Pedro Sérgio Oliveira Branco

Computer-based Facial Expression Analysis for Assessing User Experience

Setembro de 2006

Universidade do Minho
Escola de Engenharia

Pedro Sérgio Oliveira Branco

Computer-based Facial Expression Analysis for Assessing User Experience

Tese de Doutoramento
Tecnologias e Sistemas de Informação
Engenharia da Programação e dos Sistemas Informáticos

Trabalho efectuado sob a orientação dos Professores
Doutor Adérito Fernandes Marcos
Doutor Luis Miguel Eggeling da Encarnação

Setembro 2006

Acknowledgements

Throughout the last few years I have had the great privilege to work with friends on engaging and interesting projects, involving exciting technologies. I have to be thankful for the opportunity I have had to work in such an inspiring environment and to be influenced by a number of people that contributed directly or indirectly to this work.

I would like to express my thanks and appreciation to Prof. Dr. L. Miguel Encarnação for all his support throughout the years, mentoring, supporting this research, and being a friend. Thank you to Prof. Dr. Adérito Marcos for overlooking the progress, motivating and patiently awaiting the results.

Numerous students have contributed their time and knowledge in supporting this work, without them the accomplishments would have taken much longer. In particular, my special thanks to Peter Firth and Nada Jaksic who actively collaborated in the studies presented in this work.

Thank you to Dr. Paolo Bonato and his staff from the Motion Analysis Laboratory at Spaulding Rehabilitation Hospital, for making his laboratory available for our work. I am also grateful to Dr. Louis Tassinary for letting us use the facial electromyogram system. Thank you also to Dr. Laura Richterich from the Usability group at Fidelity Investments eBusiness for her collaboration and time.

This work would have not been possible without financial support. I am grateful to the INI-GraphicsNet Foundation Fellowship Program, for partially sponsoring this work. My thanks also to SAP Research Center in Palo Alto, and in particular to Dr. Frankie James, and to the Fraunhofer IGD Virtual Human Project, as well as the NSF Research Experiences for Undergraduates (REU) Program (NSF03-599), for all sponsoring this work.

Without doubt there will be errors, omissions and over-simplifications, for which I take absolute responsibility, while hoping that the rest of the material will be enough to stimulate insights and new trains of thought.

To my Family and my Mother for all the comforting and supportive messages, to the Special Friends around that made the time so much more pleasant,

Thank You!

Abstract

For the majority of the users, computers are difficult and frustrating to use. The proliferation of computers in the daily life in all sort of shapes and forms becomes a significant factor for potentially aggravating and thus degrading the users' acceptance of the technology. Traditional user observation methods, aiming at improving human-computer interaction, concentrate on monitoring users within usability laboratories. To accompany the users and follow their difficulties in daily interaction with the computer or other interaction devices would present a solution to better form a picture of the quality of the users' experience. Such strategy, though, has to rely on natural ways of assessing the users' reactions, with the risk of otherwise becoming an intrusive and undesirable feature what would be turned off. Our approach borrows from natural human social interactions where body language is so much assertive of one's appraisals and reactions. We, in particular, concentrate on facial expressions as a clue to infer the users' perceived positive and negative situations.

We start by reviewing the literature on psychology and affective computing regarding users' emotions and specifically the interpretation of facial expressions. We then present a comprehensive experiment and study asserting the relation of facial expressions with the user task difficulty while performing a word processing task. Subsequently, we discuss different techniques to monitor user state, and as a result focus on machine vision as a solution for unobtrusive facial expression analysis. Consequently, further investigate computer interfaces that react to users' facial expressions as a method to provide users' assistance. A

related application-oriented experiment and study is presented assessing users' reactions to that monitoring modality within the context of interacting with a virtual shopping assistant.

We conclude by acknowledging that while current technology certainly limits the scope of the research and the applicability of this form of user monitoring, the promising results presented strongly motivate future use of facial expressions analysis to assess the users' experience with interactive systems.

Resumo

A maioria dos utilizadores vê o computador como uma ferramenta de difícil manuseamento que por vezes conduz a experiências frustrantes. Nesse sentido, a proliferação de computadores no dia-a-dia, em todas as suas formas e variedades, é um factor potencial para degradar a aceitação da tecnologia.

Os métodos tradicionais de observação do utilizador, no sentido de melhorar a interacção Humano-Computador, conduzem a monitorização em laboratórios de usabilidade. Em contraste, o acompanhamento do utilizador e das suas dificuldades nas interacções diárias com o computador, ou outros sistemas de interacção, apresenta-se como uma melhor solução para se perceber a qualidade da experiência. Essa estratégia, no entanto, tem que ter como base métodos naturais de avaliação da reacção do utilizador, correndo, caso contrário, o risco de ser visto como uma funcionalidade intrusiva e desnecessária. O nosso método tem por base formas de comunicação não verbal que estão presentes nas interacções sociais onde são frequentemente usadas para avaliar a compreensão e dificuldade do outro. Especificamente, focamos nas expressões faciais como pista para inferir sobre situações positivas e negativas sentidas pelo utilizador.

Neste trabalho, começamos por apresentar uma discussão da literatura na área da psicologia e computação afectiva em relação às emoções e interpretação de expressões faciais. Apresentamos um estudo que estabelece a relação das expressões faciais com as dificuldades sentidas pelo utilizador no uso de um processador de texto. Discutimos diferentes técnicas para monitorizar o estado do

utilizador, e focamos na visão por computador como uma solução para analisar expressões faciais. Investigamos ainda interfaces que reagem às expressões faciais como forma de assistir o utilizador. É apresentado um estudo que avalia a opinião dos utilizadores a interfaces que monitorizam as suas expressões faciais, no contexto da interacção com um assistente virtual, num website de comércio electrónico.

Concluímos, reconhecendo que a tecnologia existente é um factor limitante no estudo e aplicabilidade da forma descrita de observação do utilizador. No entanto, os resultados apresentados são encorajadores tendo em conta uma aplicação futura da análise de expressões faciais do utilizador, como forma de avaliar a qualidade da experiência com sistemas interactivos.

Contents

List of Figures

List of Tables

CHAPTER 1 Introduction

Certainly, computers should be designed in ways that make the user experience more pleasant for everyone, including users of various ages, cultural backgrounds, and economic situations, which is known as universal usability. But while universal usability is the end goal for technology, it is important to first examine the root causes of user frustration, from a social psychology point-of-view. It is possible that there are other techniques, aside from improved computer design, that could impact or lessen user frustration.

— Katie Bessiere et. al. 2002

1.1 Motivation

Behind us are the days where computers were only accessible by experts and designed to satisfy the needs of engineers and scientists. Computers infiltrated all aspects of our lives widening the age, educational background and the cultural diversity of users. In 1980, the number of computers in use worldwide was estimated to be five million, by 2005 that number has multiplied by a factor of 200, to an estimated total of 1,044 million (etForecasts 2006). The drop in hardware cost has promoted computing devices that extend beyond the desktop: cell phones, pocket-based computers, digital video recorders, entertainment appliances, navigation systems, are increasingly present on our daily activities and occupying a significant part of our time. And it does not stop there, information/ web appliances, mobile information appliances or the web cellular devices are the new emerging technologies. Facing this plethora of offers, the consumer focus is on the degree to which those technologies satisfy a specific need, their design, and more importantly, their ease of use.

Unfortunately, this latter desirable quality, the ease of use, seems to be often violated. The growing complexity and variety in software and hardware solutions, on one hand, and the broadening of the users' pool on the other, are certainly contributing factors. These factors expose challenges that, if not dealt with, may hamper some of the technology benefits. The difficulty and consequent frustration experienced interacting with an interface, and computing devices in general, can be personally disturbing and socially disruptive (Bessiere et al. 2002). It can potentially lead to rejection, and contribute for technological illiteracy. The following numbers are cause for concern: Shneiderman (2002, 25) reports a survey of six thousand computer users where it was found that users wasted an average 5.1 hours per week in trying to use computers; in another study nearly one-third to one-half of the time spent in front of the computer was determined wasted due to frustrating experiences (Lazar et al. 2003); a subsequent study of specific causes and effects of user frustration (Lazar et al. 2004), found that 42% of student and 58% of workplace users reported being angry at the computer as a result of a technical problem encountered during daily tasks.

The paradigm of user-centered design, being advocated for some years now by the Human-Computer Interaction (HCI) community, is certainly an approach to avoid some of the pitfalls that lead to the frustration which users experience. In particular, observational methods were developed to monitor the users experience and capture their opinion. Nonetheless, the reality previously enumerated renders that task increasingly difficult: the broadening range of users with different skills and set of requirements as well as the diversity of contexts and devices, render implausible that all the potential problems can be observed in a usability testing laboratory. A one-solution-fits-all approach falls short of the desired goal. Also, the economic pressures imposed on manufacturers to release new products at shorter cycles, to gain and maintain competitive advantage, certainly does not contribute to mitigate the problems. The consequence is a negligent oversight of the adverse aspects of ill-designed human-computer interaction as well as a lack of end-user consideration and support at runtime.

In the face of the constraints listed above, there is a need to capture the users' experience in a live environment, within the context of their activities; not in an artificial environment, with a few selected individuals. However, how can we proceed without being too intrusive, without requiring the user to explicitly state problems (which in itself is a challenge for non-computer-experts), without forcing the user to divert from their task or goal in order to diagnose and resolve a problem situation? And how can the human-computer dialogue of a large variety of users be discretely exploited for software and hardware manufacturers to pro actively improve their products *before* design flaws create a negative impact on a manufacturer's business?

There is not a single solution to all these raised issues. Most likely the path to success will depend on a variety of approaches. The path followed in this work explores the notion of perceptual interfaces, capable of automatically observing the users in their habitual computer environments and recognizing their difficulties. In the next sections we will explain the approach taken.

1.2 Problem Statement

It is striking to realize that, for the most part, despite the widespread promotion of usable interfaces and appeals for increasing usability evaluation, the only vehicle to directly observe the user experience is by monitoring it within the constrains of a usability laboratory. An ideal usability evaluation scenario requires a realistic and unrestricted observation of the user, extended in time, and diverse in its participants. Those requirements would be easier to match beyond the walls of an usability testing environment, unfortunately extending the observation methodology remotely presents also a series of difficulties though. The unconstrained environment makes harder to collect relevant data, users tasks may extend for undefined periods of time, interlaced with other activities. A large pool of users would translate in prohibitive amounts of data that would need to be collected, transmitted and analyzed in order to identify the problematic aspects of the interaction. An alternative approach, that surpasses those difficulties, is to provide the user with mechanisms to self-report usability incidents; while that strategy has been shown to provide data comparable to usability experts (Hartson et al. 1996), it poses practical difficulties on itself. For once, there is the need for the user to disengage from their activity, that might justify why users self-reporting was found to be often significantly delayed from the time when the problem was actually encountered (Hartson and Castillo 1998). Also especially non-expert users can potentially attribute a particular adverse occurrence as their own-fault and not necessarily a usability incident; in the studies above the participants were shortly briefed to recognize such events.

Taking as a comparison the software crash feedback system, which e.g. the windows operating system offers as a feedback allowing to send in information pertaining to the context that caused the software failure. This is possible since the operating system has the knowledge when the failure occurred. In an software usability feedback system, however, there is not a obvious indication when problems occur. The problem, therefore, is how to detect the occurrence of critical incidents, "events that stand out during usability evaluation, e.g. major breakdowns in a user's performance or understanding." (Koenemann-Belliveau et al. 1994). How to detect when users' start perceiving problems?

1.3 Approach

In the previous section we discussed the difficulties in directly observing the user actions and detecting usability problems. Part of the problem lays on the fact that there is not an obvious automatic way to early detect the potential difficulties encountered by users.

In human-human dialog we are able to interpret body language and the subtle clues that make us aware about the others' comprehension, agreement, disagreement and emotions. That ability to accurately perceive emotions is intrinsic to the conduciveness of normal social interactions, and in fact its absence is impairing for normal social interactions. Salovey and Mayer definition of Emotional Intelligence (1990) acknowledges that skill:

> *ability to monitor one's own and others' feelings and emotions, to discriminate among them and to use this information to guide one's thinking and actions.*

Body language is also peoples' preferred communication channel, people react emotionally when struggling with a problem, leaking expressions and other non-verbal behavior.

This important aspect of the users behavior is currently being completely ignored by computers. Picard traces a parallel between autistic, severely impaired in social-emotional skills, and the computers inability to recognize users' emotions (2000).

Our research focuses on investigating if new approaches for detecting and anticipating usability problems can be devised from systems that understand users' display of emotions. The methodology proposed is new in the sense that we depart from the traditional methods in usability engineering and look (literally) to the user as a source of information. We discuss the importance of body language, and more specifically spontaneous facial expressions, in identifying users' reactions and present how those clues can be relevant to identify critical incidents.

We furthermore discuss the reasons why real-time recognition of users' spontaneous facial expressions can be helpful to complement traditional usability analysis process, be a basis to collect usability information through remote usability testing, and in the process, provide a natural communication modality for the user to voluntarily express satisfaction within the human-computer dialogue.

The computer recognition of negative emotions, or psychological states, would allow for a system that logs the most problematic occurrences, similar to the software crash report mentioned previously.

WARNING !

This machine is subject to breakdowns during periods of critical need.

A special circuit in the machine called a ‘critical detector’ senses the operator’s emotional state in terms of how desperate he or she is to use the machine. The ‘critical detector’ then creates a malfunction proportional to the desperation of the operator. Threatening the machine with violence only aggravates the situation. Likewise, attempts to use another machine may cause it to also malfunction. They belong to the same union. Keep cool and say nice things to the machine. Nothing else seems to work.

FIGURE 1-1. Note found hanging over a photocopier

1.4 Summary of Contributions

The argument presented in this work is that in current interactive systems there is a whole channel of communication that is being ignored despite providing precious insight in user interaction difficulties. We propose to observe users’ spontaneous facial reactions to detect critical incidents, thus anticipating usability problems. We present the arguments why facial expressions can be effective in identifying those occurrences, based on scientific studies and experiments conducted. And we propose a system in support of implementing such unobtrusive user-monitoring capabilities based on state-of-the-art computer vision technology.

1.5 Thesis Organization

The content of the remaining chapters are summarized bellow.

Chapter 2 starts with a discussion on the various types of affect and its relevance to HCI, it follows with a discussion on emotions as usability indicators.

Chapter 3 discusses measures of user physiology and how they relate with different psychological and emotional states. It looks to previous work on physiological sensing in HCI, and proceeds focusing on facial expressions.

Chapter 4 introduces a study investigating the association between facial expressions and task difficulty. It discusses its design and present the results, it serves as a base for the discussion following in the next chapters.

Chapter 5 examines the different computer vision solutions for facial expression analysis, it discusses the implementation of a system for supporting the research on monitoring users facial expression monitoring.

Chapter 6 elaborates on two study cases serving as a platform for evaluating applicability of the research conducted

Chapter 7 summarizes the work and offers a road mop for future research.

CHAPTER 2

Emotions meet Usability Analysis

Indeed any interface that ignores a user's emotional state or fails to manifest the appropriate emotion can dramatically impede performance and risks being perceived as cold, socially inept, untrustworthy, and incompetent.
— Scott Brave and Clifford Nass 2003.

The use of the words emotion, mood, feeling, and computers in the same sentence would seem to classify as an oxymoron. The intrinsic rationality behind computers, at least outside the entertainment domain, seems to reject any role of affect in computer-to-human or in human-to-computer dialogue. Notwithstanding a new paradigm has emerged, influenced by the developments of the understanding of the emotional process. As argued by Damasio (1994), Reason and Emotion are not separate processes. Emotions are not an appendix of human cognition that can be disconnected, and it is, therefore, not likely that computers or interaction with computers would be an exception. Reves and Nass (1996) in fact emphasize the predisposition towards social behaviors that users exhibit in computer interaction**.** In a surprising result they concluded that in fact users tend to treat media and in particular computers by the same rules that govern human to human interaction. This result can be explained by the fact that all the social machinery humans are equipped with evolved along million of years, through the interaction with other humans. In comparison, the very recent ability of the computers to behave as an interlocutor does not, suddenly, inhibit those innate responses.

Even if one is not interested in developing applications that explore the social aspects of human-computer interaction, and focuses on more, lets say, “mundane” applications: a spreadsheet, a word processor, an email application, the aspects intrinsic to the human nature such as satisfaction and frustration, are still present and determine to a big extent the user experience.

This chapter discusses the relevance of emotions as a component to measure the user experience and detect negative occurrences. We contextualize this work within the Usability Analysis and Affective Computing research area and proceed to discuss relevant emotional models and the emotion aspects that might emerge.

2.1 Distinguishing Emotion from other Affective phenomena

Given the range of lexicon that is ambiguously associated with the emotional process, namely emotions, feelings, sentiments, moods, it is necessary, for the sake of clarity, to establish the definitions used throughout this thesis. One should not expect though, that those words describe clearly distinctive phenomena, non-overlapping in their definitions. After all, the origin of those terms emerged with natural language development, and not from a scientific basis. Albeit, the existence of those words does provide some guidance on the distinctions that are worth making (Frijda 1994, 59).

Affect is the term that refers to the topic of emotions, feelings and all the semantically related words. We will be using “affect” when referring in general to the entire topic of emotions, including the other phenomena such as feelings, moods, and sentiments.

2.1.1 Emotion

No better statement on the complexity of this topic, than the fact that there is not an agreement among scholars on an universally accepted definition of emotion.

Ekman and Davidson (1994a, 412) list a set of characteristics of emotions that are agreed on by most scholars:

- **There is information processing and evaluation of events associated with emotion, though that process is not necessarily conscious.**
- **There are expressive and physiological changes associated with emotions, though there is divergence whether all emotions are distinguishable.** Particular emotions might demand metabolic requirements to respond to the stimulus, for instance increase in heart rate,..., It is not clear to what extent different emotions have a distinctive observable metabolic demands, for instance the heart rate also goes up with.... There is evidence though of a differentiation between positive and negative emotions.
- **Associated with the emotion process, there is the retrieval of relevant memories and thoughts for coping with the emotional stimulus**.
- **Involves a subjective experience, a feeling state**. Again, here, there is disagreement if this is a separate, distinctive process from emotion itself (Damasio 2000, 15).

A distinction can be made between primary and secondary emotions (Damasio 1994, 131). Primary emotions evolved as a fast response mechanism to prepare the organism to fly or fight. Examples of primary emotions are the startle reaction, also called “startle response” or “alarm reaction”, in response to a sudden unexpected stimulus, such as a flash of light, a loud noise, or a quick movement near the face. Primary emotions in HCI can be observed, for example, in response to the sudden appearance of a pop-up window (Brave and Nass 2003, 83).

Potentially more common in HCI are the secondary emotions. These result from higher level processing of a stimulus. For example realizing the wrong file was overwrote, or even internally generated stimuli like predicting difficulties executing a certain task. The different neurological pathways that support the primary and secondary emotions are highlighted in Figure 2-1. Low and high road pathways carry the information about external stimuli from the sensory thalamus to the amygdala, corresponding to the primary and secondary emotions respectively.

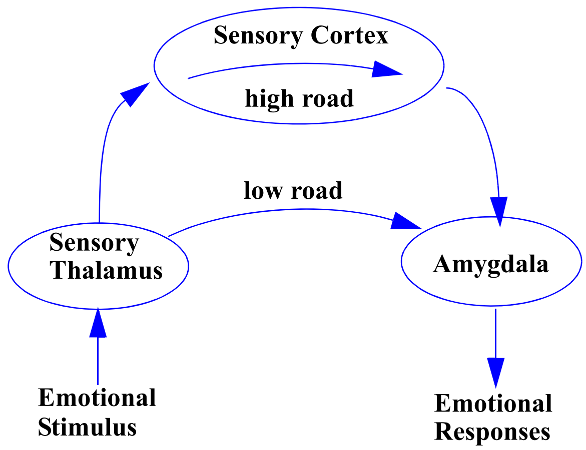


FIGURE 2-1. The low and the high roads to the amygdala. Source: LeDoux 1996, 164.

2.1.2 Mood

Mood distinguishes from emotions by its time pattern and object-directedness (Frijda 1994). While emotions last from a brief interval of seconds to minutes, moods extend for longer periods of time, lasting hours or even days. Also, though moods might be indirectly triggered by a particular object, they are not directed towards any object in particular, thus are experienced as more diffuse, global, and general than emotions (Clore and Ortony 2000, 26).

The resulting state from an emotion that occurs frequently or continuously should be referred as a mood (Damasio 2000, 16). This view is similar to Picard's model of mood where she employs the analogy of a bell (emotional stimulus) that is repetitively strike prolonging the duration of the sound and increasing its amplitude (Picard 1997, 145). This conceptual model can be translated to HCI as stating that a series of negative occurrences will most likely induce negative mood on the users.

Another feature distinguishing moods from emotions, and particular interesting to our discussion, is that moods do not own their own unique facial expression while many of the emotions do (Ekman 1994).

2.1.3 Sentiment

Sentiment is also often confused with emotion. Unlike emotions (and moods), sentiments are not states of an individual, but assigned properties of an object. While emotions and moods are fleeting-emotions lasting only seconds and moods lasting for hours or even days, sentiments can persist indefinitely and are thus responsible for guiding our propensities to seek out or avoid particular objects and situations (Brave and Nass 2003, 84). Negative sentiments towards applications, which are often reported by users. would under this definition seem to have a long term influence on users.

2.2 Affect in Human-Computer Interaction

Affective computing was defined by Picard as “computing that relates to, arises from, or deliberately influences emotion.” (Picard 1997, 3). The definition encompasses the different categories of affect, and includes both, systems that actively influence the user affective state, and systems that capture the user affect.

In general, the applications user interact with are not affective, as they do not seek to actively influence (or monitor) affect. But necessarily, and as highlighted in the introduction, the emotions are integral part of humans and the context of computer interaction does not escape that reality. To better understand the influences of the different classes of affect that were discussed in the previous section, we present a classification in Table 2-1. This is a simplistic depiction for the purpose of understanding how the different classes of affect play a role in the HCI experience. User internal stimuli, and influences external to the interaction itself were not taken into account, as we are focusing on the HCI factors that contribute to affective reactions. Included in the table is the role of affect in attention, memory, performance and assessment, compiled in part from Brave and Nass (2003).

TABLE 2-1. Category of user affect

	emotion	mood	sentiments
caused by:	event	sequence of emotions	long term usage
	content	design	predisposition
		nature of task	mood towards
influences:	arousal	biases assessment.	acceptability/disagreeable
	directs attention	attention to mood congruent stimuli	loyal /disregard
	improves memory recall (especially negative emotions)	mood dependant memory recall	
	mood	positive moods increase efficiency	
		sentiment	
measurable by:	behavior/physiological	some behavioral	
	questionnaire	questionnaire	questionnaire
interface causality:	temporarily localized amplitudes	short term temporal convergence	long term temporal convergence

Lets exemplify how the different categories of affect influence the user experience. Say a user was given the task to create a couple of graphics from the data on a spreadsheet and preparing them into a report. He might find the process of importing the graphics somewhat cumbersome and experience difficulties embedding the graphical objects. This occurrence will potentially aggravate and influence the user's emotional state. Eventually, the repeated trials and errors will negatively impact the user mood. Down the road, similar and other negative occurrences with that application might reoccur and a negative sentiment towards the application might be developed.

