clinical Psychopharmacology*, vol. 3, no. 4, pp. 417-423, 1995.

[16] K. M. Griffin and M. A. Sayette. Facial reactions to smoking cues relate to ambivalence about smoking. *Psychology of Addictive Behaviors*, vol. 22, no. 4, pp. 551-556, 2008.

[17] M. H. Winkler, P. Weyers, R. F. Mucha, B. Stippekohl, R. Stark, and P. Pauli. Conditioned cues for smoking elicit preparatory responses in healthy smokers. *Psychopharmacology*, vol. 213, no. 4, pp. 781-789, 2011.

[18] J. K. Olofsson, S. Nordin, H. Sequeira, and J. Polich. Affective picture processing: An integrative review of ERP findings. *Biological Psychology*, vol. 77, no. 3, pp. 247-265, 2008.

[19] O. AlZoubi, R. A. Calvo, and R. H. Stevens. Classification of EEG for affect recognition: An adaptive approach. *AI 2009: Advances in Artificial Intelligence*, pp. 52-61, Springer Berlin Heidelberg, 2008.

[20] R. W. Picard. *Affective Computing*. The MIT Press, 1997.

[21] P. Ekman and W. V. Friesen. Facial Action Coding System: A technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto, 1978.

[22] M. Pantic and M. S. Bartlett. Machine Analysis of Facial Expressions. *Face Recognition*, K. Delac and M. Grgic, eds., pp. 377-416, I-Tech Education and Publishing, 2007.

[23] L. I. Reed, M. A. Sayette, and J. F. Cohn. Impact of depression on response to comedy: A dynamic facial coding analysis. *Journal of Abnormal Psychology*, vol. 116, no. 4, pp. 804-809, 2007.

[24] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma. Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of Neuroscience Methods*, vol. 200, no. 2, pp. 237-256, 2011.

[25] J. F. Cohn and K. L. Schmidt. The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, no. 2, pp. 121-132, 2004.

[26] W. E. Rinn. The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, vol. 95, no. 1, pp. 52-77, 1984.

[27] M. S. Bartlett and J. Whitehill. Automated facial expression measurement: recent applications to basic research in human behavior, learning, and education. *Handbook of Face Perception*, Oxford University Press, USA, 2010.

[28] G. C. Littlewort, M. S. Bartlett, and K. Lee. Automatic coding of facial expressions displayed during posed and genuine pain. *Image and Vision Computing*, 2009.

[29] D. F. Dinges, R. L. Rider, J. Dorrian, E. L. McGlinchey, N. L. Rogers, Z. Cizman, S. K. Goldenstein, C. Vogler, S. Venkataraman, and D. N. Metaxas. Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, Space, and Environmental Medicine*, vol. 76, no. Supplement 1, pp. B172-B182, 2005.

[30] J. Hoey. Hierarchical unsupervised learning of facial expression categories. *IEEE Workshop on Detection and Recognition of Events in Video*, pp. 99-106, 2001.

[31] F. Bettinger, T. F. Cootes, and C. J. Taylor. Modelling facial behaviors. *British Machine Vision Conference*, pp. 1-10, 2002.

[32] F. De la Torre, J. Campoy, Z. Ambadar, and J. F. Cohn. Temporal segmentation of facial behavior. *IEEE International Conference on Computer Vision*, pp. 1-8, 2007.

[33] F. Zhou, F. De la Torre, and J. F. Cohn. Unsupervised discovery of facial events. *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2574-2581, 2010.

[34] M. Pantic and L. J. M. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2004.

[35] M. F. Valstar and M. Pantic. Fully automatic recognition of the temporal phases of facial actions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 42, no. 1, pp. 28-43, 2012.

[36] Y. Chang, C. Hu, R. Feris, and M. Turk. Manifold based analysis of facial expression. *Image and Vision Computing*, vol. 24, no. 6, pp. 605-614, 2006.