In this simple account for emotions, mood and sentiments, we did not discuss personalities or temper. Those concepts are, as we might suspect, important when considering emotions. They might be seen as filters of situations. For example, the user in the task described above might have had a quite calm personality, and despite the stimuli not feel particularly aggravated by the situation. On the other hand, a short-tempered user would be immediately angered at first signs of problems and much quicker develop negative mood and sentiments towards the system. This implies that the response to a HCI stimulus, otherwise capable of provoking an emotional response, will to some extent be user dependent.

2.3 Emotions as Usability Indicators

> *Nearly one-third to one-half of the time spent in front of the computer is wasted due to frustrating experiences.*
>
> — Lazar et al. 2003

> *Forty two percent of students and 58% of workplace users reported being angry at the computer as a result of a technical problem encountered during daily tasks.*
>
> — Lazar et al. 2005

> *Four in five users (83%) have experienced difficulties when using their computer within the past 12 months, and seven in 10 (70%) of these admit to shouting, swearing or being violent towards their computer when these problems arise.*
>
> — MORI 2002

Usability analysis refers to a variety of methods aimed to evaluate the quality of use of a system, or interface. It has been recognized, in the particular case of HCI, as a good practice to improve the quality of the software use. Traditionally, error rate and speed, and quality of task execution, are among the quantitative measures extracted from usability studies. The underlying assumption is that the more efficient the interface is, the more usable it is considered. In recent years, and given the increasing interest in the topic of emotion in HCI, there has been an additional set of hedonic factors added to those usability metrics: user satisfaction[1], pleasure and trustworthiness gained significant importance in assessing the user experience. Here, the underlying assumption is that besides the users performing the task efficiently, they should perceive the experience as enjoyable and fun, in order to improve task performance over time.

The integration of these hedonic factors that we just described has been, for the most part, the avenue by which emotions have been introduced in the usability field. In fact, under the definitions that were presented in the previous section, the word emotion should be replaced, in that case, by mood and/or sentiments. The discussion that we present, of emotions in usability, differs from the approach above. We refer to emotion in the strict sense of the word, we are interested in measuring the ongoing physiological responses from the user, and not the mood or sentiment that might be invoked over time interacting with the software, or from the interface. Naturally, and as previously discussed, these are not disjunctive approaches. The cascade of emotions experienced might eventually and most likely influence the users' mood, and eventually determine their sentiment towards the system. The difference is the possibility that emotions posed in being immediate and visible through non-verbal behavior, while the assessment of moods and sentiments have to be performed in much more indirect ways, and might be observable in longer term rather than the short-term reactions. In order to infer problems from the interface we have to pin point when they occur. We cannot do that with mood/sentiments. In the following paragraph, we will discuss in more detail the possibilities this approach presents.

Several studies suggest that computer problems occur too frequently and are a major source of wasted time and frustration for the users. The results presented at the beginning of this section summarize the degree of frustration incurred in daily computer use. On the face of such adverse occurrences, negative emotions emerge with potentially very extroversive consequences. The first and foremost approach to avoid such occurrences is by following an appropriate methodology for the design of software applications or interactive systems in general. The

1. in fact this is a traditional measure, but was neglected as being very subjective.

usability research has, through the years, proposed a variety of methods supporting that philosophy. The application of such user-centered strategies is a necessary step to minimize adverse occurrences but just to a certain degree. It is not possible to account for all the problems that might result from the different user expectations on how it should work, expertise levels, the variety of contexts the application is used in, its complexity, just to name a few factors. The ability to observe the users' emotional reactions, especially the occurrence of negative emotions, would provide a mechanism that allowed for the detection of adverse events, necessarily a first step to solve them. Spontaneous emotional reactions can, at least partially, be observable through non-verbal behavior; its characteristic of rapid onset allows then to pinpoint the occurrence of problems. This contrasts with the assessment of mood and sentiments which cannot be accounted for in non-verbal behavior and have a longer term effect.

This approach can be compared to the recently introduced mechanism of software crash reports. In a similar problem scenario of the usability field, the increasing variety of systems and configurations makes testing difficult and makes it unavoidable that crashes occur, despite software design and testing guidelines. The recent operating systems are capable of detecting the application crashes, compiling a log, and, if the user wishes, report those events. However, while this mechanism only reports problems after the fact, our approach holds the promise to be able to anticipate major problems before they lead to disaster.

The approach of detecting users' emotional reactions can be also applicable in traditional usability testing environments. Here, the monitoring of users' emotional responses can become an important tool to complement the traditional analysis. In usability studies observational method, the users' likes, dislikes, difficulties as they interact with the system are inferred by the usability specialist. The users' non-verbal behavior can provide in addition important clues. In fact, some usability labs employ this monitoring through video cameras. The recorded video is, though, at this point mostly used to be incorporated in the presentation of the analysis results to emphasize the users' reactions and make the point across more easily. The potential is thus created to allow machines that understand those same clues.

2.4 Discussion

The consideration of interactive systems that recognize affect, and in particular emotional reactions, raises a wide range of questions. A broad discussion of this theme deserves more space than what can be given here. We will, therefore, limit the discussion to the emotional monitoring as an usability indicator.

One important consideration, regarding emotionally aware interactive systems, is its obstructiveness, or the danger of getting too much in the way of the user. Ball (2002, 316) summarizes well what should be the underlying strategy:

> *In an emotionally aware interactive system, the recognition and simulation of emotion will play an auxiliary and probably quite subtle role. The goal is to provide an additional channel of communication alongside the spoken or graphical exchanges that carry the main content of the interaction. If the emotional aspects of the system call attention to themselves, the primary motivation of producing natural interactions will have been defeated. In fact, users that get the feeling that the system is monitoring them too closely may begin to feel anxious or resentful (of course, the emotional system, recognizing that fact, could always turn itself off!).*

The presented proposal of recognizing user emotions to detect and log the context of negative and positive emotions takes advantage of this extra communication channel without demanding attention, or interrupting the user in any way. Eventually, other more active interface responses to those emotional events would have to be sensible to that point. We will come back to this point in Chapter 6.

Another relevant point that might be raised, especially by those who oppose the idea of having a computer monitoring the user or the computer taking initiatives without user intervention, is: Why not rather ask the user to explicitly communicate when they feel frustrated; would it not solve some of the intrinsic problems being discussed? That approach was explored by Reynolds (2001) where users' explicitly communicate their frustration. Different devices were studied, ranging from a feedback interface to a pressure sensitive mouse capable of detecting how hard someone presses it. Clearly Reynolds' strategy has the advantage of being less prone to inference errors, since the users is explicitly communicating. The disadvantage of such approach is the need for the user to disengage from the task at hand to report the frustration. Depending on the severity of the problem, that extra step might not deserve the consideration from the user when more urgent matters call for attention. Also that extra step might cause frustration in itself.

Are all applications suitable for emotional monitoring? Certainly not. After all, certain communication or entertainment applications are capable of producing emotional stimulus by its content, e.g. a media player or a chat application. The users' reactions that could eventually be detected would not be necessarily resulting from interface events.

CHAPTER 3

Towards Inferring Emotional Valence through Physiological and Behavioral Observation

Your emotions affect every cell in your body. Mind and body, mental and physical, are intertwined.

– Thomas Tutko

Consider the following scenario: You are running short on time, the software does not seem to share your urgency: you need to insert a few pictures aligned vertically next to the column of text and you still need to review everything before submitting the final version. The text layout keeps shifting all over the page as you insert the pictures and resize them. You frown, you get agitated and start to sweat, it is obvious that the stress is taking over you.

If we were observing that moment, we would most likely be able to recognize the tension in our user, maybe through the agitated movements, maybe through changes in our user's facial expressions or even the perspiration. The body language, and in particular the grim face, are important clues to understand the user’s emotional distress. The ability to interpret others’ behaviors in terms of their cognitive and emotional state (mental state) is an essential skill in social interactions. However, in human-computer dialogue the computer plays an autistic role, as it is completely indifferent to the dynamics of the user state.

The external manifestations such as the ones described above, are the result of physiological activity associated with particular psychological states, triggered by an organized set of nervous system responses. Posture, eye blinking, facial expressions, blushing and perspiration, are some of the physiological manifestations we learn to observe within social contexts. Top poker players provide a good example where that skill is greatly developed. They learn to read subtle body language of their opponents to the point of being able to discern the increase of the adversary heart rate by observing the carotid artery pulsing (Bromberg 2005)!

For the most part, emotional induced physiological responses are unconscious and out of our control, they constitute a mechanism by which the organism can adapt to be better prepared to respond to stimuli. HCI stimuli are not different from any other types of stimuli. In the presence of an interface event, a cascade of physiological responses might be triggered. For example a sudden appearance of an alert box may startle the user (Brave and Nass 2003). A computer can record those physiological responses, through sensors that measure different signals and patterns of activity over time, and infer about the user cognitive and emotional state induced by the interface or interaction.

This chapter presents a brief account on how the different physiological recordings are representative of particular mental states. We will afterwards discuss the importance of facial expressions, the focal point of the presented research, and finalize discussing our proposed approach.

3.1 Monitoring Physiological Activity

Many of the physiological responses of interest to the study of emotional and cognitive behavior are under the control of the Autonomic Nervous System (ANS), see Figure 3-1. The function of the ANS is to regulate the internal balance of the organism in the presence of internal or external changes by dynamically adjusting physiological parameters such as: heart contraction force and rate, sweat gland activity, skin temperature reaction, skeletal muscle strength[1]. As suggested by its name, the ANS activity occurs automatically and involuntarily.

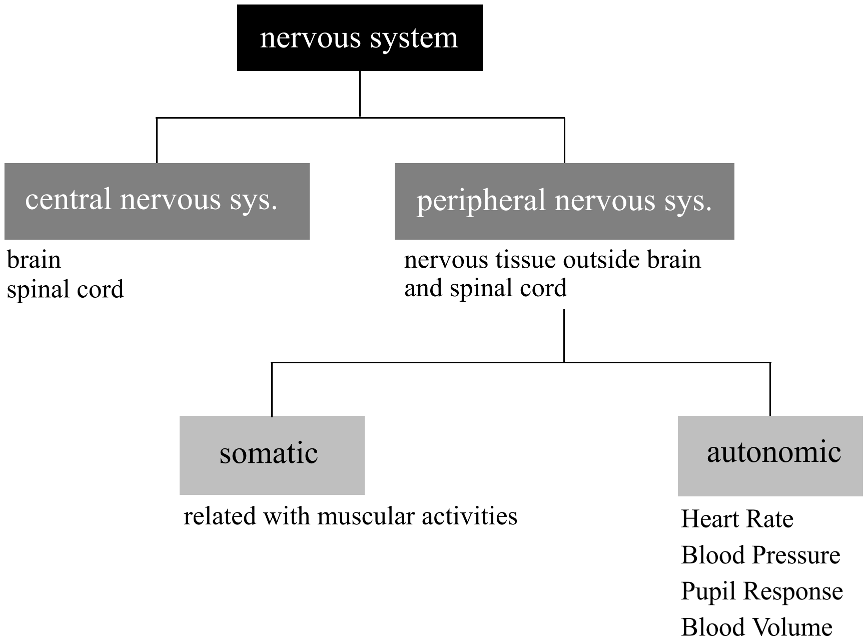


FIGURE 3-1. Simplified representation of the nervous system and the physiological responses under its control. Adapted from Andreassi (2000).

The most typical measurements of ANS activity employed in HCI research on emotional and cognitive states are listed in Table 3-1. Details on each of the physiological measures goes beyond the scope of this work; for in-depth discussion consult Andreassi (2000).

With exception of the pupillary response and eye blinks, which can now be measured with eye trackers, all other physiological measures are recorded through sensors placed on surface areas of the body. Physiological recording is, therefore, a relatively intrusive procedure, first by its nature and second because it requires physical contact with the individual. Its applicability is mostly confined to a laboratory environment. In the particular scenario of human-computer interaction, some physiological indexes can be measured attaching the sensors to devices that

1. Encyclopaedia Britannica Online, s.v. “autonomic nervous system”, http://www.britannica.com/search?query=autonomic+nervous+system+&ct= (accessed 25 August, 2006).

by their function are already in physical contact with the user, such as keyboard, mouse (Ark, Dryer, and Lu. 1999).

TABLE 3-1. Physiological variables and behavioral indicators (Backs and Boucsein 2000)

Physiological measurement	Instrument	Behavior related indicators
Electrodermal activity	Electrodes	Novelty; Intensity; Emotional content, and significance
Heart rate/period	Electrocardiograph	Fear and anger, intensive attention
Blood pressure	Sphygmomanometer	Information processing load and engagement in problem-solving tasks; Anger, Stress
Blood volume	Plethysmograph	new or unexpected stimuli; startled, fearful, or anxious
Pupillary response	Eye tracker	fatigue, workload, novelty and the emotional valence of the stimuli
Eye blink	Electro-oculogram, Eye tracker	task demands, negative emotional states, such as nervousness, stress and fatigue
Brain activity (alpha, beta, delta, theta, gamma waves)	Electroencephalography (EEG)	relaxed states (alpha); novelty, complexity, and unexpectedness, as well as during emotional excitement and anxiety (alpha blocking); active, busy or anxious thinking and active concentration (beta); higher mental activity: including perception, problem solving, fear, and consciousness (gamma)
Facial muscular activity	Electromyogram (EMG)	emotional states

Psychological changes are far from being the only source of physiological responses. Physical activity, internal and external factors to the organism concurrently influence the physiological activity. Table 3-1 highlights just the behavioral influences.

The next section discusses previous work on the application of physiological monitoring to identify emotional and cognitive states within the context of human-computer interaction.

3.1.1 Previous Work on Physiological Monitoring in HCI

Measuring users' physiological indexes during a computer related task opens a communication channel between the user and the computer, enabling the integration of human factors aspects in the design of interactive systems. We examine here the literature on the different approaches that have been explored by physiological monitoring in HCI.

3.1.1.1 Assessment of user performance

The first applications of physiological monitoring in computing systems focused on user performance monitoring. The ability to monitor user alertness, attention, cognitive load, led to research on the relation of those states with the user activity on interactive systems. Better detection performance coincided with higher levels of EEG beta activity as well as lower levels of alpha and theta activity, see discussion by Scerbo, Freeman and Mikulka (2000, 243). The EEG theta activity has been demonstrated as an indicator of fatigue due to mental load in HCI. Also there is reported evidence for reduced heart rate variability (HRV) reflecting the amount of mental effort involved in different laboratory HCI tasks, such as using a text or a line editor with full effort, or typing under time pressure, compared to relaxed use of the text editor or just copy typing (Boucsein 2000, 292). Significant increases in pupil diameter when subjects were required to process information at 75% and 100% of their capacity. However, when required to increase their processing capabilities to 125%, pupillary constriction occurred (see discussion in Andreassi 2000). With respect to task complexity, Rowe, Sibert, and Irwin (1998) study on an air traffic management application showed heart rate variability as a potential physiological measure to correlate with the visual display levels of complexity, when the user performs a monitoring task.

Delays of computer response also triggers changes of physiological signals. Boucsein (2000, 295) reviews studies investigating the impact of the computer response time delays, on the individuals physiological responses, on the presence of time pressure. Short time delays showed different results from longer delays. In shorter delays is noted a increasing of Electromyogram (EMG) frontalis power, respiration rate increased. A study by Wilson and Sasse (2000) reported increases on galvanic skin response (a measure of electrodermal activity) and heart rate and decreases on blood volume pressure when the participants where watching a video with degraded frame rate: five frames per second in comparison to 24 frames per second.

The ability to discern user mental states through physiological indices opened the possibility of exploring systems that adapt to the user. In the early eighties an adaptive computer system was presented, where the distributions of tasks between the operator and the computer could be modified in real time based on physiological workload indices. In another application, physiological measures from the central nervous system adjusted the mode of operation of a task set (manual/automated mix) based on the EEG signal reflecting user engagement (Scerbo, Freeman and Mikulka 2000, 244).

3.1.1.2 Inference of user satisfaction

The literature reviewed to this point discusses assessment of user performance metrics. This focus follows the classical view of performance as the most important metric of success of an interactive system. In the recent years, though, physiological monitoring has been revisited from a different point of view. The focus shifted from the users' cognitive effort, level of attention, alertness to more hedonic parameters such as emotion, well-being and satisfaction. Certainly this trend can be attributed to, on one hand, the neuroscience advances in emotion research and the understanding of emotions as a integral process of human cognition, and, on the other, the pioneering work of Picard exploring the users' emotional aspects.

Picard, Vyzas and Healey (2001) combine measures of respiration, blood pressure volume, skin conductance, as well as the activity of the masseter muscle to differentiate among eight emotions. Ward suggests that the galvanic skin response, a particular measure of electrodermal activity, is indicative of emotionally significant HCI events and situations (Ward et al. 2001).

Scheirer's study, offered a monetary reward and simulated random system freezes to induce users frustration while playing a computer game. The participants skin conductance and blood volume pressure were monitored throughout. The analysis of those physiological signals provided a discrimination of the moments when the system was froze (Scheirer et al. 2002).

3.1.2 Difficulties with Physiological Monitoring

The previously reviewed studies investigated the application of physiological monitoring in interactive systems. All those findings highlight one or a set of physiological indices related to a certain user behavior or state. Those studies occur within well controlled conditions, mostly monitoring the user for relative short periods of time. Outside a controlled environment though, a goal of the current research, there are a number of complications that emerge. We consider in the following paragraphs the intrinsic difficulties of physiological monitoring, and, consequently, proceed in the following section discussing the monitoring of facial expressions as an alternative.

First, most of the measures discussed require physical contact with the user, thus becoming intrusive to apply in a real-world scenario. Techniques to improve physiological readings such as skin abrasion, conductive gel, would not be practical. Eventual alternatives would be the integration of sensors in computer peripherals (Ark, Dryer and Lu 1999), or in wearables, such as jewelry and clothing (Picard and Healey 1997).

Another difficulty is the latency of the physiological signals with respect to the occurrence of the stimulus. Quite frequently the user actions succeed in a fast pace, and if one intends to establish the cause of the user state change, it is not straightforward to find that occurrence if it requires a certain amount of time for the physiological signals to react. Anttonen and Suraka (2005) found that 6 seconds of heart rate data post-stimulus were necessary to differentiate between positive and negative stimuli.

An important difficult aspect of applying physiological monitoring is its variable nature. The organism is constantly responding not only to the psychological aspects of the individual, but to numerous external variables such as time of day, meteorological conditions, physical activity. Drinking coffee, for example, trigger physiological changes that do not relate to psychological states (Quinlan, Lane and Aspinall 1997).

It is also not clear, to what extent the physiological responses differ between mental states. Different mental states may require similar organism response and, therefore, may be similar in terms of physiological patterning (Davidson and Ekman 1994). There is evidence that it is possible to discriminate, based on the response from the ANS, between positive and negative emotions (valence) and among some types of emotions. Also negative emotions may be characterized by greater autonomic activity than positive emotions (Cacioppo et al. 2000, 183) but to what extent the patterning of responses is different between workload, frustration, or other mental states is far from being clear (Davidson 1994).

Given these significant limitations, we proceed, analyzing in detail facial expressions as potential alternative indicators of cognitive/emotional states.

3.2 Monitoring Facial Expressions

The face is the most visible and expressive of all the channels for communication of emotions and, therefore, assumes an importance of its own in non-verbal communication. From the thousands of possible different facial expressions, we learn to observe in others surprise, happiness, sadness, anger, ranging from subtle to more expressive facial expressions. Focusing on facial expressions circumvents the intrusiveness of sensors in physical contact with the user, and, while it inherits the subjectiveness of physiological measures, its interpretation benefits from a large body of previous research.

3.2.1 Different Functions of Facial Expressions

The face is a multi-signal system and there are several reasons that might originate a particular expression, Figure 3-2. Beyond a vehicle for the emotional display, and cognitive processes, facial expressions also convey conversational signals. For example, in the context of a conversation, raising the eyebrows is used to lay emphasis while speaking (Ekman 1997).

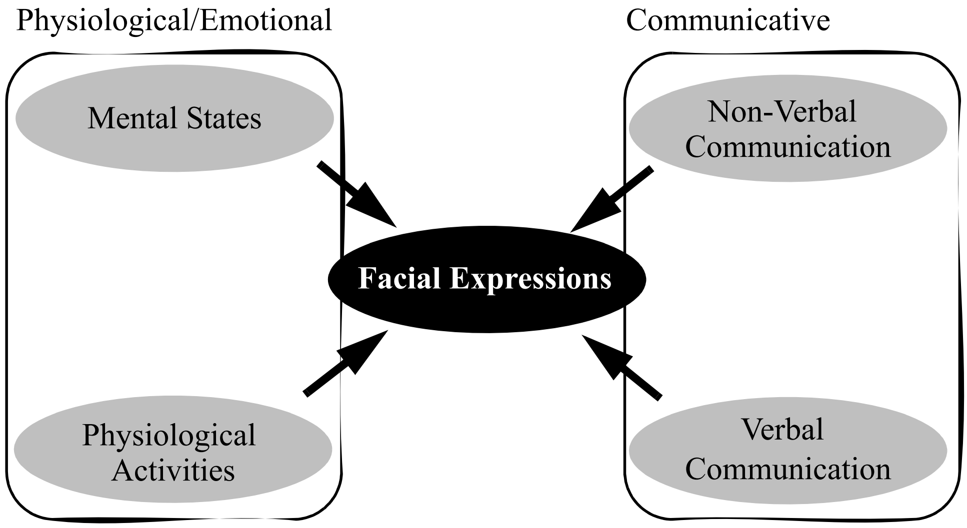


FIGURE 3-2. Different factors that influence facial expressions. Adapted from Fasel and Luettin (2003)

Wehrle and Kaiser (2000) summarize the different roles of facial expressions:

- In speech it is often used by the listener as a back-channel informing the speaker that he can go on talking and that he has been understood (regulator);
- To emphasize a particular message, or to change the meaning of verbal message where the speaker facial expression modifies or contradicts what is being said, e.g. when being ironic (illustrator).
- As a mean for installing, maintaining, or aborting a relationship, e.g., when a couple is discussing a controversial topic, a smile can indicate that although they disagree on the topic there is no “danger” for the relationship.
- An indicator for cognitive processes: e.g., frowning often occurs when somebody does some hard thinking while concentrated attending to a problem, or when a difficulty is encountered in a task.
- An indicator for an emotion (affect display)

From the list above, two major categories of facial expressions can be considered: facial expressions employed as communicative tool, or related to an emotional/cognitive process. Necessarily the decodification of a particular expression in one of the classes listed above is very much dependent on the context, for example a frown might be a deliberated posed expression, or a spontaneous display of emotion. We will go back to this point in Section 3.4, but to infer on the user state, the expressions that we will focus the interest on are related to an emotional/cognitive process.