[37] I. Kotsia and I. Pitas. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 172-187, 2007.

[38] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. *Proc. IEEE International Conference on Face and Gesture Recognition*, pp. 46-53, 2000.

[39] Y. Yacoob and L. S. Davis. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 636-642, 1996.

[40] I. A. Essa and A. P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 757-763, 1997.

[41] J. Hoey and J. J. Little. Value directed learning of gestures and facial displays. *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[42] M. Yeasin, B. Bullot, and R. Sharma. From facial expression to level of interest: a spatio-temporal approach. *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[43] S. Lucey, A. B. Ashraf, J. F. Cohn. Investigating spontaneous facial actions recognition through AAM representations of the face. *Face Recognition*, pp. 275-286, 2007.

[44] J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, vol. 58, no. 6, pp. 1233-1258, 1987.

[45] J. G. Daugman. Complete discrete 2-D gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 7, pp. 1169-1179, 1988.

[46] M. Lyons and S. Akamatsu. Coding facial expressions with gabor wavelets. *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200-205, 1998.

[47] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perception. *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 454-459, 1998.

[48] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, 1999.

[49] Y. Tian, T. Kanade, and J. F. Cohn. Eye-state action unit detection by gabor wavelets. *Proc. International Conference on Multi-modal Interfaces*, pp. 143-150, 2000.

[50] Y. Tian, T. Kanade, and J. F. Cohn. Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 229-234, 2002.

[51] G. Guo and C. R. Dyer. Learning from examples in the small sample case: face expression recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 3, pp. 477-488, 2005.

[52] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.

[53] C. Shan, S. Gong, and P. W. McOwan. Facial expression recognition based on local binary patterns: a comprehensive study. *Image and Vision Computing*, vol. 27, no. 6, pp. 803-816, 2009.

[54] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen. Local binary patterns and its application to facial image analysis: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 41, no. 6, pp. 765-781, 2011.

[55] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. *Proc. European Conference on Computer Vision*, pp. 469-481, Springer Berlin Heidelberg, 2004.

[56] Y. Chang, M. Vieira, M. Turk, and L. Velho. Automatic 3d facial expression analysis in videos. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 293-307, Springer Berlin Heidelberg, 2005.

[57] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato. A 3d facial expression database for facial behavior research. *Proc. International Conference on Automatic Face and Gesture Recognition*, pp. 211-216, 2006.

[58] J. F. Cohn, L. I. Reed, Z. Ambadar, J. Xiao, and T. Moriyama. Automatic analysis and recognition of brow actions and head motion in spontaneous facial behavior. *IEEE International Conference on Systems, Man and Cybernetics*, vol. 1, pp. 610-616, 2004.

[59] F. Dornaika and F. Davoine. Simultaneous facial action tracking and expression recognition in the presence of head motion. *International Journal of Computer Vision*, vol. 76, no. 3, pp. 257-281, 2008.

[60] J. Ahlberg. An active model for facial feature tracking. *EURASIP Journal on Applied Signal Processing*, no. 1, pp. 566-571, 2002.

[61] X. Li, T. Jia, and H. Zhang. Expression-insensitive 3d face recognition using sparse representation. *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2575-2582, 2009.

[62] Y. Tian. Evaluation of face resolution for expression analysis. *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 82-88, 2004.

[63] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Recognizing facial expression: machine learning and application to spontaneous behavior. *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 568-573, 2005.

[64] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding*, vol. 91, no. 1, pp. 160-187, 2003.

[65] M. Pantic and I. Patras. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 2, pp. 433-449, 2006.

[66] R. E. Kaliouby and P. Robinson. Real-time inference of complex mental states from facial expressions and head gestures. *Real-time Vision for Human-Computer Interaction*, pp. 181-200, Springer US, 2005.

[67] Y. Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 699-714, 2005.

[68] A. Kleinsmith and N. Bianchi-Berthouze. Affective Body Expression Perception and Recognition: A Survey. *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15-33, 2013.