We present some of the psychology work linking facial expressions with emotions and mental states. We proceed on summarizing the previous work concerning applications of facial expression recognition in human-computer interaction. We then discuss some of the challenges posed in the observation of spontaneous facial expressions in human-computer interaction.

3.2.2 Facial Expressions of Emotion

The debate on facial expressions and emotions has a long history. The validity of facial expressions as indicator of emotions and whether different cultures express, through the face, emotions in a similar way, have been the focus of much debate since Darwin's work on "The Expression of the Emotion in Man and Animals" (1872/1998). A influential work by Ekman (1972) presented a series of studies focusing on the facial behavior of difference cultures. The two major conclusions, that support the view of universality of facial expressions, are that the different cultures exhibit similar spontaneous facial expressions in response to the same emotional stimuli, and also, facial expressions are interpreted in a consistent manner across cultures. The emotions that were most agreed on were: sadness, surprise, anger, disgust, fear and happiness. Notably in these studies two pre-literate cultures were included, that lived in relative isolation, uninfluenced by movies, television, or magazines. The methodology followed in those studies was such as to observe the individuals in isolation, to ensure that cultural or social rules would affect the intensity of the display of the expression. Therefore, those facial expressions could be considered spontaneous and not modulated by any cultural or social rules that could affect the intensity or the display of the expression. In fact, for a long time the face was considered an unreliable source of emotional information; facial expressions were thought to depend exclusively on the social environment. For example there are reports of cultures that do not express anger in the face, and others that laugh at funerals (Keltner and Ekman 2000). The fact is that while there are socially learned rules that dictate which emotions and to whom they are appropriate to exhibit, the spontaneous facial expressions of emotion proved to similar in terms of appearance across cultures.

The current view, supported by other relevant studies, substantiate the link between facial expression and the experience of emotion (Keltner and Ekman 2000). Table 3-2 illustrates prototypical faces of each of the six universal emotions: surprise, fear, anger, disgust, sadness, and happiness, and the appearance changes associated with each. By prototypical is meant that the expression exhibits at full intensity the features that are characteristic of each facial expression; it is therefore a reference expression.

One source for the difficulty in the study of facial expressions was the lack of an unambiguous methodology to describe an expression. Ekman and Friesman (1977) developed the Facial Action Coding System, known as FACS, to accurately encode the description of facial expressions. FACS provides a method to measure and describe the appearance changes in the face resulting from the muscular activity. FACS is purely descriptive and just based on the face appearance; it does not attempt to interpret the facial expression in terms of psychological or affective state. All facial expressions are described through a combination of Action Units (AU) with associated intensities. Each AU represents an action of a facial muscle or group of muscles. Table 3-3 on page 49 lists each AU and the corresponding description, it also list action descriptions that are also part of the FACS, and differentiate from AU as there in not a specific muscular basis for the actions.

TABLE 3-2. Facial expressions of emotion

Basic Emotions	Appearance	Facial Cues
Surprise		brows raised (curved and high) skin below brow stretched horizontal wrinkles across forehead eyelids opened and more of the white of the eye is visible jaw drops open without tension or stretching of the mouth
Fear		brows raised and drawn together forehead wrinkles drawn to the center upper eyelid is raised and lower eyelid is drawn up mouth is open lips are slightly tense or stretched and drawn back
Disgust		upper lip is raised lower lip is raised and pushed up to upper lip or it is lowered nose is wrinkled cheeks are raised lines below the lower lid, lid is pushed up but not tense brows are lowered
Anger		brows lowered and drawn together vertical lines appear between brows lower lid is tensed and may or may not be raised upper lid is tense and may or may not be lowered due to brows' action eyes have a hard stare and may have a bulging appearance lips are either pressed firmly together with corners straight or down or open, tensed in a squarish shape nostrils may be dilated (could occur in sadness too) unambiguous only if registered in all three facial areas
Happiness		corners of lips are drawn back and up mouth may or may not be parted with teeth exposed or not a wrinkle runs down from the nose to the outer edge beyond lip corners cheeks are raised lower eyelid shows wrinkles below it, and may be raised but not tense crow's-feet wrinkles go outward from the outer corners of the eyes
Sadness		inner corners of eyebrows are drawn up skin below the eyebrow is triangulated, with inner corner up upper lid inner corner is raised corners of the lips are drawn or lip is trembling

TABLE 3-3. FACS Action Units (AU) and Action Descriptors (AU)

AU/AD	Description	AU/AD	Description
1	Inner corner of eyebrow raised	31	Jaw clench
2	Outer corner of eyebrow raised	32	Bite
4	Brow lowerer	33	Blow
5	Upper Lid Raiser	34	Puff
6	Cheek raiser and lid compressor	35	Suck
7	Lid tightener	36	Tongue bulge
8	Lips toward	37	Lip wipe
9	Nose wrinkler	38	Nostril dilate
10	Upper lip raiser	39	Nostril compress
11	Nasolabial furrow deepener	45	Blink
12	Lip corner puller	46	Wink
13	Sharp lip puller	51	Head turn left
14	Dimpler	52	Head turn right
15	Lip corner depressor	53	Head up
16	Lower lip depressor	54	Head down
17	Chin raiser	55	Head tilt left
18	Lip pucker	56	Head tilt right
19	Tongue show	57	Head forward
20	Lip Stretcher	58	Head back
21	Neck tighten	61	Eyes turn left
22	Lip funneler	62	Eyes turn right
23	Lip tightener	63	Eyes up
24	Lip presser	64	Eyes down
25	Lips part	65	Walleye
26	Jaw drop	66	Cross eye
27	Mouth stretch		
28	Lips suck		
29	Jaw thrust		
30	Jaw sideways		

FACS scoring requires the experts' careful observation of all the facial changes. They rely on photographs, but preferably on video recordings to annotate the minute changes on the face morphology. A substantial level of training is required, over 100 hours, and in general the coding effort is very time consuming, one minute of video takes one hour to score (Bartlett et al. 1999). Other methods to describe facial expressions have been proposed throughout the years.

FACS is nevertheless the most widely applicable, and influenced inclusively the MPEG-4 standard on facial animation.

3.2.3 Facial Expressions in Relation to Other Mental States

The link between facial expressions and emotions, and in particular basic emotions, has been the most studied. There is nevertheless work presenting evidence that more complex mental states, such as interest, boredom and confusion, to name a few, can also be inferred from visual cues. Baron-Cohen (1995) presents a test where individuals attribute mental states to faces. There is a significant inter-individual agreement of which images reflect what mental states.

The recognition of cognitive mental states requires the fusion of multiple information sources such as changes of pose and head movements, eye gaze in addition to facial expressions (El Kaliouby and Robinson 2004b). Each of those events occur in different time intervals, requiring therefore the analysis of those events at multiple time levels. This approach seems therefore to add another layer of complexity in comparison to an interpretation of the facial display in terms of emotions. There is also the question to what extent the different classes of mental states are differentiable and unambiguously. El Kaliouby and Robinson (2004b, sec. 6.2) found that an automatic system to infer mental states results in misclassification for example between *agreement* and *concentrating* and multiple classifications, for example *unsure* and *thinking*. The agreement on the meaning of labels such as *agreement, concentrating, disagreement, thinking, unsure* and *interested*, might be even more questionable than a set of basic emotions that can be justifiable in evolutionary terms. It can be argued that if other labels were chosen, those facial expressions could be classified differently, and still be as valid interpretations as the firsts.

3.3 Facial Expressions Monitoring in HCI Research

With the advent of multimodal interfaces, multiple channels of communication between the computer and the user as well as between the user and the computer have emerged. The user to computer communication, which relied mostly on the hands as the input mechanism, saw speech being added to the mixture. Today, though the hands are still the most dominant way to control computers, speech control is also present in desktop, pda, cell phones, for example. While the face is a central part of human-human communication, it has not been equally important in human-computer interaction. This is no surprise, since the interaction with the computer is still mostly compared to the use of a tool, and, therefore, in that role, the computer attention to the user face would not be seen as important.

In this section we discuss the previous work concerning the application of facial expressions monitoring in HCI context. Ward, Bell and Marsden (2003) identify different application domains that can be divided between along two axes, first concerning the facial expressions of interest: voluntary versus spontaneous, and second the adaptation of software to the facial expression: adapt versus not adapt, Table 3-4.

TABLE 3-4. Classification of the different applications of facial expressions monitoring in HCI.

	Spontaneous facial reactions	Intentional Facial movement
Software does **not** adapt its behavior	Usability evaluation	Controllable software - virtual puppeteering
Software does adapt its behavior	Adaptive software	Management of collaboration

Source: Ward, Bell and Marsden (2003)

The distinction between spontaneous and intentional facial expressions might seem *a priori* somewhat superfluous, both might occur interchangeably or in sequence. The intent with that distinction is to classify the manner by which facial expressions are employed: if as a conscious input from the user, for the purpose of interface control or communication, or as a more subtile hint on the user state. Our interest focuses on the latter, though intentional facial expressions might emerge as a communication channel between the user and a system, if this is capable of detecting facial expressions.

There is a considerable amount of work on the computer vision aspects of facial expression tracking, which we will summarize in Chapter 5. The same cannot be said on its application which is left for the most part unaddressed in the literature.

When it is addressed, the emphasis is placed on intentional facial movement. Examples can be found in facial character animation (virtual puppeteering) (Pyun, Shin, Kim, and Shin. 2001; Chandrasiri, Naemura, and Harashima 2001), controlling the interface with face gestures as an alternative for differently-able people to interact with the computer, for example the use of voluntary frowning as a counterpart for the mouse button press (Surakka, Marko and Isokoski 2004; Lombardi and Betke 2002), controlling interfaces when the hands are occupied, for example to interact with a car navigation system (Ko, Kim, and Ramakrishna 1999), and using intentional facial expressions to communicate affect in Instant Messaging (El Kaliouby and Robinson 2004a).

Less work has specifically addressed the monitoring of spontaneous facial expressions. Baker discusses a facial expression tracking algorithm to estimate drivers mental state (Baker et al. 2004). The emphasis though is placed on the computer vision system; the classification of the drivers mental state is not discussed. In a similar application, Gu, Ji and Zhu (2002) presents a facial expression monitoring system, for drivers drowsiness detection, but also here the classification of expression in drowsiness/non-drowsiness was not addressed. An instance of work specifically concerned with the identification of an individual’s confusion and interest states is presented by Scheirer, Fernandez and Picard (1999). A pair of eyeglasses with embedded sensors detects the activity of the corrugator muscle (see Figure 3-3).

Hazelett (2003) specifically addressed the topic of usability evaluation in a study where the corrugator muscle activity was recorded while participants interacted with websites with different usability level. The overall level of corrugator activity during all the period the participants interacted with the website was higher for the less usable websites. In a recent study, Hazelett (2006) extended his findings, having participants playing a car racing video game and recording the corrugator and zygomatic muscle. The corrugator muscle EMG was significantly greater during game negative events, such as being passed by other cars, or running off the course. The zygomatic muscle (responsible for smiling) was found to be significant greater during positive event such, overpassing a car, winning.

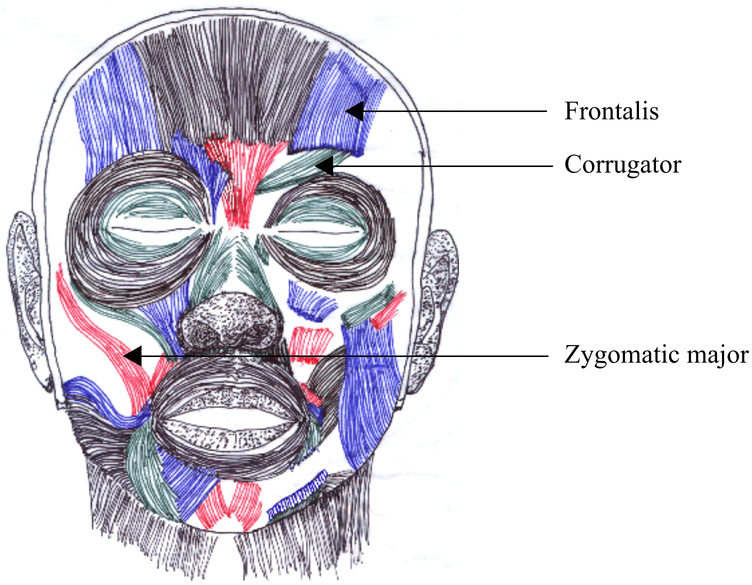


FIGURE 3-3. Three groups of facial muscles often monitored in the emotion psychology research. Corrugator: pulls the inner half of each eyebrow downward; Zygomatic: controls movement of the mouth and produce expressions like smiles; Frontalis: brings the eyebrows upward.

Ward, Bell and Marsden (2003) conducted one of the few studies in HCI measuring users' spontaneous facial expressions employing a video-based facial expression tracker. Participants, filling in a web quiz on driving theory, were exposed to two types of stimuli. The first, a surprise alert, consisted of a standard browser alert box with the corresponding sound that stated "Reaction Time Test. Click OK to Proceed", was considered a strong stimuli. The second consisted of a question with high affective content such as how to deal with accident injuries or bad behavior in other drivers, which was considered a weaker stimuli. A first analysis, performed by observers judging the video recording of the faces, concluded that the increase on facial movements correlated well with an increase on the physiological arousal for the stronger stimuli, the surprise event. With respect to the weaker stimuli, only 1/3 of the arousal responses were accompanied by greater facial movements. A second analysis was performed, using the animation of the virtual face created by the face tracking software. Observers judged the amount of movement of the virtual face, which, in ideal face tracking conditions, should mimic the expression of the participants. For the strong stimulus, there was an agreement for 80% of the participants: the virtual face appeared as not having facial movement in 20% of the participants (false negative). With respect to the weaker stimuli there was agreement in 47% of the participants, with occurrences false negatives and also false positives, where the virtual face appeared increasing the facial movement which didn't correspond to the reality.

3.4 Inferring User's Valence Towards the Interface from Facial Expressions.

Previously in this chapter, we discussed different modalities to infer about the emotional state of the user: physiological and facial expressions. Physiological measures are an effective and well studied indicator of emotional states. Its posing limitations are the intrusiveness of the measures. Facial expressions' emotional meaning has also been a rich topic of research in the psychology field. The interpretation of facial expressions in terms of emotions has been a rich topic of research in the psychology field. In comparison to physiological monitoring, facial expressions offer an opportunity to explore natural channels of communication, with lesser impact in terms of intrusiveness.

A crucial aspect of the interpretation of facial expressions is the context in which the expression is observed. Recalling the different roles of facial expressions discussed in Section 3.2, grouped as conversational or physiological/emotional relevant, naturally the expressions of interest to identify particular users' positive/negative reactions are within the latter group.

Analyzing the types of situations that might trigger facial expressions within HCI, we considered the domains of information to which a facial expression of emotion can relate to (Ekman 1997), and propose three that seem more likely in a HCI context:

- The antecedents, the events which brought about the expression.
- The person's thoughts; plans; expectations; memories.
- The internal physical state of the person showing the expression.

Table 3-5 illustrates occurrences that can occur within a HCI context, for each of those categories.

TABLE 3-5. Elicitors of facial expressions in HCI

Antecedents/event	
	Error occurred
	A particular action has unexpected results
	Finding difficult to complete a task
	Not understanding a particular computer message
	Unexpected occurrence
	An inadvertent action
	Unable to find a particular option
	Unexpected behavior from an application
	Some message/text
	Reading content of emotional nature
Thoughts/plans	
	Unable to plan the steps to execute a task
	Expecting to complete or fail a task
	Recalling a memory related to the task
	Recalling an unrelated memory
Internal physical state	
	stress
	alertness, satisfaction
	tiredness

While most of the situations listed in Table 3-5 do in fact indicate situations of interest to detect the user state, there are some notable exceptions. Reading content that is emotional in nature, for example a bad news communicated through email, or having some unrelated thought that triggers some facial display are not events that are intrinsic to the dynamics of the user-computer interaction. Also left from Table 3-5 are communicative facial expressions, where the user might be interacting with another user.

This ambiguity is important to consider when designing studies to research on users' facial expressions in a HCI environment. While we are interested in monitoring users' spontaneous facial expressions, signaling positive and negative reactions within unconstrained real-world environments, an initial step calls for a more contained approach, where we can monitor users' facial expressions with a reasonable degree of confidence that it relates to the interaction aspects.

The interpretation of spontaneous facial expressions raises also another challenge. Ekman's model of facial expression of emotion proposes a certain number of prototypical expressions, and related family of expressions associated with a certain number of innate basic emotions. Table 3-6 lists the most common prototypical AU combinations and major variants which are associated with emotion.

In a first naive attempt to decode emotions from facial expressions one might attempt to identify the action units involved and identify one of the six emotions that represent the best match. This has been to most extent the approach followed by software advertised as emotion expression recognizers (Fasel and Luettin 2003). Unfortunately, this methodology is not quite right, spontaneous expressions don't necessarily exhibit all the features associated with the prototypical facial expressions. Instead, the precise morphology of the face for those spontaneous facial expressions varies around a core configuration, likely reflecting the degree and control of emotion (Keltner and Ekman 2000). For example, a face of surprise might occur with the brows raised and eyelids open without the jaw open. The list of the AUs for each of the prototypical faces of emotion, can just be used as a reference since there is not a direct translation to facial expressions that occur naturally. Mapping tables, such as Table 3-6, illustrating the combinations of AU associated with prototypical expressions, should, for the reasons above, not be employed to interpret FACS scores.

Also, there is not a complete evidence for all the table entries and its effective relation with emotions (Ekman, Friesen and Hager 2002, chapter 12). An attempt to apply those tables to the interpretation of spontaneous facial expressions of individuals interacting with a computer game, revealed that those patterns of expressions were rarely observed (Kaiser, Wehrle and Schmidt 1998).

TABLE 3-6. Prototypical expressions and AU combinations.

Emotion	Prototypes	Major Variants
Surprise	1+2+5B+26	1+2+5B
	1+2+5B+27	1+2+26
		1+2+27
		5B+26
		5B+27
Fear	1+2+4+5*+20*+25, 26 or 27	1+2+4+5*+L or R20*+25, 26 or 27
	1+2+4+5*+25, 26 or 27	1+2+4+5*
		1+2+5Z, with or without 25, 26, 27
		5*+20* with or without 25, 26, 27
Happiness	6+12*	
	12C/D	
Sadness	1+4+11+15B with or without 54+64	1+4+11 with or without 54+64
	1+4+15* with or without 54+64	1+4+15B with or without 54+64
	6+15* with or without 54+64	1+4+15B+17 with or without 54+64
		11+15B with or without 54+64
		11+17
	25 or 26 might occur with all prototypes or major variants	
Disgust	9	
	9+16+15,26	
	9+17	
	10*	
	10*+16+25,26	
	10+17	
Anger	4+5*+7+10*+22+23+25,26	Any of the prototypes
	4+5*+7+10*+23+25,26	without any one of the
	4+5*+7+23+25,26	following AUs:
	4+5*+7+17+23	4, 5, 7, or 10
	4+5*+7+17+24	
	4+5*+7+23	
	4+5*+7+24	

Adapted from Ekman, Friesen and Hager (2002). *means in combination the AU may be at any level of intensity

A more extensive collection of facial expressions and their psychological interpretations is being collected under a database, the Facial Action Coding System Affect Interpretation Dictionary. In this database, the facial expressions are described with FACS and its interpretation in terms of affective states, for which there is agreement among experts, are added. The cataloguing of the expressions is under ongoing development. To this point not all the emotion expressions are known; though there are about 6000 catalogued different expressions and information on how they relate to emotions, it is not nearly a complete set (FACSAID). Therefore, it doesn't seem feasible to train a computer to recognize all the emotion expressions because there is no complete training set. If the computer program is only able to distinguish between the different prototypes of expressions, from posed videos or pictures like the ones illustrated above, this capability is little useful because in practical situations those prototypical expressions don't occur.

Factoring these elements, we adopted the strategy of monitoring users' valence. Valence, in psychology, is defined as the positive or negative emotional value associated with a stimulus[2]. A classification of facial expressions in terms of positive/negative expressions lifts the burden on the interpretation of spontaneous facial expressions; facial expressions that otherwise would be ambiguous can more easily be at least distinguished according to its valence. Also the cost of that simplification for the purpose of detection adverse events is not substantial. After all, the point is understanding if the user reacted negatively or positively in any given situation.

There are a number of open research question that this approach raises. The studies discussed in Section 3.2.3 present evidence that users do exhibit facial expressions while interacting with the computer. It is though questionable to what degree facial expressions are sensitive to a subtile stimuli that might emerge from usability problems, for instance having difficulty finding a particular element in the interface, performing a step with unintended effect. Will facial expressions be indicative of such problems? Are they related with the task difficulty?

2. Wikipedia, s.v. “valence”, http://en.wikipedia.org/wiki/Valence (accessed 25 August, 2006).

3.5 Discussion

While computers can recognize patterns of physiological activity, identifying what provoked those sets of reactions is a much harder problem. Humans find it often hard to identify the cause of someone else's emotional reactions, computers with further less capabilities to understand language, situations and even common sense, are much more restricted to address the problem. Despite these limitations the recognition of emotional states is still useful. The following example highlights well how the discrimination of emotional states associated with approval or disapproval is useful even if unable to understand what cause the reaction (Picard 1997, 26):

> *Dogs can recognize vocal affect, even though they presumably cannot understand what is being said. If Fido is on the sofa and you yell angrily, 'Get down off the sofa!' he may not only get down, but he will probably acknowledge the emotion physically, with the position of his ears, tall, and head.*

In a similar manner the recognition of the users' positive or negative attitudes by the computer could lead to acknowledgement, eventually learning on what caused that reaction.

The focus on the physiology of emotion and the recognition of the physical displays associated with the each emotional state might seem at first sight to neglect language as a rich mechanism to inform other about mental states. Why not consider language to inform about one's emotions? After all, people rely on the language to transmit their emotions in ways that go beyond patterns of physiological reactions. That technique is commonly applied in usability studies in the form of the think-aloud protocol (Lewis and Reiman 1993), where users verbalize their thought process revealing their understanding and assessment of the application. Outside an usability lab environment, mechanisms for the user to report back their problems are harder to implement, it would require the user to quit their task at hand in order to express themselves. Also, when users have a chance to self report their problems, it is significantly delayed from the time of its occurrence (Hartson and Castillo 1998).