[69] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175-1191, 2001.

[70] S. Mota and R. W. Picard. Automated posture analysis for detecting learner's interest level. *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, vol. 5, 2003.

[71] S. D'Mello and A. Graesser. Automatic detection of learner's affect from gross body language. *Applied Artificial Intelligence*, vol. 23, no. 2, pp. 123-150, 2009.

[72] M. F. Valstar, H. Gunes, and M. Pantic. How to Distinguish Posed From Spontaneous Smiles Using Geometric Features. *Proceedings of the 9th International Conference on Multimodal Interfaces*, pp. 38-45, ACM, 2007.

[73] H. Gunes and M. Piccardi. Automatic Temporal Segment Detection and Affect Recognition from Face and Body Display. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 39, no. 1, pp. 64-84, 2009.

[74] J. A. Russell, J. A. Bachorowski, and J. M. Fernández-Dols. Facial and vocal expressions of emotion. *Annual Review of Psychology*, vol. 54, no. 1, pp. 329-349, 2003.

[75] T. Iohnstone and K. R. Scherer. Vocal communication of emotion. *Handbook of Emotion*, pp. 220-235, Guilford Press, 2000.

[76] J. Liscombe, J. B. Hirschberg, and J. J. Venditti. Detecting certainness in spoken tutorial dialogues. *Proc. Ninth European Conference on Speech Communication and Technology (INTERSPEECH)*, 2005.

[77] O. W. Kwon, K. Chan, J. Hao, and T. W. Lee. Emotion recognition by speech signals. *Proc. Eighth European Conference on Speech Communication and Technology (EUROSPEECH)*, 2003.

[78] A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. How to find trouble in communication. *Speech Communication*, vol. 40, no. 1, pp. 117-143, 2003.

[79] S. Steidl, M. Levit, A. Barliner, E. Nöth, and H. Niemann. Of all things the measure is man: Automatic classification of emotions and inter-labeler consistency. *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 1, pp. 317-320, 2005.

[80] J. A. Russell, J. A. Bachorowski, and J. M. Fernández-Dols. Facial and vocal expressions of emotion. *Annual Review of Psychology*, vol. 54, no. 1, pp. 329-349, 2003.

[81] S. T. Shivappa, M. M. Trivedi, and B. D. Rao. Audiovisual information fusion in human-computer interfaces and intelligent environments: A survey. *Proceedings of the IEEE* vol. 98, no. 10, pp. 1692-1715, 2010.

[82] Z. Zeng, J. Tu, P. Pianfetti, M. Liu, T. Zhang, Z. Zhang, T.S. Huang, and S. Levinson. Audio-Visual Affect Recognition through Multi-Stream Fused HMM for HCI. *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 967-972, 2005.

[83] N. Sebe, I. Cohen, T. Gevers, and T.S. Huang. Emotion Recognition Based on Joint Visual and Audio Cues. *Proc. 18th International Conference on Pattern Recognition*, vol. 1, pp. 1136-1139, 2006.

[84] S. Petridis and M. Pantic. Audiovisual discrimination between laughter and speech. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5117-5120, 2008.

[85] T. Dalgleish, B. D. Dunn, and D. Mobbs. Affective neuroscience: Past, present, and future. *Emotion Review*, vol. 1, no. 4, pp. 355-368, 2009.

[86] J. P. Dunning, M. A. Parvaz, G. Hajcak, T. Maloney, N. Alia-Klein, P. A. Woicik, F. Telang, G. J. Wang, N. D. Volkow, and R. Z. Goldstein. Motivated attention to cocaine and emotional cues in abstinent and current cocaine users-an ERP study. *European Journal of Neuroscience*, vol. 33, no. 9, pp. 1716-1723, 2011.

[87] A. T. Sohaib, S. Qureshi, J. Hagelbäck, O. Hilborn, and P. Jerčić. Evaluating classifiers for emotion recognition using EEG. *Foundations of Augmented Cognition*, pp. 492-501. Springer Berlin Heidelberg, 2013.