The discussion on users' body language monitoring is not exhausted in their facial expressions. Posture in conjugation with other non-verbal signals, could potentially be used to infer on users' attention, interest, and emotions, for example the rigid posture associated with anger (Knapp and Hall 2006, 9). Beyond non-verbal monitoring, user vocalization might also have correlations with particular emotions or mental states (Knapp and Hall 2006, 10). Naturally an ideal monitoring system would be able to account for those and other modalities, in a similar manner to what humans are capable of.

CHAPTER 4

Creating a Base Line: A Case Study on Users' Facial Expressions

The movements of expression give vividness and energy to our spoken words. They reveal the thoughts and intentions of others more truly than do words, which may be falsified.

— Darwin 1872/1998

4.1 Study Background

In Chapter 2 we discussed the importance of non-verbal communication in HCI, and in the previous chapter we presented different measures from the nervous system associated with particular mental states, namely stress, teariness, cognitive workload, and emotional states. The face, in particular, was discussed as a important channel of non-verbal communication. In comparison to other physiological indexes, namely electrodermal or heart rate measures, spontaneous facial expressions exhibit a low latency in response to stimuli, a fact which is important to assess in-time users' responses to HCI events.

In this chapter, we present a study correlating facial expressions with the occurrence of adverse interaction events as perceived by the user. The goal of this experiment was to establish a correlation between task difficulty and emotional valence and to create a reference baseline on user facial response for subsequent studies. We chose, therefore, to use the well established Electromyogram (EMG) to measure facial muscular activity in a first stage, before exploring a camera-based approach. Also this study extends the previous work associating facial responses with usability assessment, as we expose the users to an open-end task where most of the induced stimuli emerge from the natural adverse events occurring during the interaction. Strong and blunt stimuli have been used in the past to provoke measurable physiological responses, in particular, audible alerts or popup windows have been used in HCI contexts (Ward and Marsden 2004). The choice to employ stimuli emerging from the difficulties users experienced during an HCI task, allows us to assess the validity of the face in identifying particular problematic events in real-life HCI scenario. With respect to the task, we intended it to be as familiar as possible. Since one of the most common activities performed with the computer is word processing, we targeted this particular application also because it offered the opportunity for a more elaborated and diverse set of tasks when compared to other popular activities such as email or web browsing.

4.2 Study Setup

The study was conducted by us at the Motion Analysis Laboratory, Spaulding Rehabilitation Hospital, which conducts studies employing EMG equipment. Twenty one participants were recruited through the Spaulding Hospital volunteering mailing list, 5 male and 16 female, mostly workers at that hospital or from the nearby Massachusetts General Hospital. The participants were required to have a minimum familiarity with Microsoft® Word. At their arrival, the participants were briefed about the generic purpose of the study: "to develop methods of assessing computer user awareness and emotional state in order to improve the efficiency of human-computer interfaces", see "Research Consent Form #1" on Appendix I. The participants were situated at an office desk, a video camera was placed on the back of the desk with a view onto the user's face. Three pairs of EMG sensors were placed symmetrically on the participant's face, recording the activity of three muscle groups: corrugator, frontalis, and zygomatic (Figure 4-1). The relevance of these muscles was previously discussed in "Monitoring Facial Expressions" on page 44.

We chose to perform bilateral recordings (on both sides of the face) to ensure that unilateral expressions would be recorded, see for example Figure 4-1. To allow the participants to become as comfortable as possible to the sensors, they were asked to play the computer game Tetris for about 10 minutes before starting the main task.

FIGURE 4-1. Participant monitored with EMG sensors exhibiting an unilateral expression.

Once the participants were ready, the investigator opened a Microsoft® Word file containing a page of text and requested the participants to format that text according to the printed page that was provided (Figure 4-2 on page 65). The subtasks covered a range of difficulty levels, ranging from basic to more advanced formatting, ensuring this way that participants with different levels of expertise could potentially experience a challenging situation. The participants were further instructed if they could not complete a particular component of a task, to skip it and move on. They worked until they finished all required steps, or felt they had completed as much as they could. Table 4-1lists the subtasks required. Notice that there was no specific list of subtasks provided to the participants to be executed in some predetermined order. Instead it was an open-ended task where each user would follow their own approach. While this made the data analysis more difficult, it was aimed at simulating a realistic task, where problem solving is involved making users engaged. This is an important point since if users were just following a list of tasks, they might not become engaged and therefore less likely to react.

TABLE 4-1. List of actions required to complete the task

select font	edit footer
select size	create table
select style	shade table cells
change text color	insert picture in table cell
insert word art	insert autoshape
format text in columns	rotate autoshape
insert picture	insert symbol
insert picture caption	

In addition to the EMG and the video recording of the participants' faces, the screen was also video-captured while participants worked on this activity. This recording provided a detailed account on the user's interaction and permitted an analysis on the difficulty experienced executing each subtask. In the design of the experiment, among the employable methods to gain insight into the subjects' assessment of the task, the think-aloud protocol was set aside since the subjects' conversation would have greatly affected their non-verbal facial expressions and thus rendered the conducted experiment much more difficult. To ensure the synchronization between all the recordings, the computer that acquired the EMG recording and the participants' computer clock were synchronized through the network time protocol service. A small LCD display connected to the participants computer, and visible in the video recording of the face, served to synchronize the video. At the end of the session we provided participants an opportunity to express their difficulties executing the task in a questionnaire, see "Questionnaire Study #1" on Appendix II.

My Newsletter

August 2, 2002

Welcome to the first issue of **My Newsletter**, a bimonthly publication dedicated to spread *all my news to all my friends*. You will find here all my stories and opinions about the everyday events. I count with your contribution with articles, jokes and other amusement news.

My Vacations in Rhode Island

On summer nights, as the sun rolls over the western horizon, the park simmers with visitors and tourists. It is home to a summer-long concert series, festivals and even a gondola ride. But the single biggest draw to this four-acre urban park is the sounds and sights of *WaterFire*. Below, the river is dotted with braziers crackling with burning wood, throwing off an aromatic smoke and atmosphere that draws visitors back repeatedly.

View from the Park

My Calendar in September

S	M	T	W	T	F	S
1	2 *	3	4	5	6 *	7 *
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

2 → Labor Day
6 → Rosh Hashanah begins at sundown
7 → Rosh Hashanah
30 → Autumn begins

FIGURE 4-2. Formatted page of text - the goal for the participants' task.

4.3 Data Analysis

The purpose of the data analysis was to inspect an eventual correlation between the facial activity observed, signaling users' emotional valence, and the difficulty experienced during segments of the task. The questionnaires the participants answered in the end of the session provided insights in the situations they found the most difficult. Nevertheless we were interested in a much more detailed account of the users' experience. We proceeded by creating a measuring scale and using the screen-captured video to collect information about the perceived degree of difficulty each user experienced for each task. All the subtasks were scored following a 6 point scale:

- (1) Fine: no problem;
- (2) A minor problem, for example failing at a first attempt to locate a menu entry, but immediately resolved;
- (3) Problematic, more time was spent but quickly found a way through;
- (4) Some effort spent; user experienced several missteps;
- (5) Considerable effort executing the subtask, was repeated several times, before completing;
- (6) Failed to complete the subtasks;

A log file for each participant listed all their actions, the time of occurrence and the corresponding score.

The raw EMG signal was recorded at 1000 Hz. Before processing the EMG recording the video of the participants was inspected. The EMG signal segments where participants touched their faces, the sensors, or diverted their attention from the screen, were discarded. This process ensured that eventual spikes, that don't relate to a reaction to the task, were removed from the EMG recording. The remaining of the EMG recording was high-pass filtered with a 20Hz cut-off frequency to eliminate movement artifacts. Movement artifacts are characterized by lower frequencies caused by the motion of the sensor on the skin; though the sensors are attached with stickers the succession of movements and perspiration can cause them to move slightly.

The analysis of the EMG signals focused on detecting the onsets of the muscular activity. The muscle onset or excitability is defined by an increase of the signal amplitude. The root mean square (RMS) is commonly used measure of amplitude, proportional to the power of the signal (Smith 1997, chap. 2).

For a discrete signal, the RMS is defined as:

$$\chi_{RMS} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \chi_i^2}$$

where $\{\chi_i\}$ represents the signal samples

The RMS was computed based on a 30ms window. Determining periods of muscle activity (onset) is a function of many criteria, usually a threshold is set based on some measure of muscle resting activity. The EMG signal must be over this threshold by some arbitrary measure usually 2-3 standard deviations by a certain minimum amount of time (Hodges and Bui 1996). We chose to define muscle onset as an increase of 3 standard deviations, with duration of at least 1second, over the mean. Based on this definition the muscle onset was calculated for each participant for each subtask.

4.4 Results

The main research question being investigated by this study was whether there is a correlation between facial expressions, signaling negative emotional valence, and the occurrence of adverse interaction events. As explained in the previous section, the rating of the users' perceived subtask difficult was subjective, based on the observation of the screen recording. Participants, nevertheless, provided their own assessment of the problematic events after the session. While we didn't expect the participants to recall the degree of difficulty of every subtask, we asked them to indicate the tasks they perceived as most difficulty. We proceeded by comparing the rating we attributed to those subtasks against the rating of the remaining subtasks. If our rating agreed, at least to some degree, with the participant's own assessment, the subtasks indicated by the user would also be the ones we rated higher. To proceed with that analysis we have considered the following null hypothesis:

H0: The average difficulty rating for the subtasks indicated by the participants is less or equal than the average difficulty rating for all the other subtasks.

A paired t-test, on the average rating of the subtasks the participants indicated on the questionnaire versus the average score of remaining tasks, showed significant differences $t = 7.73524$ ($df = 15$), $p < 0.001$. This indicates that the null-hypothesis above can be rejected and, therefore, confirms our expectations regarding our assessment of the subtasks difficulty.

For the analysis of the main research question, whether there is any correlation between muscle activity and task difficulty, we aggregated the data for all the subjects and analyzed the frequency of tasks with and without muscle activity. Table 4-2 on page 69 shows the contingency table for the *corrugator* and *zygomatic* muscles. It should be noted that the *frontalis* muscle group was not factored into these results since recurring problems with the sensors caused the recording to be unusable. The null hypothesis established was the following:

H0: muscle activity and task difficulty are independent.

A chi-square analysis shows significant correlation between task difficulty and corrugator muscle activity (χ (5,2238) = 284.763, $p < 0.001$, Cramer's V =.357) and also between task difficulty and zygomatic muscle activity(χ(5, 2238) = 472.479, $p < 0.001$, Cramer's V =.459). These results lead us to reject H0 and conclude there is a relationship between the two variables.

To examine the direction of the relationship, a column-equalized Goodman-Kruskal lambda was calculated considering the task difficulty as the independent variable (Goodman and Kruskal, 1979). This analysis reveals a moderate relationship between corrugator activity and task difficulty ($\lambda = 0.133$), and a moderately strong relationship between zygomatic activity and task difficulty ($\lambda = 0.230$).

TABLE 4-2. Contingency table for muscle activity (A) versus no muscle activity (~A) for each task difficulty level.

Corrugator						
	1	**2**	**3**	**4**	**5**	**6**
A	168 [338.0]	110 [100.2]	58 [45.6]	76 [44.3]	122 [63.0]	185 [127.9]
~A	884 [714.0]	202 [211.8]	84 [96.4]	62 [93.7]	74 [133.0]	213 [210.1]
Zygomatic						
	1	**2**	**3**	**4**	**5**	**6**
A	244 [486.5]	174 [144.3]	82 [65.7]	109 [63.8]	162 [90.6]	264 [184.1]
~A	808 [565.5]	138 [167.7]	60 [76.3]	29 [74.2]	34 [105.4]	134 [213.9]

At first sight, this association of task difficulty with zygomatic activity seems to contradict the fact that this muscle is associated with positive stimuli (smile Proclivity). Nevertheless, this phenomenon has been previously observed multiple times in computer-game settings where increased zygomatic activity was reported in negative or obstructive events (van Reekum 2000, sec. 2.2.2).

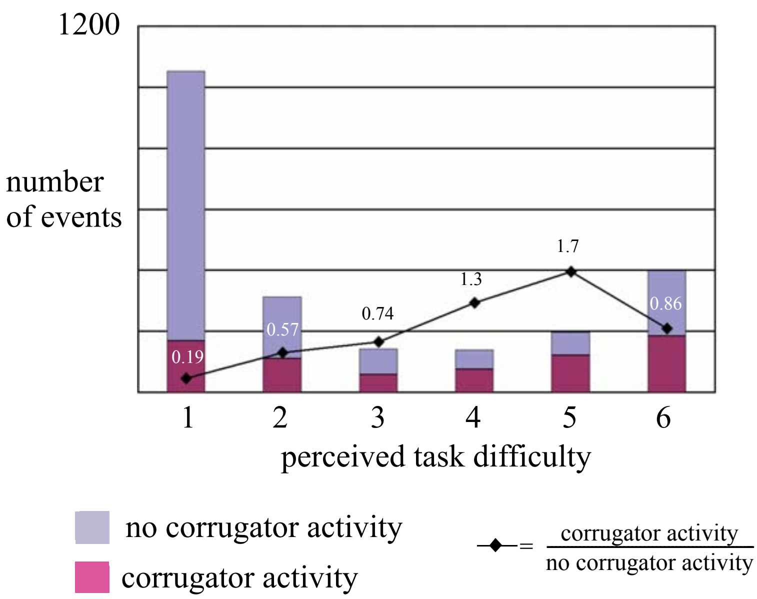


FIGURE 4-3. portion of tasks with associated corrugator muscle activity.

Figure 4-3 and Figure 4-4 show the distribution of tasks with and without associated muscle activity, as well as their ratio, for each task difficulty level.

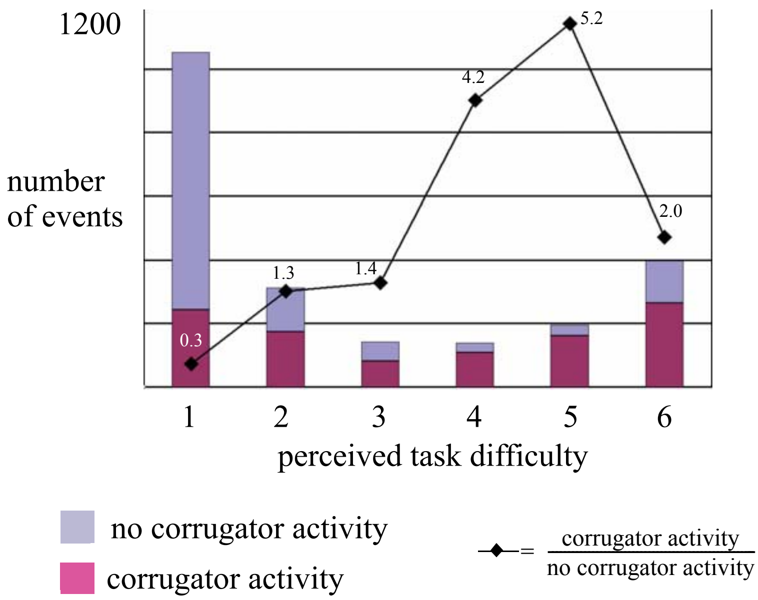


FIGURE 4-4. Proportion of tasks with associated zygomatic muscle activity.

It is evident from the graphs that, in general, the proportion of tasks with muscle activity increases with the increase on the task difficulty. Notably, for tasks scored as (6), there's a decrease in the muscle activity. Though the scientific study and evaluation this phenomenon requires further investigation, we have anecdotal evidence that this behavior emphasizes the subjects' resignation at that point, as those events correspond to tasks the subjects were not able to complete and consequently gave up on.

4.5 Discussion

The described study suggests that an increase in the difficulty of the user's experience with an interface is overall associated with an increase in particular facial muscles activity. While the results of this study are not surprising and confirm what has been shown in HCI studies (Hazelett 2003; Ward, Bell and Marsden 2003), it is the first time that the analysis has been done in level of resolution that considers all the users actions in a complex task. With this approach we aimed at demonstrating that, in general, spontaneous facial reactions occur naturally during the course of the interaction reflecting users' perceived difficulty. Further, it provides a user-behavioral baseline for experimenting with other approaches to capture users' facial responses in HCI.

One of the difficulties analyzing the data is the assessment of the users' difficulty performing the task. As explained previously, the assessment was done per each subtask based on the screen recording of the user actions, and was in part validated by the fact that the subtasks judged most difficult by the users where also the subtasks rated most difficult by our scale. Eventually other techniques could have been followed, for example using the time to complete the subtask to assess its difficulty; this was not straightforward though, as participants often paused and glanced to the instructions for directions while executing the subtask and therefore the duration to complete the task was potentially biased. An alternative approach would have been for the participants themselves, after the session, observe the recordings of their own actions and rate their difficulty in each subtask. This would introduce though an variable component since each individual could apply a different set of rules by which to assess their own actions. Also, often such process requires playing back several times the recording and would have been an unpractical time consuming process.

It is also interesting to notice that some of the facial expressions observed in response to negative occurrences were smiles. It is revealing that even in the relative controlled environment we can already witness the complexity of the facial responses. This same phenomenon was evidenced by Kaiser and Wehrle (2001) in a computer game setting. They observed smiles and smile controls (a smile that is inhibited or obscured) which occurred more frequently in negative situations that were perceived as self-caused as compared to other-caused negative situations. They argued that the function of the smile in those cases serves as a coping strategy to handle negative situations. It is, therefore, an important point to keep in mind since the interpretation of smile as a result of a positive event might be misleading.

EMG sensors posed several restrictions that limit its applicability beyond an experimental setup: it requires the physical contact with the user, needs expert input to find the optimal location for the sensors to be applied over the muscles, requires cleaning the skin and preparing the sensors for better recordings. Its advantages is the relative simplicity by which the facial activity can be measured, and the ability to detect minute changes on the muscles responses, that might not be visible. Nevertheless, the applicability of the results discussed in this chapter are dependent on the ability to carry facial expressions monitoring through less intrusive technology. Video-based monitoring presents itself as the only other alternative, with the major disadvantage being the robustness the computer vision algorithms. Nevertheless, the increasing presence of web cameras in desk-top-computing environments and the emergence of commercial software capable of facial expression analysis make this the most promising approach. In the next chapter this topic will be discussed in depth.

CHAPTER 5

FACEit - A System for Facial Expression Analysis

Dave, although you took very thorough precautions in the pod against my hearing you, I could see your lips move.

– HAL 9000 "*2001: A Space Odyssey"*

In 1968, at the time Arthur C. Clark envision HAL 9000 in its memorable novel *2001: A Space Odyssey*, it seemed plausible that computers forty years later would be capable of a number of complex social functions. By 2001, computers revealed much more modest in those skills; not only they are not able to interpret art, but they are also lag far behind in respect to facial expression analysis or interpreting emotions, capabilities fully exercised by HAL. Even basic skills like discerning if someone is sitting in front of them, is mostly missed by current computers.

Before proceeding, we summarize the argument presented so far. We have hypothesized that the classification of the facial expressions valence is an appropriate approach for classifying users' perceived problems. The study in the previous chapter discussed a study that examines the correlation of facial expression activity and task difficulty. Electromyogram (EMG) sensors presented the most accessible method to measure facial expressions, and while it is a solution appropriate for a study where users are bounded to a laboratory environment, it is not practical to extend the results to a generic environment. Video-based monitoring of the face is, in contrast, less intrusive with potential to be more widely applicable.

This chapter focuses on computer vision approaches to monitor users' facial expressions. We describe the requirements, design and implementation of the image-based facial expression monitoring system - FACEit. We present the motivation behind its implementation, followed by a discussion on the requirements of a system that successfully monitors spontaneous facial expressions. The chapter follows with an overview of the numerous computer vision methods available and proceed in more detail on the methods FACEit implements.

5.1 Motivation

The low cost and good quality of computer video cameras and an increase on the popularity of video chat applications such as MSN Messenger[1] and Skype[2], have fueled the growth of computer cameras to millions of users worldwide. Some computers, for example Apple's iMac, are equipped with an integrated video camera. The current application of computer video cameras is exclusively for chatting, but the increase on the cameras and computer hardware sophistication allows for complex image processing, and computer vision applications. The recent Logitech web cameras support face tracking to keep the user in frame, and for an entertaining video chat, the user face is replaced by a character that mimics the user pose and face movements (Logitech 2006). As is the case with numerous technologies, the computer game industry is pioneering the integration of video cameras in new forms of game interaction. The eye toy from Sony Playstation® (Sony 2005), and a soon to be released version from Xbox® enable the user to interact through movements and gestures in video games. Specifically in usability monitoring, a few exhibitors in CHI'06, the conference on human factors in computing systems, displayed products incorporating video based monitoring: SeeingMachines presented the faceLAB system for real-time tracking of head pose, gaze direction and eyes' behavior (SeeingMachines 2006), Noldus[3] offered a sneak preview of an ongoing software development for facial expressions monitoring. Those systems are at this point targeted for expensive usability lab set-

1. http://get.live.com/messenger
2. http://www.skype.com
3. http://www.noldus.com

ups, and might require special camera hardware; but it is a question of time that those developments, supported by increasingly complex vision algorithms and inexpensive hardware, will be incorporated in common computing devices.

Research on video-based facial expression analysis has resulted in a significant body of knowledge, but the availability of ready-to-use solutions is still quite limited. At the time this research has been conducted, [4] provided the only commercial system available for video-based facial expression analysis. The company has just been recently acquired by Google. Other systems are been developed by researchers working on computer-vision algorithms for facial expression tracking (e.g. Tian, Kanade and Cohn 2001) but are not yet robust enough to be deployed for generic use. The development of a whole facial expression analysis system is a multidisciplinary effort, and, as will become evident in the next sections, quite a complex task. The existence of a shared platform open to the research community would potentially catalyze further developments and eventually provide an open and ready-to-use platform.

5.2 The Ideal Facial Expression Analysis System

Image-based facial expression analysis refers to the process of detecting and measuring the visible changes caused by facial muscular activity. Throughout the research it proved useful to identify and consider the characteristics of an ideal facial expression monitoring system to serve as a reference for existing and future systems. Pantic and Rothkrantz (2000) identified a list of properties of an ideal system for facial expression analysis. They address three different levels of requirements, corresponding to different applications: general, behavioral science research application and multimodal HCI Applications. We enumerate them bellow focusing on multimodal HCI application:

- **Individual independency**: The human face shares common traits among individuals, but there are a diverse range of appearances dictated by the age, ethnicity, eyewear, facial hair, grown-together eyebrows, etc. An ideal facial

4. http://www.nevenvision.com/

expression monitoring system should work independently of those individual characteristics.