[88] M. Soleymani, M. Pantic, and T. Pun. Multimodal emotion recognition in response to videos. *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 211-223, 2012.

[89] M. A. Parvaz, N. Alia-Klein, P. A. Woicik, N. D. Volkow, and R. Z. Goldstein. Neuroimaging for drug addiction and related behaviors. *Reviews in the Neurosciences*, vol. 22, no. 6, pp. 609-624, 2011.

[90] M. C. Van de Laar, R. Licht, I. H. Franken, and V. M. Hendriks. Event-related potentials indicate motivational relevance of cocaine cues in abstinent cocaine addicts. *Psychopharmacology*, vol. 177, no. 1-2, pp. 121-129, 2004.

[91] X. Liu, D. B. Vaupel, S. Grant, and E. D. London. Effect of cocaine-related environmental stimuli on the spontaneous electroencephalogram in polydrug abusers. *Neuropsychopharmacology*, vol. 19, no. 1, pp. 10-17, 1998.

[92] D. J. Kim, J. Jeong, K. S. Kim, J. H. Chae, S. Jin, K. J. Ahn, H. Myrick, S. J. Yoon, H. R. Kim, and S. Y. Kim. Complexity changes of the EEG induced by alcohol cue exposure in alcoholics and social drinkers. *Alcoholism: Clinical and Experimental Research*, vol. 27, no. 12, pp. 1955-1961, 2003.

[93] V. Knott, M. Cosgrove, C. Villeneuve, D. Fisher, A. Millar, and J. McIntosh. EEG correlates of imagery-induced cigarette craving in male and female smokers. *Addictive Behaviors*, vol. 33, no. 4, pp. 616-621, 2008.

[94] R. N. Ehrman, S. J. Robbins, M. A. Bromwell, M. E. Lankford, J. R. Monterosso, and C. P. O'Brien. Comparing attentional bias to smoking cues in current smokers, former smokers, and non-smokers using a dot-probe task. *Drug and Alcohol Dependence*, vol. 67, no. 2, pp. 185-191, 2002.

[95] M. Field, K. Mogg, and B. P. Bradley. Eye movements to smoking-related cues: effects of nicotine deprivation. *Psychopharmacology*, vol. 173, no. 1-2, pp. 116-123, 2004.

[96] W. G. Shadel, R. Niaura, and D. B. Abrams. Effect of different cue stimulus delivery channels on craving reactivity: comparing in vivo and video cues in regular cigarette smokers. *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 32, no. 4, pp. 203-209, 2001.

[97] C. Tong, D. H. Bovbjerg, and J. Erblich. Smoking-related videos for use in cue-induced craving. *Addictive Behaviors*, vol. 32, no. 12, pp. 3034-3044, 2007.

[98] M. A. Sayette and M. R. Hufford. Urge and affect: a facial coding analysis of smokers. *Experimental and Clinical Psychopharmacology*, vol. 3, no. 4, pp. 417, 1995.

[99] P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual. NIMH, Center for the Study of Emotion & Attention, 2005.

[100] N. D. Volkow, G. J> Wang, F. Telang, J. S. Fowler, J. Logan, A. R. Childress, M. Jayne, Y. Ma, and C. Wong. Cocaine cues and dopamine in dorsal striatum: mechanism of craving in cocaine addiction. *The Journal of Neuroscience*, vol. 26, no. 24, pp. 6583-6588, 2006.

[101] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49-59, 1994.

[102] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, vol. 23, no. 12, pp. 1080-1093, 2005.

[103] M. B. Stegmann, B. K. Ersbll, and R. Larsen. FAME-a flexible appearance modeling environment. *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, pp. 1319-1331, 2003.

[104] M. Pantic, M. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. *IEEE International Conference on Multimedia and Expo*, 2005.