- **Robust to light variations**: Similar to the human visual system, that adapts from day light, indoor light, to dark environments, an automatic facial expression analyzer should be capable of working under a variety of illumination conditions.
- **Robustness to occlusion**: In real-life situations, very often the face is observed in less than ideal conditions. The head pose, or a hand for example, can partially obstruct the view. Within certain bounds, those situations should not cause a system to fail; naturally if the head pose is so extreme that obstructs the face view, or if an object covers most of the face, it is impossible to recover the expression.
- **Ability to recover from lost tracking**: Even the perfect system would have moments where the face tracking is lost, for example when a face goes out of the camera view, its view is obstructed, or due to limitations of the video capturing hardware such as light saturation. The system should be able to return to normal operation as soon as conditions render appropriate.
- **No markers/make-up required**: As will be discussed in the next section, some systems require the use of markers attached on the face. Make-up applied to the skin is another technique to improve the system robustness to light reflections over the face. These approaches are obviously not applicable for a generic system use.
- **Robustness to rigid movements**: In a real-life, individual often exhibit facial expressions as their pose changes, for example laughing and leaning the head back. An ideal system should be able to keep tracking the face despite this rigid head movements.
- **Robustness to fast movements**: Video capturing hardware poses a limit on how fast an image frame is acquired. Typical computer cameras capture video at a rate of 15 frames per second (fps), while new models might operate at up to 30 fps. The constant rate of image acquisition means that the faster the head or facial expression occur, the larger is the displacements between consecutive

video frames. The system should be able to follow large displacement of the face and the facial features.

- **Fully automatic**: No manual intervention should be required to locate in the initial position of the face, or facial features, or correct the position of the tracking points. The system should detect and locate a face in the image and proceed with the facial analysis processing without manual intervention.
- **No calibration**: No need for special procedures to be carried in the beginning of the video acquisition, or after detecting the face, for example to be in a frontal pose, or the need to adjust initial
- **Distinguishes all possible expressions**: All possible face movement should be detected and categorized. This includes for instance facial movements that are only unilateral.
- **Recognize microexpressions:** A microexpression is a brief facial expression that lasts no more than a 1/25th second. Because they are hard to be consciously controlled these microexpressions leak emotions that individuals try to conceal or that they are not even aware of (Ekman 2003). The ability to detect microexpressions is dependent on the speed of the video capturing hardware and the sensibility of the system to detect minute facial movements.

The next set of desirable properties are dependent on the application pursuit. Classifying the users' expressions in terms of generic behavioral classes, will not necessarily require the level of detailed description that the Facial Action Coding System (FACS, discussed in Section 3.2) provides. Instead the following requirement can be considered as being sufficient:

- **Allows different number and categories to be identified**: The translation of facial expressions in categories should be configured to match the information one is trying to analyze from the expression. For instance, certain applications might be interested in acquiring just the valence of the expression, others might want to assign a mental state.
- **Assign quantified interpretation labels:** The intensity of the facial expressions is individual dependent, nevertheless a measure of the intensity of the expression is potentially useful.

- **Allows blended expression categories:** The face is subtle and often ambiguous. Therefore, it would be to simplistic to assume a particular expression would at any given point fall definitely in one interpretation category. When that is the case the system should classify expressions as falling within more than one category.
- **Real-time processing**: Facial expression analysis output within the time is required for the next frame to be acquired.

None of the systems reviewed in the literature by Pantic and Rothkrantz (2000) satisfied all the requirements, and despite more recent developments, the complete list of characteristics of an ideal system is still beyond the reach of a current computer vision solutions. Partially views of the face, sensibility to subtle expressions, robustness to fast movements, and different illumination conditions are still present challenges. Numerous implementations work under restricted conditions, for example requiring a manual intervention, a stable frontal view of the face (Littlewort et al. 2002), or tracking posed expressions. Those systems, therefore, become seriously impaired in a setting that requires facial expression analysis within unconstrained and spontaneous situations.

5.3 Review on Video Based Facial Expression Analysis Methods

There are three major steps involved in the process: locate the face in the image, model the configuration of the facial features/face morphology and interpret or describe the facial configuration. Figure 5-1 illustrates each step, highlighting the major methodological approaches. There is a substantial amount of work in the computer vision field addressing each of those processes. The next sections discuss each of those processes organized around the methodological approach.

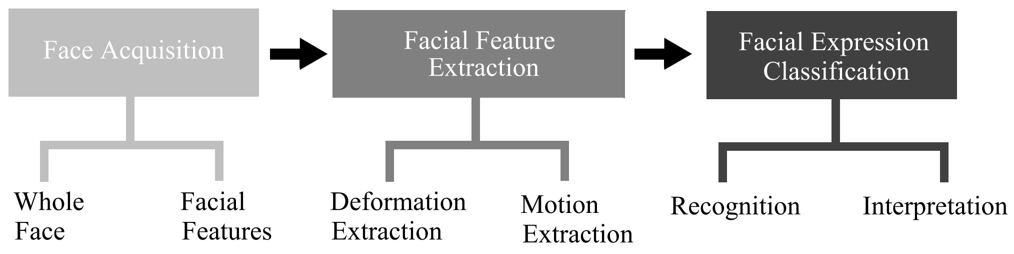


FIGURE 5-1. Generic facial expression analysis framework. Adapted from (Fasel and Luettin 2003).

5.3.1 Face Acquisition

Face Acquisition refers to the process of locating the face in a scene. The face location indicates a region of interest where the subsequent processes of facial expression analysis will concentrate. The face is located on a scene either by detecting it on its entirety, or locating particular facial features and infer the face location from it. In either case, automatic detectors are essential when considering that there should not be a manual intervention to locate the face in the video frame. We highlight here some references to the previous work on this topic, a more in depth discussion on the techniques is discussed by Hjelmas (2001).

Morimoto presents a system that relies on the physiological properties of the eyes to locate the pupils (Morimoto et al. 2000), which can be then used to infer the face location in the video imagery (Davis and Vaks 2001). Morimoto's approach relies on the red-eye effect that is often observed in flash photography. An infrared light source is placed on-axis with the camera (aligned to the camera lens), illuminating the face. The retina reflects the light back to the camera producing a bright area on the camera image. Triggering the on-axis light at every other frame, the pupils alternate between bright and dark spot respectively. Simple image differencing reveals its location. This approach while robust, presents some difficulties with reflections from eye glasses, specular reflections from objects and fluorescent light sources.

A different approach to localize the face, based on a color tracking algorithm, is discussed by Bradski (1998). The algorithm operates on a probability distribution of the skin color. The image is converted to the HSV color model that separates the hue (color) from saturation and brightness. The flesh color occupies a segment of the hue range, independently of the person race; a darker skin individual simply has greater flesh color saturation than a light-skinned individual. A color-based approach for face tracking tends to work well under controlled light environments, but the illumination from different types of light sources (day light, fluorescent or incandescent) or excessive or dim light will degrade the performance of this approach.

Face tracking can be seen as a particular instance of visual object tracking, and within this context there are number of solutions that borrow from machine learning algorithms. Viola and Jones (2001) introduced an algorithm that selects important visual features of the object from a larger database of the object images and creates efficient classifiers. A *cascade* of classifiers allows increasingly complex selection of the regions in the image that are promising candidates to be the target object. This algorithm yield performances comparable to the best face detection systems and it runs at 15 frames per second. An improvement on this algorithm by Lienhart and Maydt (2002) is applied in the FACEit system for face detection.

The output of the face detection process provides a region of interest that will be subsequently processed along the pipeline. Therefore, a robust face detector is a crucial part of the system.

5.3.2 Facial Feature Extraction

After locating the face, the next step is to process the facial image and focus on the facial features of areas that are prone to change with facial expressions. Most facial expression analysis systems address the deformation of intransient facial features, that is the features that are always present in the face, for example, eyelids, eyebrows, mouth. Permanent furrows and transient features, such as wrinkles and bulges are not considered, since its presence and morphology is person dependent and, therefore, difficult to model. Exceptions are the discussions by Lien (1998, chapter 5) and Barlett et. al (1996, sec. 4).

5.3.2.1 Motion extraction

Motion-based approaches compute the movement of the facial features from a sequence of image frames. Dense optical flow is one of the methods to estimate the motion field over the whole face or particular face areas. The motion vectors are computed at each pixel or at every block of pixels between consecutive frames. Mase (1991) applied region-based dense optical flow to estimate the movement of 12 of the totally 44 facial muscles. Each muscle region was manually selected as well as an axis along which each muscle expands and contracts. Dense optical flow motion was quantified into eight directions and allowed for a coarse estimation of muscle activity. Since there is no distinction between the motion caused by the facial expression from motion resulting from head movement, that approach is not robustness to rigid head movement, one of the above mentioned requirements.

Motion-based approaches or optical flow approaches can be improved by the use of motion models. Motion models introduce constrains in the motion field, increasing its stability, and allowing both the head movements and facial expression movement to be accounted for. It also increases the stability, since the model restricts arbitrary motions or discontinuities. Black and Yacoob (1995) proposes a collection of local parametric models for the face, mouth, eyebrows, and eyes. The image motion of those regions are modeled using image flow models with few parameters. Essa (1995) couples an optical flow approach with a physical model describing skin and muscle effects. The analysis was not real time, the most time consuming part was the flow computation taking about 60 seconds per frame for 380 by 450 images (Essa 1995, 83).

In general, dense optical flow methods require high processing times; textureless parts of the face, and moving illumination reflections provoke noise in the motion field. Thresholds on the magnitude of the flow can be imposed to reduce the noise, but it decreases simultaneous the sensibility to subtle facial expressions.

Feature point tracking estimates motion from a set of features along the high contrast areas of the face like the eyebrows or lips. Those features are followed at every frame, the relative motion in respect to each other or to a neutral pose is used to infer about expressions. This method relies on the initial selection of the facial features to track. Lien (1998) discusses the use of feature point tracking for coding FACS. Feature points manually selected on the first frame were automatically tracked across the video sequence. As previously discussed, a manual selection in the initial frame is not a desirable approach in an automated system. Also a limitation of feature point tracking is its sensibility to illumination changes and specular reflections, since the tracking of the features across frames relies on some constant luminance property, like color distribution or contrast.

To circumvent those limitations, the natural feature points of the face might be replaced by artificial dot markers, attached to predetermined locations on the face where the muscles interact, highlighting the muscles actions. Emphasizing the dot markers through distinctive color, or reflective infrared light material, will increase the robustness of the tracking. The markers are placed in the regions of the face where muscles interact, highlighting the muscles actions. This approach was followed by Himer et al. (1991) and Kaise and Wherle (1992). That technique has been applied in the Polar Express movie to facially animate a character based on a real actor’s facial expressions. Applying dot markers to the face, though arguably less intrusive that EMG sensors, it is not a practical approach.

5.3.2.2 Deformation extraction

So far we discussed motion-based approaches to extract the visible changes induced by the facial expression. Deformation-based approach focus on retrieving the shape and/or texture changes that are caused by the deformation of the facial features. Under this category we highlight two methods that have been applied to the problem of facial expression modeling: Active Shape Models (ASM) and Active Appearance Models (AAM). The ASM consists of a statistical model of the shape variations of objects. The model is composed by a set of vertices, defining a contour (shape), and information about the image appearance around each vertex, such as the gradient. The AAM is a generalization of the ASM, instead of using only the information near the vertices of the shape, it uses all image region covered by the modeled object. An AAM consists of a statistical model of the shape and grey-level appearance of the object of interest. In contrast to the motion extraction approaches, these methods can be applied to a single image or to a video sequence by processing each frame independently.

The aim of these methods is to synthesize the range of the object deformations, and appearance, in the case of the AAM, and through an iterative process fit the model to an example of the object in a new image. FACEit implements an AAM of the face to extract facial expression information. Since the AAM uses the full image information, the procedure is more robust than ASM search (Edwards, Taylor and Cootes 1998). An in depth discussion on both ASM and AAM is presented by Cootes and Taylor (2004).

An AAM can be used to retrieve a facial expression by finding the best parameters that best match the image. That process of fitting an AAM to the image consists of minimizing the error between the input image and the closest model instance, a nonlinear optimization problem. Mathematically, minimizing the following expression:

$$(I - M(p, \lambda))$$

Where I is the input image, M the image modelled by the AAM.

This is a high dimensional search problem, where different numerical approaches can be applied, in general requiring significant computational time. An extension of the basic AAM algorithm by Baker and Matthews (2004) allows the AAM fitting process to run in real-time.

A disadvantage with these methods is the effort required to create a good model that can account for the variations in expression and pose. We describe the process in Section 5.4.2.

5.3.3 Facial Expression Classification

The recognition of the facial expression is the last step in the facial expression analysis process. Two methodological approaches can be distinguished to describe the patterns of facial movement: a *judgement* based and a *sign vehicle* based approach (Ekman 1982). Judgement based approaches interpret the expression according to its meaning, typically emotional states. The sign vehicle approach describes facial expressions in terms of the pattern of movements that compose it, without attempt for interpretation. It is an preferred method when the measurements of the individual facial actions is of prime importance, for example for coding FACS in behavioral research, or performing facial animation.

Facial expression analysis systems, influenced by Ekman's proposed model of six basic universal emotions, typically interpret expressions in emotional categories: happiness, sadness, surprise, disgust, anger, and fear. Often those systems are inaccurately denominated as emotion recognizers; the recognition of facial expression deals purely with the visual information without regard for context or other influences, unless there is other parallel sign of the individual experience the emotion, the classification of the expression in one of those categories does not measure the individual emotion. Emotions don't necessarily occur with a facial expression, and the occurrence of a facial expression, even if similar to an expression of emotion, does not signify that the individual is indeed experiencing that emotion, see discussion by Ekman, Friesen and Hager (2002). The interpretation of facial expressions in those categories consists in the distinction of a number of prototypical expressions. Prototypical expressions corresponding to those basic emotions are in fact rare occurrences when observing spontaneous facial expressions (Kaiser, Wehrle and Schmidt 1998; see discussion in Section 3.4), often facial expression classifiers are trained with databases of individuals posing a particular expression, which is substantially different that recognizing spontaneous expressions, that are composed of brief and subtle movement.

5.4 FACEit System Design and Implementation

FACEit is a facial expression analysis tool developed to support the research on the study of facial expressions in a HCI environment. It was not the goal from the start on its development to devise new methods or improve existing facial expression monitoring algorithms. It was considered to be too ambitious in the context of this work. After all this topic, just by itself, can constitute one or more dissertations on its own. In contrast, the effort of constructing a system was thought to be useful for several other reasons. First and foremost would provide us with access to a system that could be tested in the proposed scenario. Existing commercial solutions are not targeted to the classification of user behavior or emotional states and are quite expensive. Research systems focus on the implementation and evaluation of computer vision algorithms, without the concern for a ready to use solution. The implementation of FACEit would allow for a better analysis of the feasibility of our approach: monitoring the valence of facial expressions for the identification of critical incidents; eventually difficulties that were initially overseen would be highlighted. It would also constitute a research tool through which newer improved algorithms, and different approaches to the problem of facial expression analysis, could in future be integrated.

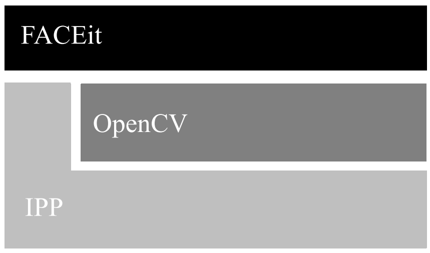


FIGURE 5-2. FACEit System layers.

FACEit was developed in C and is implemented on top of the Intel Open Source Computer Vision Library (OpenCV) 5.0 and Intel Integrated Performance Primitives 4.1 (IPP). OpenCV is an opensource library offering a range of computer vision algorithms aimed at real-time video processing (Intel 2006a). IPP is a commercial library from Intel (Intel 2006b). Both of these libraries are royalty free and can be distributed with the application. IPP provides highly optimized software functions for a variety of data processing functions (IPP). The image processing and matrix algebra are the two library components used by us. IPP algorithms are optimized for a range of Intel CPUs enabling a substantial speedup of the applications. The OpenCV library dispatches some of its functions to the corresponding IPP implementation for faster processing. In the most time critical operations FACEit invokes directly the IPP image processing functions.

FACEit is composed of the three necessary components previously discussed: *Face Detection*, *Facial Features Alignment* and *Facial Expression Analysis*. The next diagram presents the output of each of the FACEit modules.

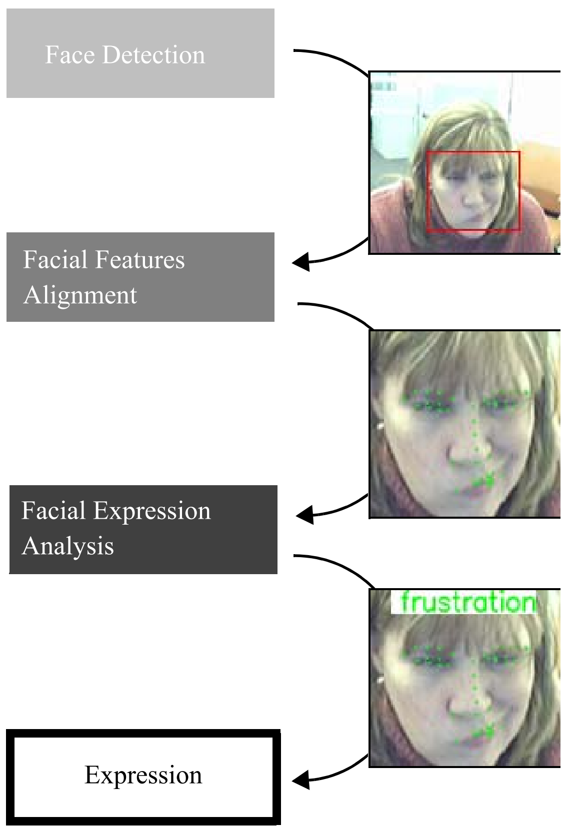


FIGURE 5-3. Output of each of the FACEit modules.

A detailed discussion follows on the implementation of each module described above, with particular emphasis on the face features alignment algorithm, which is a core component of the system. Figure 5-8 on page 94 presents an overall execution diagram.

5.4.1 Face Detection

The Face Detection is implemented using the object recognition algorithm in the OpenCV library. The algorithm has been initially proposed by Viola and Jones

(2001) and improved by Lienhart and Maydt (2002). In a training stage simple visual features are selected from a larger pool which provided the best object discrimination to build a classifier. At each level of the training stage increasingly complex classifiers are combined (“cascade”) to yield a more complex a robust classifier increasing the discrimination of the object of interest from a background. The classifier is trained with a few hundreds of sample views of a particular object, so called positive examples, that are scaled to the same size, for example 20x20 pixels, and negative examples - arbitrary images of the same size. A classifier trained for faces is provided in OpenCV. We used the classifier trained with frontal face views.

The face detection finds rectangular regions in the given image that are likely to contain frontal view faces and returns those regions as a sequence of rectangles. The function scans the image several times at different scales. Each time it finds overlapping regions in the image it applies the classifiers to those regions. After it has proceeded and collected the candidate rectangles, it groups them and returns a sequence of average rectangles for each large enough group. Only the largest rectangle is considered, corresponding to the face nearest to the screen. The face detector can process a 384 by 288 pixel image in about 0.067 seconds, so running at approximately 15 frames per second (Viola and Jones 2001).

5.4.2 Facial Features Alignment

At the heart of the FACEit system is a fast Active Appearance Model (AAM) algorithm (Matthews and Baker 2004). The AAM algorithm belongs to a group of statistical methods for computer vision, where an object is represented by a base shape s_0 corresponding to a set of vertices, a base appearance A_0 corresponding to the pixels intensities, and a set of parameters that modify the shape and appearance according to a model defined at a training stage. Independent AAMs model shape and appearance separate, allowing a linear shape and appearance variation:

$$s = s_0 + \sum_{i=1}^{n} p_i \cdot s_i \qquad A = A_0 + \sum_{i=1}^{m} \lambda_i A_i$$

Changing the parameters p_i and λ_i in the equation above yields different configurations of face shape and appearance. The shape vectors s_i model the changes of the face shape caused by posture and expression, while the appearance vectors A_i models the appearance changes caused by illumination, pose and identity.

Given as input the location of the face in an image, from the face detection module, the goal of the AAM algorithm is to search for the pose and expression. The search of those parameters is performed through an efficient gradient descendent algorithm, inverse compositional image alignment algorithm (Baker and Matthews 2004), allowing the process to run in real-time.

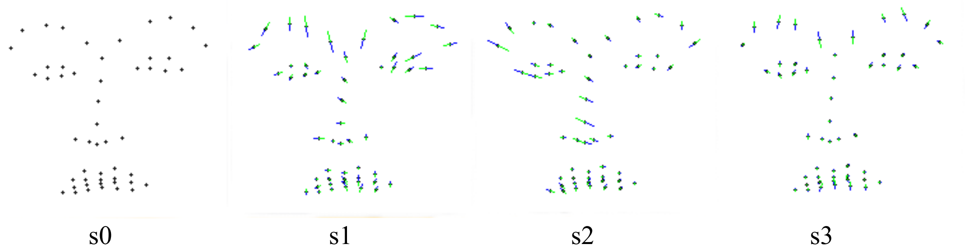
5.4.2.1 Training a face AAM

The base shape s_0 and shape vectors s_i as well as the base appearance A_0 and appearance vectors A_i of the face AAM are defined at a training stage. The process consists in computing the shape and appearance average, as well as the major variations, over a set of training face images. Landmark points are placed over the features that are easily identified, consistently across different examples of the face, Figure 5-4.

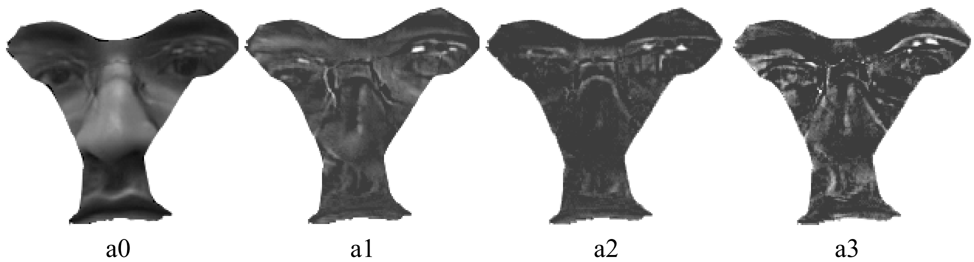
All the images are consistently annotated with the same number of vertices and the vertices have to correspond to same facial features over all the images. The vertices are placed over the edges and corners of the facial features such as eyebrows and lips, where it is easier to observe correspondence between images. The more comprehensive the images in the training data are in terms of identity, pose, expression and illumination, the better the AAM will be able to describe a wider variety of faces, for example the set of images in Figure 5-4 could be extended with images from different individuals. Theoretically, a large enough training set could account for all the shape and appearance variations. In practice, however, the larger the model the more instable the algorithm behaves (Gross, Matthews and Baker 2005). In the various face AAM we trained, we limited the training dataset to the images of the specific user to monitor the expressions. This obviously reduced the generalization of each trained AAM, but on the other hand limited the effort of hand labeling the images to tens of images, rather than hundreds or thousands. Since we intended the AAM to model facial expressions and be robust to different poses, the images in the training dataset included faces under different poses and expression.

FIGURE 5-4. Hand labelled training images.

The vertices for all the images in the training set are aligned (Cootes and Taylor 2004, sec. 4.2). After that process, the difference between the set of vertices in each image is due to expression or pose. Principal component analysis (PCA) is applied to the aligned vertices to compute the components of shape that account for the most variation in the dataset, resulting in a mean shape (s0) and shape variations (s1,...,sn), Figure 5-5a. A similar process is performed for the appearance, for each image in the training set, the pixels inside the shape defined by the hand labelled vertices, are warped back to the base shape and PCA is applied to the collection of those images to calculate the mean appearance (a0) and appearance variations (a0,...,an), Figure 5-5b.



(a) base shape and shape vectors overlaid on top of the base shape



(b) base appearance and variations

FIGURE 5-5. Linear shape and appearance model of an independent face AAM.

This process allows an AAM to describe new instances of the face as a sum of a base shape and base appearance plus a linear combination of shape and appearance variations.

The process described is performed once in a training stage. The real-time process of fitting the AAM to the image is described next.

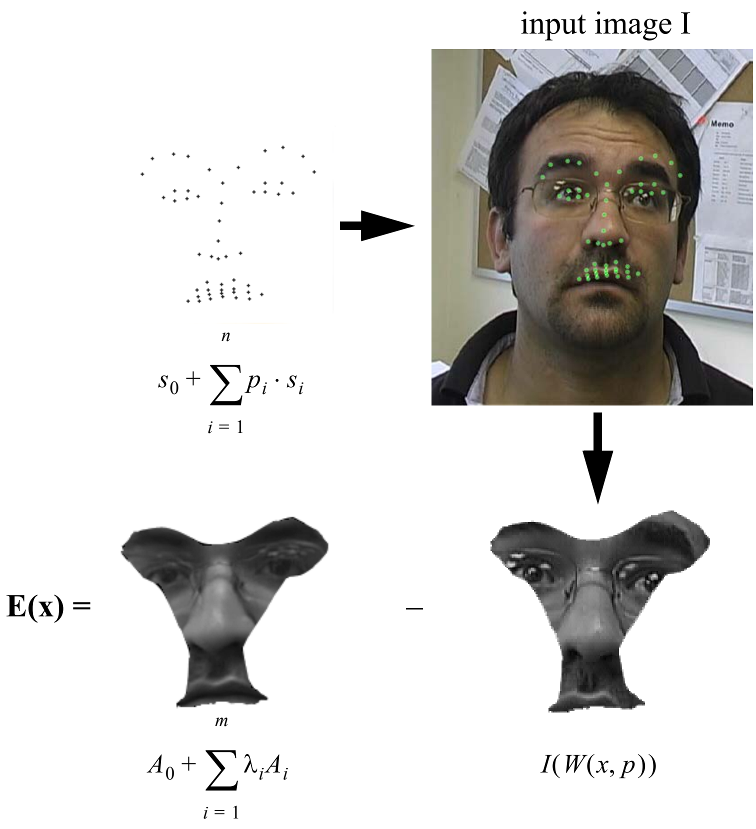
5.4.2.2 Fitting an AAM model

We describe now how the AAM is used to describe the face shape and appearance in a video frame. Given as input the location of the face in an image, from the face detection module, the goal of the AAM algorithm for fitting the model is to find the best linear combination of shapes $\{s_0, \ldots, s_n\}$ and appearances $\{A_0, \ldots, A_m\}$ that resembles the input image. In other words, the goal is to minimize the pixels intensity difference between the input image and a linear combination of {s0,..., sn} and {A0,…,Am}.

The difference is calculated in respect to the base shape s0, therefore the need for a warping (W) from the input image I pixels inside s, back to s0. Mathematically we want to minimize the following expression:

$$\sum_{x \in s_0} \left[A(x)_0 + \sum_{i=1}^{m} \lambda_i \cdot A_i - I(W(x, p)) \right]^2$$

Different gradient descent algorithms are available to solve that expression, we implemented the method by Matthews and Baker (2004) for an independent AAM. The tracking at each frame is considered successful when the difference between the input image and the modeled image falls bellow a given threshold. Since the AAM does not perfectly recreate all possible face shape and face appearance variations, there is no guarantee that the algorithm will converge to the right solution. The AAM model might not be capable of approximating a certain input image - lack of representation power, or might fall in a local minimum.



W(x,p) warps the mesh defined by the vertices in shape s back to the base shape s_0

FIGURE 5-6. Fitting AAM.

As long as the AAM fitting algorithm converges within a certain number of steps, defined by an error threshold, FACEit continues fitting the AAM to the new frames**.** When there is a failure to converge, it resets back to the face location mode, Figure 5-8 on page 94.

5.4.3 Facial Expression Analysis

The set of parameters p_i, computed by the AAM fitting algorithm previously described, uniquely describes a facial expression. FACEit interprets those parameters as a description of the facial expression in terms of neutral, positive and negative expressions, relating to the positive of negative experiences of the user. In order for the AAM to model those facial expressions, it is required to collect a images representing each of those expressions. Since each AAM is user specific, it would be too cumbersome to ask each user to track to portrait a set of those expressions. Instead, we assembled a training set from face photographs collected from the FACS (Ekman and Friesen 1977). Those photographs portray one individual exhibiting prototypical facial expressions corresponding to the activation of each facial muscle, or combination of facial muscles. FACS is by no means a range of all the expressions that are observable in spontaneous settings. The face is estimated to be able to exhibit tens of thousands of different expressions, even discounting the methodological difficulties in collecting such a range of expressions to train an AAM, the model would be too big to be numerically stable. Again here the approach was to reduce the training set and include only a small subset encompassing typical expressions such as frown, dimples and smiles. The selection on which expressions to include was based on observations from facial expressions recorded in previous studies during the interaction with a computer.

The shape vectors computed from the AAM trained from FACS images, are combined with the appearance vectors computed from an AAM trained for a specific user. This way, an AAM is created capable of representing the shape variations corresponding to FACS expressions and the appearance of the specific user.

The set of parameters found by the AAM fitting algorithm previously described above uniquely describes a facial expression. The last step is to classify the combination of warp parameters as a particular expression. As mentioned previously, we chose to perform the analysis of the face in terms of positive and negative expressions, relating to the positive or negative experience of the user.

If we think of the set of shapes $\{s_0, \ldots, s_n\}$ as a basis of a (n+1)-dimensional space, the AAM fitting parameters $(p_0, \ldots, p_n)$ is a vector $\vec{p}$ in that space corresponding to neutral, positive or negative expression. So the problem is to know in which subspace $\vec{p}$ falls.

The approach followed was to project each of the hand labelled shapes, used for training the AAM, in that space and obtain a set of vectors. Since it is known for each expression in the training set if it is a positive or negative we can associate each of those vectors with a label. This process is performed at a training stage just after calculating the shapes PCA.

Now given the set of parameters $(p_0, \ldots, p_n)$ calculated by the AAM fitting algorithm, we can find the nearest vector and attribute the corresponding label. In the figure bellow the warp parameters represented by the vector p would correspond to a negative expression. The intensity is given by the dot product between both vectors.

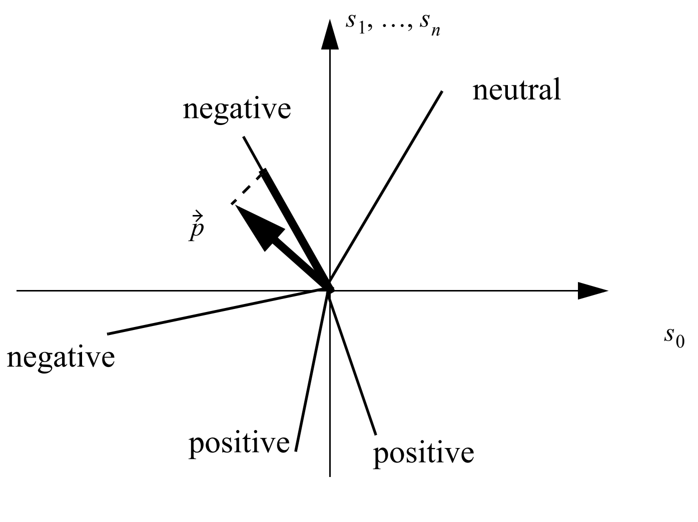


FIGURE 5-7. Classifying the expression and intensity. Simplified representation of the space $\{s_0, \ldots, s_n\}$

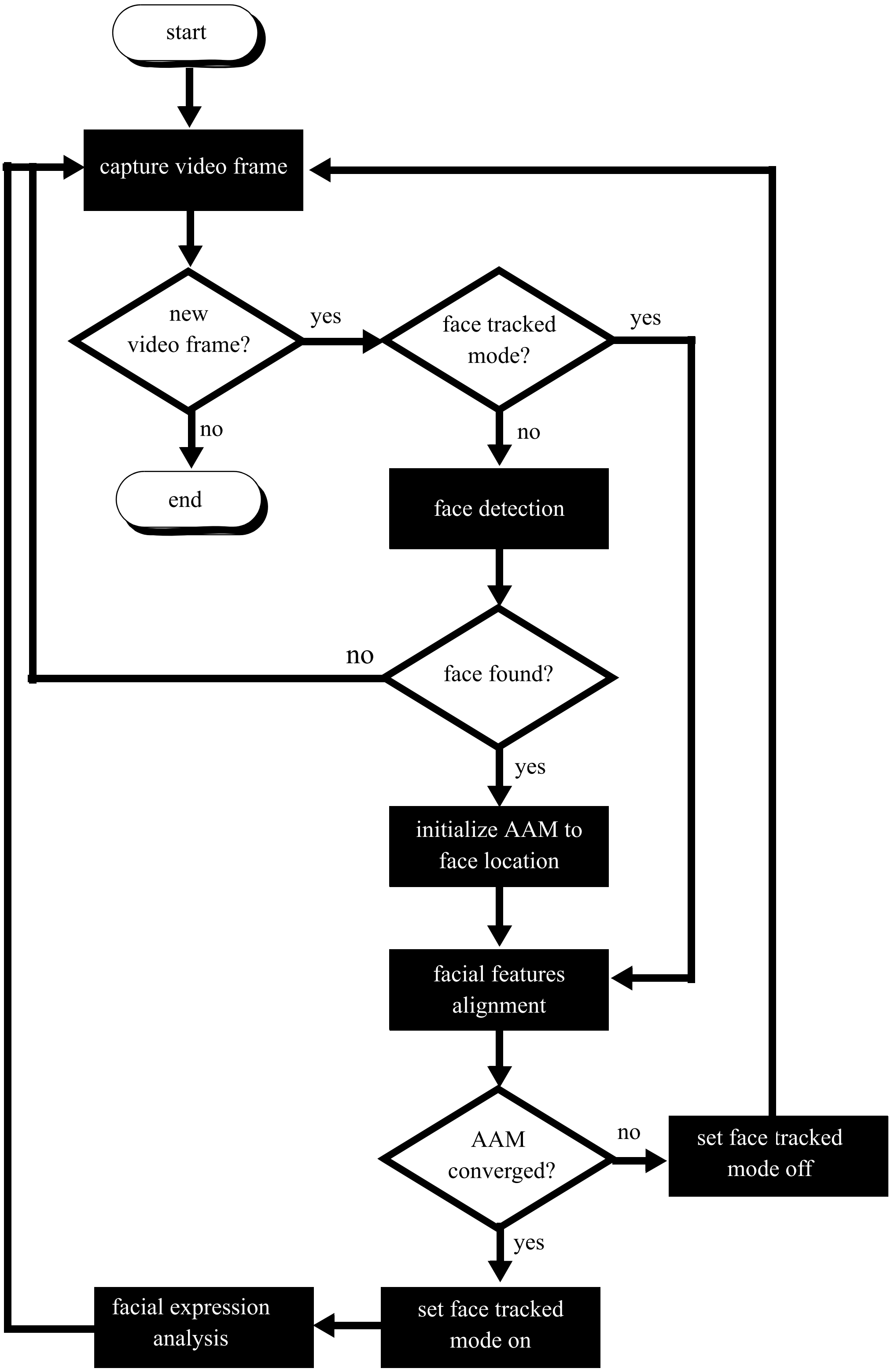


FIGURE 5-8. Overall FACEit Algorithm.

5.4.4 Evaluation

Overall measures of the system performance are directly influenced by the performances of the facial features alignment and the facial expression analysis. The quality of the AAM model 'fit' in the image will necessarily dictate the ability of the system to recognize an expression. If the AAM doesn't correctly perform the facial features alignment, the facial expression analysis performances will necessarily deteriorate. The analysis of the AAM performance is by itself a complex task: different number of shape and appearance models, image size, and the image itself can significantly alter the system performance for the same algorithm. Numbers on the AAM performance, and in particular on the inverse compositional algorithm implemented in FACEit, can be found on Matthews, Baker (2004). A comparison on the performance of an individual trained AAM, as used in FACEit, versus a generically trained AAM can be found on Gross, Matthews, and Baker (2005).

In an informal evaluation we used two video clips from two different users. Images were extracted from each video clip to create an appearance model for each user, while the shape model was created from two images from the FACS manual portraying a smile and a frown. The shape model, therefore, described only those two expressions. The testing was done with those same videos.

The first video contains 4 positive expressions (smile with different intensities), where head movement accompanies one of them. The second video contains 3 positive expressions with some head movement and 3 negative expressions. Note, that we just account the expressions that we can clearly identify, some subtle expressions can be ambiguous and therefore hard to classify even for a human observer. The number of expressions correctly identified and the number of mismatches were counted. An expression is considered identified when it was reported by the system at least once while the expression is shown.

In the first video clip, the smile with most head movement is mismatched as negative expression, and a neutral face is mismatched as negative. The remaining expressions are reported as positive. In the second video clip, 2 positive expressions and 2 negative expressions are correctly reported, and one frown is mismatched as positive. A neutral face is mismatched as negative.

Again, this is an informal evaluation, since there is a wide range of variables that can be adjusted, so that subtle changes have an affect on the system performance. Also, ideally, we should evaluate the system on a separate video stream, but since the training images were all taken under a similar illumination conditions, the AAM would perform poorly under illumination conditions different from the video.

Another aspect through which we can compare the performance of FACEit system is to compare its functionality against the features of an ideal system, based on the discussions in Section 5.2. Table 5-1 shows the comparison feature by feature.

TABLE 5-1. Comparison of the features of an ideal facial expression analysis system and FACEit

Features	Ideal	FACEit
Individual independency	y	p
Robust to variations in lightning	y	y
Robustness to occlusion	y	n
Recover from lost tracking	y	y
No markers/make-up required	y	y
Robustness to rigid movements	y	p
Robustness to fast movements	y	n
Sensibility to microexpressions	y	n
Automatic face detection	y	y
FACS coding	p	n
Allows different number and categories to be identified	y	y
Multiple interpretation labels	y	y
Quantified interpretation labels	y	y
Blended expression categories	y	y
No calibration	y	y
No manual intervention	y	y
Real-time	y	y

(y) - yes; (n) - no; (p) - partially

The most limiting factors of FACEit system are its lack individual independency, and the lack of robustness to occlusion. The lack of robustness to occlusion in practice translates in the inability of the algorithm to analyze facial expressions if the face is partially cover, for instance when the hand is in front of the mouth, or when the hairs cover partially the face.

CHAPTER 6

Research Context and Evaluation

Thinking is easy, acting is difficult, and to put one's thoughts into action is the most difficult thing in the world.

– Johann Wolfgang von Goethe

The intuitiveness of use and quality of information technology solutions is of prime interest for companies developing online services, software, and hardware solutions. Those qualities affect the consumers' perception of the brand or company, inevitably influencing their decisions. Companies, aiming to match the consumers' expectation, realize the importance of investing in usability practices and researching in new methods to evaluate the quality of the experience. The development of new products or improved versions certainly benefits from the knowledge on the users daily experience and their difficulties interacting with then. The development of widely accessible and non-intrusive user-monitoring solutions constitutes therefore a potential competitive advantage. This chapter discusses two study cases serving as a platform for evaluating applicability of the research conducted. The first was carried in collaboration with the usability group at Fidelity investments, a provider of financial services with online solutions where clients research investment products and execute financial operations. The second study was sponsored by SAP Research labs, USA, and evaluates an online e-commerce website prototype where virtual assistants assist the user through their shopping experience.

6.1 Complementing a Traditional Usability Lab Setting

Usability testing of software, websites, and interactive systems are crucial within the development cycle, and many more companies realize its potential return of investment (Black 2002). Usability testing is nevertheless an expensive endevour. Testing sessions can last for relative long periods of time, where at least one or two usability engineers are occupied conducting the protocol and annotating the session. The analysis process following the data collection is also time consuming, it requires browsing, observing and annotating recording logs. Necessarily, any approaches that reduce the time in any of those stages, bring an economical benefit, and are therefore welcome by the companies. Monitoring and recognizing users' facial expressions could assist in that process. For example, the participants recordings from a usability study could be indexed by the amount of negative expressions to quickly find the most problematic incidents during the session; also the total amount of users' negative facial expressions would constitute a fast process to establish a usability metric. This approach does not replace human usability engineers that are capable of retrieving more qualitative information, but could, in the manner described, assist and complement other usability measures.

Necessarily the implementation of such scenario must rely on robust facial expression analysis systems. The discussion on facial expression analysis systems in Chapter 5, highlighted the difficulty of current systems to match the required robustness. Nevertheless, we conducted two informal sessions on the use of the FACEit system in a usability lab. Fidelity eBusiness usability group at Fidelity Investments were interested in testing the concept of users' facial expression monitoring as a complement or alternative to their current protocols. Their goal is to conduct faster and less expensive studies, even if it meant trading off the quantity of information extracted from each session.

We analyzed video from usability session participants, recorded from a video camera placed over their computer screen, Figure 6-1. A couple of lessons were learned from those informal evaluations. When participants were using the think-aloud protocol or had the opportunity to engage in conservation with the observer, it made the facial expression analysis difficult. The successions of facial movement due to speech dramatically decreased the robustness of the FACEit system. While eventually a more robust system could circumvent that difficulty, it would still be hard to differentiate between expressions due to conversation from expressions signaling emotional reactions. Also the fact that the users' could vent their frustrations through speech could inhibit what otherwise would users' facial displays of frustration.

FIGURE 6-1. Participants' video during a testing session. The camera was placed over the monitor.

The users tasks consisted in searching for information on the website, for example finding what an annuity was, one of the financial products offered by the company. It required scanning and reading webpages for information, and navigating the website. This contrasts with task used in the study in Chapter 4, which required more interaction from the participants. That might justify why participants we observed in a second session, which were explicitly instructed not to speak, did not for the most part exhibit facial expressions, or were otherwise too subtile. This led us to conclude that the nature of the task, specifically the different levels of engagement, might also influence the users' predisposition to exhibit facial expressions.

6.2 Interfaces Responding to Users' Expressions

To this point we have discussed the identification of the user' perceived problems, and the collection of when those moments occurred to identify usability incidents. Implicit in the discussion is that the data collected on the events that were seen as most problematic, over a significant number of users, will be used to adjust and improve the interface. In the limit, one could consider that the ultimate goal of interfaces that monitor users' expressions is to provide users' with on-time assistance in response to negative expressions, or acknowledge and recorded what provoke users' satisfaction as suggested by positive facial displays. The development of interfaces that provide adaptive user guidance and support has been a topic of research of its own within the HCI community (e.g. Encarnação 1997; Encarnação and Stoev 1999). Situations, during the course of interaction, where problems arise and the user is in need of additional support, constitute a sweet spot to tailor the interface and provide assistant to the user. Monitoring users' facial expressions is a mechanism by which users' perceived problems could be identified, and since it is based on the users' own reactions, it might be less error prone than approaches that infer problems by analysis of interface events, e.g. Hilbert and Redmiles (2000). Eventually it could also complement those techniques.

We explored that scenario within the context of an online furniture shopping website based on the "Kliving" e-commerce demonstrator from SAP Research (Jaksic 2005). The site contains a humanlike sales assistant (social agent) capable of speech output and subtle facial expressions, Figure 6-2. While the presence of social agents in an application is not always appropriate, they have been shown to work well in motivating and engaging users in some computer activities that are social in nature, such as training, presentation and sales (Doyle 1999). We implemented and evaluated a social agent that reacts to the users' spontaneous display of emotions, in an attempt to reduce the frustration caused by usability problems. We present the study and discuss the results in the following sections.

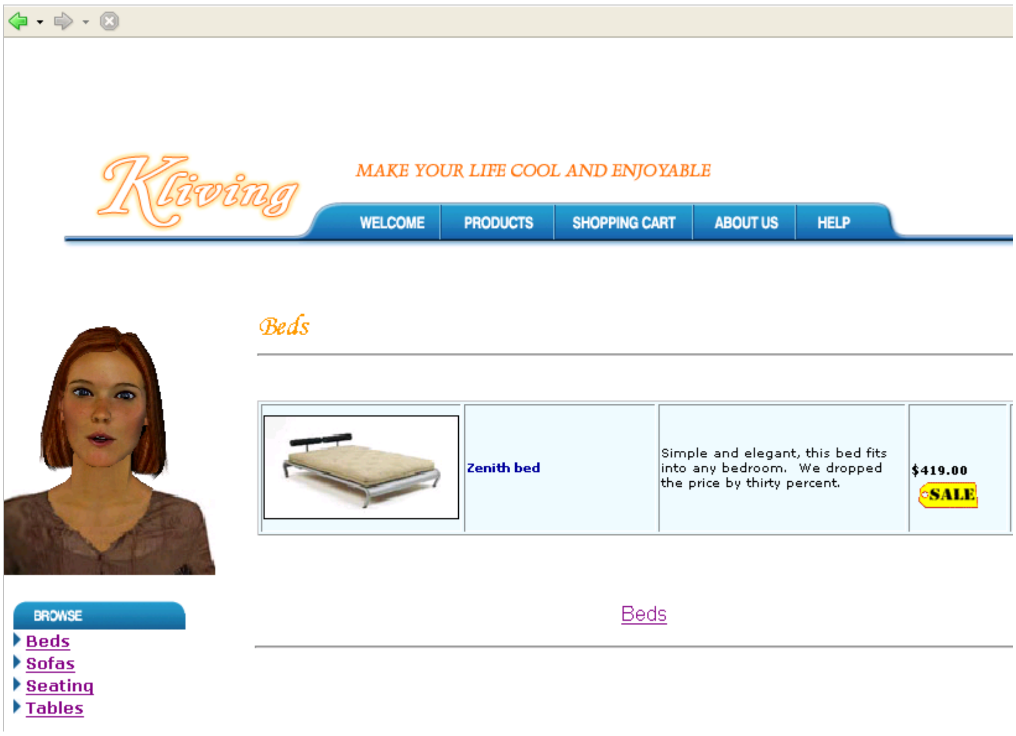


FIGURE 6-2. Furniture Shopping Website.

6.2.1 Study Design

We have conducted a study to evaluate how users respond to social agents that react to their facial expressions, as captured by a web camera. We investigated whether a virtual assistant capable of addressing users' emotions and mimicking affective responses has the ability to reduce frustration levels, engage users, and enhance the overall computing experience. More specifically the research questions investigated were:

- How do users assess virtual assistants that react to their facial expressions?
- Do they consider the interaction more positive and feel less frustrated when problems occur if the virtual assistant intervenes as a reaction to the user's display of negative affect?
- Similarly, do they find the interaction more pleasing and calming if the virtual assistant reacts positively to their smiling?

- If so, which conditions have to hold for this model of interaction to work and in which context?

Thirty users (18 females and 12 males), with at least a basic knowledge of computers, such as experience with email and Internet browsing, were recruited to represent users with different ranges of experience in online shopping. More than half of the users were recruited through an electronic posting on Brown University's graduate student bulletin, while the rest responded to a web posting on Craig's list for Providence, RI[1], a community electronic bulletin board. The announcement was posted in the volunteers section of this website. Since both recruiting methods involve an online posting, the participant pool was guaranteed to be comprised of those who use computers and the Internet. The majority of users (78%), were in their twenties.

The study was conducted as a two-condition experiment, where the participants were assigned randomly to each condition. In the control condition, the virtual assistant is programmed to speak solely based on current page location and user actions. Most of these comments are related to the products. In the experimental condition, the assistant also speaks in reaction to the participant's spontaneous facial expressions. The participants were informed that their face would be monitored with a camera and that in case they were randomly chosen to be in the experimental group, the social agent may respond to their facial expressions.

After hearing the navigation instructions from the social agent, users could begin browsing the website and shop furniture items. They were expected to purchase a minimum of ten items, at least one from each of the seven furniture categories. Two different versions of the website were used. In the first, the highly frustrating version, some of the products could not be purchased and several links were broken or led to the incorrect page. The second version, which was used after the first twelve users, eliminated some of the issues and was left with a purchasing problem in one product category and with one broken link. In both versions the website was fairly easy to navigate, though some of the product pages contain

1. http://providence.craigslist.org

small graphics of poor quality, and often very terse descriptions. Originally, there was also an unexpected shopping cart error that we later found to be exclusive to the Internet Explorer browser, which caused all of the purchased items to disappear randomly from the cart, and all subsequent purchases to never appear in it. This unpredictable shopping cart behavior was initially an unintended problem, which we, however, decided to keep for the first version of the website because it provoked frustration in our users and it represents a real-life scenario that causes annoyance and inconvenience to the costumer.

As previously mentioned, during the session participant's positive and negative facial expressions were monitored. Given the usability problems present in the website, designed to resemble real-life usability problems, we expected to observe mostly negative expressions. A perfectly robust facial expression monitoring was required in order to assure the quality of the data collection, beyond what FACEit could guarantee. Therefore, we used the Wizard of Oz technique to simulate an automated facial expressions analysis system. The Wizard of Oz technique enables not implemented or immature technology to be evaluated by using a human to simulate the response of a system. This technique can be used to test device concepts and techniques before it is implemented. A description of the technique is provided by Kelley (1984).

The investigators, who were remotely monitoring the participant's screen over the network and face via a web cam (Figure 6-3), controlled the animated character by triggering prerecorded phrases in response to the users' positive and negative facial expressions. Also within the definition of facial expressions we included head-nods and head-shakes. This are communicative signals, and not expressions of emotion, but were also included since it signal participants' agreement or disagreement. There were a total of five responses to negative facial expressions, see Table 6-1 on page 107.

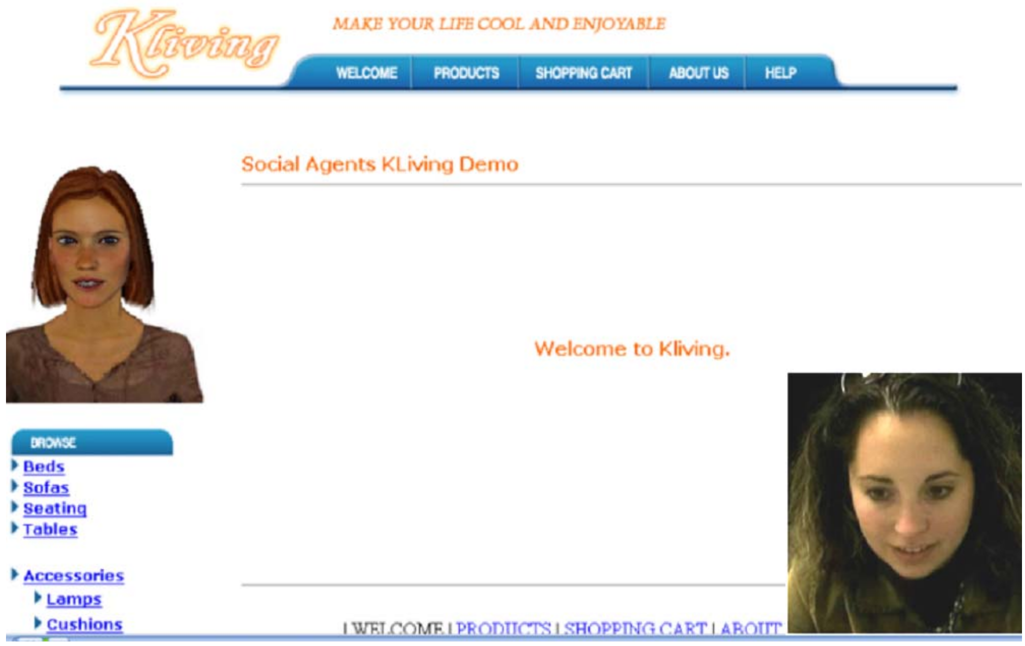


FIGURE 6-3. The investigator's view.

As is evident from the nature of the responses, the social agent does not try to provide a specific solution to the problem the user was experiencing. That task would require context awareness of the user task or the specific problem encountered. Nevertheless, that intervention can still be useful as argued by Klein, Moon and Picard (2002). They demonstrated that a computer agent capable of showing empathy and support, is enough to allow users to recover after a frustrating experience, and still walk away with a positive feeling about their interaction. Further note that in this approach the social agent doesn’t respond when the users’ encountered a problem *per se*, but only intervenes when a facial expression occurs. This is a more selective intervention, based on the users’ reaction, and eventually more acceptable, than trying to be apologetic all the time some incident arises. Given the results discussed in Chapter 4, the user is more likely to exhibit facial expressions for the events perceived as most difficult, therefore this form of intervention is more likely to relate to the users' effective perception of a problem.

TABLE 6-1. Scripted responses from the social agent.

Facial Expression	Social Agent Responses
Negative	I apologize if you are having troubles. Please be patient as we continue to improve our site
	Sorry if you are having troubles. I suggest writing the problem in the feedback form and moving on.
	Oh no, you seem to be having more problems. We will look into fixing them as soon as we can.
	If you are unhappy with our site, please let us know. We carefully consider all customer suggestions.
Positive	I hope you are having a good time!
	Looks like you are enjoying the shopping!

When finished, users filled out surveys that collected qualitative data about their experience with the application and the virtual sales assistant. At the end of the session, the subjects were debriefed and informed whether they participated in the control or experimental condition. Several participants offered comments and suggestions, which often provided additional meaning to survey results. The results are discussed next for each of the versions, comparing the control and experimental groups.

6.2.2 Results

As mentioned above, we ran two versions of the experiment. The first version, with the more frustrating website, had 6 users in the control and 6 users in the experimental condition. The second version was created in order to assess response to a slightly less frustrating application, after realizing that the first version was highly frustrating and resulted in overall negative assessment of the social agent in both conditions. Of the 18 participants in this version, two scores were not used due to factors that may have influenced their ratings. One of the participants did not purchase the minimum required number of items, resulting in an interaction that was too short to properly evaluate the application. The other user experienced a delay, which influenced the speech and movements of the social agent. As a result, there were 16 participants in the second version, 8 in the control condition and 8 in the experimental.

The questionnaire presented to the users at the end of the session offer a number of questions regarding the users' assessment of the social agent, those results were reported by Jaksic (2005). We focus the discussion on the results relate specifically with the facial expression monitoring.

Participants were asked to report the highest level of frustration that they experienced during their interaction with the application. The following scale was used:

1. Not frustrated
2. Slightly frustrated
3. Somewhat frustrated
4. Frustrated
5. Very frustrated

As expected participants using the first version of the website reported being more frustrated (3.8) than the second version, which was designed to be less frustrating (2.6). An independent samples t-test showed this to be a significant difference, with p=0.02.

Comparing the experimental and control group for the first version of the website, the two groups yielded averages of 4.3 and 3.2 respectively. While the sampling for this group is not big enough to yield statistical significance, that result suggests that users that interacted with social agent that reacted to their facial expressions seemed to have felt more frustrated.

On average, the participants of the second version reported being between slightly annoyed and somewhat frustrated (2.5), with experimental group being somewhat frustrated and the control just slightly annoyed.

Next, to assess if the users' preferred the assistant that reacted to their expressions, we asked participants to rate the influence of the assistant's behavior during the frustrating times of the interaction. The following scale was used:

1. Reduced frustration
2. Slightly reduced frustration
3. Did not have an influence
4. Slightly added to the frustration
5. Added to the frustration.

We expected the users in the experimental condition to report a more significant reduction in their frustration.

In the first version of the website (more frustrating), participants had an average score of 3.5, which is between "neutral" and "slightly added to frustration", there was no significance difference in the effect of the social agent whether it reacted to users' facial expressions or not.

In the second version, the control group users reported that the agent had "no influence" (3.1) but the experimental group reported the agent to slight reduce the frustration (1.7).

TABLE 6-2. Average ratings and effect sizes for how the assistant affected users' moods during frustrating times

	Control		Experimental		Effect size
	Mean	σ	Mean	σ	
Website version 1	3.3	1.3	3.7	1.4	0.3
Website version 2	3.1	0.7	1.7	0.8	2.0

The large effect size of 2.0 in the second version prompted us to perform a Mann-Whitney U test for significance between the average ratings of the two conditions. Unlike the t-test, the Mann-Whitney test does not assume normally distributed answers for the two independent groups, or a large sample size. The results of this test, Table 6-3, indicate that there is a significant difference between the average ratings of the two versions, with p=0.011.

The results from the second version indicate that when users experience moderate frustration, the social agent intervention to the users' facial expressions seem to help lower the frustration the user experienced. To assert this hypotheses we correlated the number of responses that each user received in reaction to negative facial expressions with how much the social agent influenced their mood. For users in the control condition, this number was always 0; the number varied between 1 and 4 for the experimental condition.

TABLE 6-3. Results of the Mann-Whitney Test for 2 independent samples

Rank	Condition	N	Mean Rank	Sum of Ranks
Reduce Frustration	0	7	9.43	66.00
	1	6	4.17	25.00
	Total	13		

Test Statistics	Reduced Frustration
Mann-Whitney U	4.000
Wilcoxon W	25.000
Z	-2.532
Asymp. Sig. (2-tailed)	0.011
Exact Sig. [2*(1-tailed Sig.)]	.014
Grouping Variable:	Condition

The result of this correlation yields a statistically significant positive correlation (0.73) between the number of interventions of the social agent and the effect that it had on reducing frustration, Table 6-4. It illustrates that there is a significant positive correlation between the number of affective phrases triggered and the effect that the assistant had on reducing frustration.

TABLE 6-4. Correlation between number of responses triggered due to negative facial expression and the reported positive effect of social agent on mood during frustrating times

		Negative Responses	Reduced Frustration
Negative Responses	Pearson Correlation	1	.730(*)
	Sig. (2-tailed)		0.003
	N	16	14
Reduced Frustration	Pearson Correlation	.730(*)	1
	Sig. (2-tailed)	0.003	
	N	14	14

(*) Correlation is significant at the 0.01 level (2-tailed).

6.2.3 Summary

In the first version of the website, most users reported being frustrated or very frustrated. For these users the intervention of the social agent seemed to further aggravated their frustration: slight trends indicated that those in the experimental group assessed the agent more negatively than those in the control group.

Although the second version of the experiment still caused some frustration to most users, most reported only a moderate amount of frustration. Furthermore, for those in the experimental group, the agent was reported to slightly reduce frustration, in contrast to the control group, which reported that the agent had no influence on their mood during frustrating times.

Those results suggest that in situations when users' are highly aggravated with problems with the interaction, the strategy to react, or at least to be apologetic, might in fact further aggravate users. On the other hand, when users' do not experience only moderated levels of frustration, they seem to be more receptive and appreciate the intervention. It is important to note that this results relate specifically with the use of virtual assistants, and that some users seem to react negatively to virtual assistants in general, as indicated by not only their negative assessment of this particular assistant but also of any previous interactions with virtual assistants. That attitude might biased negatively the perception of the social agents that respond to facial expressions.

6.3 Users' Acceptability of Facial Expression Monitoring

There are a number of ethical issues that need to be considered, and are often raised, regarding computers' ability to read or infer about one's emotional state. In particular regarding facial expressions, since video monitoring of the face can be very well seen as an invasion of the privacy. With that specific question in mind, we addressed the users in the study that was described in Section 6.2. We asked the participants:

How would you feel about having your facial expressions monitored by the computer during use of an application, for the purpose of detecting and addressing possible frustrations?

1. *Don't mind at all*
2. *Don't care*
3. *Should be fine in some situations*
4. *Prefer to never be monitored*

We were surprise to find that only 3 participants (10.7%) stated that they prefer no never be monitored. The remaining participants, 14 (50%) didn’t mind or didn’t care if they would be monitored, and 11 (39.3%) raised some reserves depending on the situation.

While certainly that group of participants in the study does not represent a sample of the generic population, and is certainly not representative of different cultures, it still holds as an interesting result within a certain segment of the population. The attitude suggested by this results, may reflect the degree of comfort with the presence of cameras that are ever more ubiquitous these days, extending its presence from stores to public places, to ATM machines.

CHAPTER 7

Conclusions and Future Research

It is better to ask some of the questions than to know all the answers.
– James Thurber

This last chapter summarizes the results discussed and the contributions presented. We proceed acknowledging the limitations of the work and pointing to numerous possible avenues for further work.

7.1 Summary of Results and Contributions

We started this work creating a motivation for the need to assess outside usability laboratories, in real-world environments, the quality of the users' interaction experience. We proposed to focus on the facial language and its natural relevance in social interactions as a mechanism to infer about user problems. Facial expressions occupy a central role in non-verbal communication and became, therefore, the focus of the discussion. Within the scope of the conducted research, of studying users' facial expressions as an indicator of their reactions researched, the following contributions are highlighted:

- The relevant psychology literature of the role of facial expressions and its significance for assessing the valence of users reactions were analyzed and discussed.
- A fundamental study was conducted to research the occurrence of users' facial expressions in a typical everyday computing scenario, i. e. while performing

word processing tasks. This study contributed to establish a correlation between users' perceived task difficulty and exhibited facial activity (Branco et. al 2005; Branco, Encarnação and Marcos 2006).

- A review on the different phsyiological monitoring sensors was conducted, leading to the focus on non-obtrusive camera-based approach.
- Different machine vision methods for the analysis of facial expressions were reviewed, and a tool was implemented to assist the research on unobtrusive monitoring of facial expression. We tested the applicability of that tool in a usability laboratory and concluded about its limitations.
- An original study was conducted with a real-world eCommerce application assessing the effects and acceptability of interfaces that respond to the user positive and negative facial expressions (Branco, Encarnação and Marcos 2006; Jaksic et al. 2006). The results suggested that this form of interaction is appropriate when users exhibit moderate levels of frustration.

7.2 Limitations and Future Work

Throughout this thesis we laid out the argument for monitoring users facial expressions as a mean to infer on the users emotional valence towards the interaction. Certainly there are a number of limitations in the research presented which create opportunities for continuing research. In Chapter 5 we reviewed the state-of-art in machine vision, and proposed a tool for monitoring users' facial expressions valence. The lack of robust tools capable of recognizing different expressions and discerning different user emotional valence is a limiting factor in the wide applicability of this monitoring modality, and is certainly limiting the related research. Significant research in robust, user and environment independent computer-vision technologies is needed to overcome these limitations.

In our research we studied users within laboratory environments for the solely purpose of monitoring their expressions. The unconstrained monitoring of users in real-world environments as they interact with a system, would allow for better assessment of the validity of our approach. Better machine vision solutions will be available in a near by future, enabling this effort to proceed.

However, even if computers will get better at analyzing our facial expressions, it is less likely that they do so well in analyzing the context of when they occur. This was a topic raised in Chapter 3, while discussing the interpretation of facial expressions. Being able to differentiate among situations where the users' expression is indicative of a problem, versus just a communicative signal or an irrelevant display of emotion requires some degree of context inference. The fusion of different types of information such as what is the user's active task, the focus of attention, the timing of the reaction with respect to the time of the action can all make for context inference that might help to disambiguate the interaction context. Research needs to be conducted in inferring context to disambiguate the relevance of the users' facial expression, mimicking to some extent the human skills in interpreting correctly a particular expression based on the situation.

Facial movement are just one of many other nonverbal behaviors - body movements, posture, gaze, pitch are additional ones. Humans combine effortless all those clues to interpret about others emotional and mental state. Understanding each of those modalities in itself and, preferably, in combination is certainly an important future research subject.

We focused on the role of facial expressions as an indicator of emotional valence. Its applicability goes beyond that purpose. It ranges from its potential as an alternative interaction modality for different-able people, to monitoring operators attention levels operating critical systems, to augmenting anthropomorphic interfaces to identifying users' expressions as an attempt to mimic more human-like responses. All these are by their own right significant research areas worth pursuing.

A vital topic that has been just briefly addressed in Chapter 6 is the acceptability and the associated ethical aspects of facial expression monitoring. To what extent the society and the different sub-cultures allow and accept that machines that will be capable to some degree of capturing the user emotions, will necessarily be a controversial topic.

7.3 Final Remarks

Machines that are able to interact socially, this work represents just one of many contributions in that direction necessary to reach that vision. It is not necessary to invoke the image of anthropomorphic interfaces to realize the potential of social interaction skills integrated in the technology that surrounds us. The emerging diversity of computing devices that we carry around, used by one specific user, opens the opportunity to develop one-on-one relationships, in the sense that those devices get to know intimately the user. The relevant quality of those devices would be their ability to integrate seamlessly within our daily-lives, similar to the perfect assistant that knows our likes and adapts to our requests. The qualities users will appreciate in ever ubiquitous computing will not be described in terms of speed or memory capacity, but instead on their ability to connect with the user, understand him and anticipate him and the surrounding environment. With the physical dimensions of those devices constantly decreasing there is the possibility for those to be incorporated in clothing, accessories and objects we interact with, imperceptibly monitoring our reactions and all relevant factors that reveal our preferences and physical state, and disambiguating them through the contexts of our environment and history.

References

Andreassi, J. L. 2000. *Psychophysiology: Human behavior and physiological response.* 4th ed. Mahwah, NJ: Lawrence Erlbaum Associates.

Anttonen, J., and V. Surakka. 2005. Emotions and heart rate while sitting on a chair. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI '05), 2–7 April, 2005, Portland, Oregon, USA*, 491–499. New York: ACM Press.

Ark, S. W., D. C. Dryer, and D. J. Lu. 1999. The Emotion Mouse. In *Proceedings of the 8th International Conference on Human-Computer Interaction: Ergonomics and User Interfaces* 1, 818–823.

Backs, R. W., and W. Boucsein, eds. 2000. *Engineering Psychophysiology: issues and applications*. Mahwah, NJ: Lawrence Erlbaum.

Baker, S. and I. Matthews. 2004. Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision*, 56 (3):221-255. Also available online: http://www.ri.cmu.edu/pubs/pub_4031.html.

Baker, S., I. Matthews, J. Xiao, R. Gross, T. Ishikawa, and T. Kanade. 2004. Real-time Non-rigid Driver Head Tracking for Driver Mental State Estimation. In *Proceedings of the 11th World Congress on Intelligent Transportation Systems, 18-22 October, 2004*, Nagoya, Japan. Also available online at http://citeseer.ist.psu.edu/baker04realtime.html

Ball, E. 2002. A Bayesian Heart: Computer Recognition and Simulation of Emotion. In *Emotions in Humans and Artifacts,* ed. R. Trappl, P. Petta and S. Payr, 303–332. Cambridge, MA: MIT Press.

Baron-Cohen, S. 1995. *Mindblindness: an essay on autism and theory of mind.* Cambridge, MA: MIT Press.

Bartlett, M. S., J. C. Hager, P. Ekman, and T. J. Sejnowski. 1999. Measuring facial expressions by computer image analysis. *Psychophysiology*, 36(2):253–263.

Bartlett, M. S., P. A. Viola, T. J. Sejnowski, B. A. Golomb, J. Larsen, J. C. Hager, and P. Ekman, P. 1996. Classifying facial action, *Advances in Neural Information Processing Systems* 8, MIT Press, Cambridge, MA. p. 823-829. Also available online: ftp://ftp.cnl.salk.edu/pub/marni/nips95.ps.Z.

Bessiere, K., I. Ceaparu, J. Lazar, J. Robinson, and B. Shneiderman. 2002. *Social and Psychological Influences on Computer User Frustration*. Tech. Report HCIL-2002-19. Maryland: University of Maryland Human-Computer Interaction Lab. http://www.cs.umd.edu/localphp/hcil/tech-reports-search.php?number=2002-19 (accessed 21 August, 2006).

Black J. 2002. Usability Is Next to Profitability. BusinessWeek Online. http://www.businessweek.com/technology/content/dec2002/tc2002124_2181.htm.

Black, M., and Y. Yacoob. 1995. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion, *Fifth International Conf. on Computer Vision, ICCV'95*, June 1995, Boston, MA, 374-381. Also available online: http://www.cs.brown.edu/~black/Papers/iccv95-copyright.html.

Boucsein, W. 2000. The Use of Psychophysiology for Evaluating Stress-Strain Processes in Human-Computer Interaction. In Backs and Boucsein 2000, 289–309.

Bradski, G. R. 1998. Computer Vision Face Tracking For Use in a Perceptual User Interface. *Intel Technology Journal* Q2: 15. Also available online: http://citeseer.ist.psu.edu/bradski98computer.html.

Branco, P., L. M. Encarnação, and A. Marcos. 2006. It's all in the Face: Studies on Monitoring Users' Experience. In *Proceedings of SIACG 2006 Ibero American Symposium in Computer Graphics, July 5-7, Santiago de Compostela, Spain*.

Branco, P., P. Firth, L. M. Encarnação, and P. Bonato. 2005. Faces of Emotion in Human-Computer Interaction. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'05)* Extended Abstracts, ACM Press, 1236-1239.

Brave, S., and C. Nass. 2003. Emotion in Human-Computer Interaction. In *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications*, ed. J. A. Jacko and A. Sears, 81–96. Mahwah, NJ: Lawrence Erlbaum Associates.

Bromberg, M. G. 2005. The Poker Corner: Lying about the Nuts. *Barstool Sports* 19 October, 2005–1 November, 2005. Also available online: http://www.barstoolsports.com/article/poker_corner/509/.

Cacioppo, J. T., G. G. Berntson, J. T. Larsen, K. M. Poehlmann, and T. A. Ito. 2000. The Psychophysiology of Emotion. In Lewis and Haviland-Jones 2000.

Chandrasiri, N. P., T. Naemura, and H.Harashima. 2001. Real time facial expression recognition system with applications to facial animation in MPEG-4. *IEICE Transactions on Information and Systems* E84-D (8):1007–1017.

Clore, G. L., and A. Ortony. 2000. Cognition in Emotion: Always, Sometimes, or Never? In Lane and Nadel 2000, 24–61.

Cootes, T. F. and C. J. Taylor. 2004. Statistical Models of Appearance for Computer Vision. http://www.isbe.man.ac.uk/~bim/Models/app_models.pdf.

Damasio, A. R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam Publishing Group. Page references are to the paperback edition New York: Harper Collins, 1995.

———. 2000. A Second Chance for Emotion, In Lane and Nadel 2000, 12–23.

Darwin C. 1872/1998. *The expression of the emotions in man and animals*. Third Edition, with Introduction, Afterwords and Commentaries by Paul Ekman. New York: Oxford University Press.

Davidson, R. J. 1994. Complexities in the Search for Emotion-Specific Physiology. In Ekman and Davidson 1994b, 237–242.

Davidson, R. J., and P. Ekman. 1994. Afterword: Is There Emotion-Specific Physiology? In Ekman and Davidson 1994b, 261–262.

Davis, J. W., and S. Vaks. 2001. A perceptual user interface for recognizing head gesture acknowledgements. Proceedings of *2001 workshop on Perceptive user interfaces*, *Orlando, Florida*, 1–7.

Doyle, P. 1999. When is a communicative agent a good idea?. In *Proceedings of Agents-99*, Workshop on communicative Agents.

Edwards, G. J., C. J. Taylor, and T. F. Cootes. 1998. Interpreting Face Images using Active Appearance Models. In *International Conference on Face and Gesture Recognition*, 300–305.

Ekman, P. 1972. Universals and cultural differences in facial expressions of emotion. In *Nebraska symposium on motivation, 1971,* ed. J. Cole, 207–283. Lincoln, Nebraska: University of Nebraska Press.

———. 1982. *Emotion in the human face*. New York: Cambridge University Press.

———. 1994. Moods, Emotions, and Traits. In Ekman and Davidson 1994b, 56–58.

———. 1997. Should we call it expression or communication? *Innovations in Social Science Research* 10, 333-344.

———. 2003. Darwin, Deception and Expression. In *Emotions Inside Out*, ed. P. Ekman, J. Campos, R.J. Davidson, and F.W. de Waal, Annals of the New York Academy of Sciences.

Ekman, P., and R. J. Davidson. 1994a. Affective Science: A Research Agenda. In Ekman and Davidson 1994b, 411–430.

———, eds. 1994b. *The Nature of Emotion.* New York: Oxford University Press.

Ekman, P., and W. V. Friesen. 1977. *Facial Action Coding System*. Consulting Psychologists, Palo Alto, CA, 1977. Also available as CD-ROM by Ekman, P., W.V. Friesen, and J. C. Hager. 2002. *The Facial Action Coding System*. Salt Lake City: Research Nexus.

Ekman, P., W.V. Friesen, and J. C. Hager. 2002. *Facial Action Coding System: Investigator's Guide*. 2nd ed. Salt Lake City: Research Nexus CD-ROM.

El Kaliouby, R., and P. Robinson. 2004a. FAIM: Integrating Automated Facial Affect Analysis in Instant Messaging. In *Proceedings of the 9th international conference on Intelligent user interface table of contents, 13 - 16 January, Funchal, Madeira, Portugal,* 244–246. New York: ACM Press.

———. 2004b. Mind Reading Machines: Automated Inference of Cognitive Mental States from Video. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*.

Encarnação, L. M. 1997. Multi-Level User Support through Adaptive Hypermedia: A Highly Application-Independent Help Component. In Proceedings of the International Conference on Intelligent User Interfaces (IUI'97) January, Orlando, FL., 187-194.

Encarnação, L. M. and S. L. Stoev. 1999. An application-independent intelligent user support system exploiting action-sequence based user modeling. In *Proceedings of the 1999 International Conference on User Modeling (UM'99)*, 20-24 June 1999, Banff, Canada, 245-254.

Essa, I. A. 1995. *Analysis, Interpretation, and Synthesis of Facial Expressions*. Ph.D. diss., MIT Department of Media Arts and Sciences. http://citeseer.ist.psu.edu/essa95analysis.html.

etForecasts. 2006. Computers-in-use by Country: Historical estimates and forecast of the number of PCs-in-use and total computers-in-use for over 55 countries. Executive Summary. http://www.etforecasts.com/products/ES_cinusev2.htm (accessed 22 August, 2006).

FACSAID. Facial Action Coding System Affect Interpretation Dictionary (FACSAID). http://face-and-emotion.com/dataface/facsaid/description.jsp

Fasel, B., and J. Luettin. 2003. Automatic Facial Expression Analysis: A Survey. *Pattern Recognition* 36 (1):259–275. Also available online: http://www.idiap.ch/publications/fasel03a-art.bib.abs.html.

Frijda, N. H. 1994. Varieties of Affect: Emotions and Episodes, Moods, and Sentiments. In Ekman and Davidson 1994b, 59–67.

Goodman, L. A. and W. H. Kruskal. 1979. *Measures of association for cross classifications.* New York: Springer-Verlag.

Gross, R., I. Matthews, and S. Baker. 2005. Generic vs. Person Specific Active Appearance Models. *Image and Vision Computing* 23(11):1080–1093.

Gu, H., Q. Ji, and Z. Zhu. 2002. Active facial tracking for fatigue detection. In *Proceedings Sixth IEEE Workshop on Applications of Computer Vision, 2002 (WACV 2002), 3–4 December, Orlando, FL USA*, 137–142. IEEE Computer Society.

Hartson, H. R., and J. C. Castillo. 1998. Remote evaluation for post-deployment usability improvement. In *Proceedings of Advanced Visual Interfaces (AVI 98), 24–27 May, 1998, L'Aquilla, Italy*, 22–29.

Hartson, H. R., J. C. Castillo, J. Kelso, J. Kamler, and W. C. Neale. 1996. Remote Evaluation: The Network as an Extension of the Usability Laboratory. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'96), 13–18 April, 1996, Vancouver, BC, Canada*, 228–235. New York: ACM Press.

Hazelett, R. L. 2003. Measurement of User Frustration: A biologic approach. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI '03), April 5–10, 2003, Ft. Lauderdale, Florida, USA,* Extended abstracts 734–735. New York: ACM Press.

———. 2006. Measuring Emotional Valence during Interactive Experiences: Boys at Video Game Play. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI '06)*, *22–27 April, Montreal, Quebec, Canada*, 2:1023–1026. New York: ACM Press.

Hilbert, D. M., and D. F. Redmiles. 2000. Extracting usability information from user interface events. *ACM Computing Surveys (CSUR)* 32(4): 384–421.

Himer, W., F. Schneider, G. Kost, and H. Heimann. 1991. Computer-based analysis of facial action: A new approach. *Journal of Psychophysiology* 5(2): 189–195.

Hjelmas, E. 2001. Face Detection: A Survey. *Computer Vision and Image Understanding* 83(3): 236–74.

Hodges, P. W., B. H. Bui. 1996. A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography. *Electroencephalography and Clinical Neurophysiology* 101, 511–519.

Intel. 2006a. Open Source Computer Vision Library. http://www.intel.com/technology/computing/opencv/index.htm.

———. 2006b. Intel Integrated Performance Primitives. http://www.intel.com/cd/software/products/asmo-na/eng/perflib/ipp/index.htm.

Jaksic, N. 2005. Assessing User Response to Affective Social Agents. Master's Thesis. University of Rhode Island.

Jaksic, N., P. Branco, P. Stephenson, and L. M. Encarnação. 2006. The Effectiveness of Social Agents in Reducing User Frustration. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'06)* Extended Abstracts, 917–922.

Kaiser, S., and T. Wherle. 1992**.** Automated coding of facial behavior in human-computer interaction with FACS. *Journal of Nonverbal Behavior* 16(2): 65–140.

———. 2001. The Role of Facial Expression in Intra-individual and Inter-individual Emotion Regulation. In *Emotional and Intelligent II: The Tangled Knot of Cognition.* Papers from the 2001 AAAI Fall Symposium, ed. D. Cañamero, 61–66. Menlo Park, CA: AAAI Press. Available online: http://citeseer.ist.psu.edu/499853.html

Kaiser, S., T. Wehrle, and S. Schmidt. 1998. Emotional episodes, facial expression, and reported feelings in human-computer interactions. In *Proceedings of the Xth Conference of the International Society for Research on Emotions*, ed. A. H. Fischer, 82–86. Würzburg: ISRE Publications.

Kelley, J. F. 1984. An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Office Information Systems*, 2 (1):26–41.

Keltner, D., and P. Ekman. 2000. Facial Expression of Emotion. In Lewis and Haviland-Jones 2000.

Klein, J., Y. Moon, and R. W. Picard. 2002. This computer responds to user frustration: theory, design and results. *Interacting with Computers* 14:119-140.

Knapp, M. L., and J. A. Hall. 2006. *Nonverbal Communication in Human Interaction,* 6th ed. Belmont, CA: Thomson.

Ko, J. G., K. N. Kim, and R.S. Ramakrishna. 1999. Facial feature tracking for eye-head controlled human computer interface. In *Proceedings of IEEE TENCON'99, 15–17 September, Cheju, Korea*, 72–75.

Koenemann-Belliveau, J., J. M. Carroll, M. B. Rosson, and M. K. Singley. 1994. Comparative usability evaluation: critical incidents and critical threads. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '94)*, *24–28 April, Boston, MA*. 245–251. New York: ACM Press.

Lane., R. D., and L. Nadel, eds. 2000. *Cognitive Neuroscience of emotion*. New York: Oxford University Press.

Lazar, J., K. Bessiere, I. Ceaparu, J. Robinson, and B. Shneiderman. 2003. Help! I'm Lost: User Frustration in Web Navigation. *IT&SOCIETY* 1 (3): 18–26.

Lazar, J., A. Jones, K. Bessiere, I. Ceaparu, and B. Shneiderman. 2004. *User Frustration with Technology in the Workplace*. Tech Report HCIL-2004-12. Maryland: University of Maryland Human-Computer Interaction Lab. http://www.cs.umd.edu/localphp/hcil/tech-reports-search.php?number=2004-12 (accessed 21 August, 2006).

Ledoux, J. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster. Page references are to the paperback edition New York: Simon & Schuster, 1998.

Lewis, C., and J. Reiman. 1993. *Task-Centered User Interface Design: A Practical Introduction*. Boulder, CO: University of Colorado.

Lewis, M., and J. M. Haviland-Jones, eds. 2000. *Handbook of Emotions*, 2nd ed. New York: Guilford Press.

Lien, J. 1998. *Automatic Recognition of Facial Expressions Using Hidden Markov Models and Estimation of Expression Intensity.* Doctoral dissertation, Carnegie Mellon University Technical Report CMU-R1-TR-31. http://www.ri.cmu.edu/pubs/pub_468.html (accessed 28 August, 2006).

Lienhart, R., and J. Maydt. 2002. An extended set of Haar-like features for rapid object detection. In Proceedings of the *International Conference on Image Processing (ICIP'02),* September, Rochester, NY, 1:900–903.

Littlewort, G., I. Fasel, M. Stewart Bartlett, and J. R. Movellan. 2002. *Fully automatic coding of basic expressions from video*. Technical Report, UCSD INC

MPLab. Also available online: http://mplab.ucsd.edu/techreports/mplab2002.03.pdf.

Logitech. 2006. Visual effects. http://www.logitech.com/index.cfm/products/videoeffects/welcome/US/EN,crid=2354.

Lombardi, J., and M. Betke. 2002. *A Self-initializing Eyebrow Tracker for Binary Switch Emulation*, Tech. Report 2002-023. Boston, MA:Boston University. http://www.cs.bu.edu/techreports/pdf/2002-023-eyebrow-tracker.pdf (accessed 23 August, 2006).

Mase, K. 1991. Recognition of facial expression from optical flow. *Institute of Electronics Information and Communication Engineers Transactions* E74: 3474–3483.

Matthews, I. and S. Baker. 2004. Active Appearance Models Revisited. *International Journal of Computer Vision*, 60 (2):135–164. Also available online: http://www.ri.cmu.edu/pubs/pub_4601.html.

MORI. 2002. Computer Users Suffer 'PC Rage'. http://www.mori.com/polls/2002/bthomecomputing.shtml.

Morimoto, C.H., D. Koons, A. Amir, and M. Flickner. 2000. Pupil detection and tracking using multiple light sources. *Image and Vision Computing* 18: 331–335.

Pantic, M., and L. J. M. Rothkrantz. 2000. Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (12): 1424–1445.

Picard, R. W. 1997. *Affective computing*. Cambridge, MA: MIT Press. Page references are to the paperback edition Cambridge, MA: MIT Press, 1998.

———. 2000. Toward computers that recognize and respond to user emotion. *IBM Systems Journal* 39 (3&4):705–719. Riverton, NJ: IBM Corp.

Picard, R. W, and J. Healey. 1997. Affective Wearables, *Personal Technologies* 1 (4):231–240.

Picard, R. W., E. Vyzas, and J. Healey. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (10): 1175–1191.

Pyun, H., H. J. Shin, T. H. Kim, and S. Y. Shin. 2001. *Real-time Facial Expression Capture for Performance-driven Animation*. Tech Report CS-TR-2001-167. Korea Advanced Science and Technology. http://www.cs.wisc.edu/~joony/RESEARCH/face_capture.pdf (accessed August 23, 2006).

Quinlan P, J. Lane, L. Aspinall. 1997. Effects of hot tea, coffee and water ingestion on physiological responses and mood: the role of caffeine, water and beverage type. *Psychopharmacology* 134(2): 164–73.

Reeves, B. and C. Nass. 1996. *The media equation: how people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.

Reynolds, C. 2001. *The Sensing and Measurement of Frustration with Computers*. Master's thesis, MIT. http://vismod.media.mit.edu/pub/tech-reports/TR-559.pdf (accessed 21 August, 2006).

Rowe, D., J. Sibert, and D. Irwin. 1998. Heart Rate Variability: Indicator of User State as an Aid to Human-Computer Interaction. In Proceedings of the *Conference on Human Factors in Computing Systems (CHI '98), April 18–23, Los Angeles, CA.* 480–487. New York: ACM Press.

Salovey, P. and Mayer, J. D. 1990. Emotional Intelligence. *Imagination, Cognition, and Personality* 9:185–211. Baywood Publishing Co., Inc.

Scerbo, M. W, F. G. Freeman and P. J. Mikulka. 2000. A Biocybernetic System for Adaptive Automation. In Backs and Boucsein 2000, 241–253.

Scheirer, J., R. Fernandez, J. Klein, and R.W. Picard. 2002. Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers* 14(2): 93–118.

Scheirer, J., R. Fernandez, and R. W. Picard. 1999. Expression glasses: a wearable device for facial expression recognition. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI '99) Extended Abstracts, 15–20 May, Pittsburgh, Pennsylvania*. 262–263. New York: ACM Press.

SeeingMachines. 2006. FaceLAB 4. http://www.seeingmachines.com/facelab.htm.

Shneiderman, B. 2002. *Leonardo's laptop: human needs and the new computing technologies*. Cambridge, MA: MIT Press.

Smith, S. W. 1997. *The Scientist and Engineer's Guide to Digital Signal Processing*. San Diego, CA: California Technical Publishing.

Sony. 2005. EyeToy. http://www.eyetoy.com.

Surakka, V., I. Marko, and P. Isokoski. 2004. Gazing and Frowning as a New Human-Computer Interaction Technique. In ACM Transactions on Applied Perception 1 (1): 40–56. New York: ACM Press.

Tian, Y., T. Kanade, and J. F. Cohn. 2001. Recognizing Action Units for Facial Expression Analysis. In IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2): 97–116.

van Reekum, C. 2000. Levels of Processing in Appraisal: Evidence From Computer Game Generated Emotions. Doctoral dissertation, University of Geneve.

Viola, P., and M. Jones. 2001. Rapid Object Detection using a Boosted Cascade of Simple Features. In Proceedings of *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)* 1:511–518. Los Alamitos, CA: IEEE Computer Society.

Ward R D, D. Bell, and P. H. Marsden. 2003. An exploration of facial expression tracking in affective HCI. In *People and Computers XVII: Designing for Society; Proceedings of HCI 2003*, ed. P. Palanque, P. Johnson and E. O'Neill. 383-399. London, UK: Springer-Verlag.

Ward, R. D., and P. H. Marsden. 2004. Affective computing: problems, reactions and intentions. *Interacting with Computers* 16, 707-713.

Ward, R. D., P. H. Marsden, B. Cahill, and C. Johnson. 2001. Using skin conductivity to detect emotionally significant events in human-computer interaction. In *Proceedings of BCS IHM-HCI 2001, 10–14 September, 2001, Lille, France,* 2: 20–25.

Wehrle, T. and Kaiser S. 2000. Emotion and Facial Expression. In *Affective interactions: towards a new generation of computer interface,* ed. A. M. Paiva, 49–63. Berlin: Springer-Verlag.

Wilson, G. and M. A. Sasse. 2000. Investigating the Impact of Audio Degradations on Users: Subjective vs. Objective Assessment Methods. In *Proceedings of OZCHI 2000:Interfacing Reality in the New Millennium*, ed. C. Paris, N. Ozkan, S. Howard and S. Lu, 135-142.

APPENDIX I

Research Consent Form #1

All the participants in the study presented in Chapter 4 were given the following approved research consent form. This study was conducted at the Motion Analysis Laboratory in Spaulding Rehabilitation Hospital, Boston, MA, USA.

Research Consent Form

Brigham and Women's Hospital
Massachusetts General Hospital
Newton-Wellesley Hospital
North Shore Medical Center
Spaulding Rehabilitation Hospital
Partners Community HealthCare, Inc.

Version 5.1: March 2003

Subject Identification

Protocol Title: EMG-Based Methods for Design of Human-Computer Interface

Principal Investigator: Paolo Bonato, PhD

Site Principal Investigator: Paolo Bonato, PhD

Description of Subject Population: Healthy Subjects

PURPOSE
We would like permission to enroll you as a participant in a research study**.** The purpose of the study is to develop methods of assessing computer user awareness and emotional state in order to improve the efficiency of human computer interfaces. We will use recordings of the electrical signals from some of the muscles (called electromyograms or EMG's) of your face during the accomplishment of tasks performed during computer use. This project is expected to lead to an enhanced computer environment where software can adapt a computer's media outlets to complement the attention given by the user.

PROCEDURES
The testing session will be divided into two stages. Surface EMG electrodes will be positioned over several facial muscles before beginning the first stage and will remain in the same positions until completion of the second stage. You will be situated at a computer station and during the first stage of the testing will be asked to simply "play" the video game Tetris for 30 minutes. The second part will involve working with Microsoft Word to format a prescribed printed page. Your face will also be video taped during the session and you will be asked to complete a short questionnaire regarding the experience. The overall procedure will take approximately three hours to complete.

For your participation in this study you will receive $40.00 for your time and transportation costs.

STUDY CONTACTS
If you have any questions about this study, you may contact the Motion Analysis Laboratory staff at (617) 573-2728 or Dr. Paolo Bonato at (617) 573-2745. If you have any medical questions or think you may have experienced a research related injury, you may contact Dr. Frontera at (617) 573-7180 or Dr. Krivickas at (617) 573-2761.

COSTS
There will be no charge to you for these tests.

Page 1

Research Consent Form

Brigham and Women's Hospital
Massachusetts General Hospital
Newton-Wellesley Hospital
North Shore Medical Center
Spaulding Rehabilitation Hospital
Partners Community HealthCare, Inc.

Version 5.1: March 2003

Subject Identification

RISKS AND DISCOMFORTS
The test does not imply any serious risk of injury. There is a slight risk of an allergic skin reaction from the adhesive tape used to secure the electrodes. This risk has been minimized by using materials that minimize allergic reactions.

BENEFITS
You are not expected to personally benefit from your participation in this study. The information collected in this study will be useful in developing more effective ways to assess risk factors for musculoskeletal disorders.

ALTERNATIVES
You may refuse to participate in this study without prejudice to your current or future medical care at Spaulding Rehabilitation Hospital.

Page 2

Research Consent Form

Brigham and Women's Hospital
Massachusetts General Hospital
Newton-Wellesley Hospital
North Shore Medical Center
Spaulding Rehabilitation Hospital
Partners Community HealthCare, Inc.

Version 5.1: March 2003

Subject Identification

PRIVACY AND CONFIDENTIALITY

Federal law requires Partners HealthCare System, Inc. and its affiliated hospitals, researchers, health care providers, and physician network to protect the privacy of information that identifies you and relates to your past, present, and future physical and mental health and conditions ("protected health information"). If you enroll in the research described in this consent form, your "protected health information" will be used and shared with others as explained below.

1. **What protected health information about me will be used or shared with others during this research?**
   - ☐ Existing medical records.
   - ☒ New health information created from study-related tests, procedures, visits, and/or questionnaires.

2. **Why will protected health information about me be used or shared with others?**
   - The main reasons include:
     - to conduct and oversee the research described earlier in this form;
     - to ensure the research meets legal, institutional, and accreditation requirements; and
     - to conduct public health activities (including reporting of adverse events or situations where you or others may be at risk of harm).
   - Other reasons may include for treatment, payment, or health care operations. For example, some medical information produced by this study may become part of your hospital medical record because the information may be necessary for your medical care. (You will also be given the Partners' Notice for Use and Sharing of Protected Health Information which provides more information about how Partners and its affiliates use and share protected health information.)

3. **Who will use or share protected health information about me?**
   - Partners and its affiliated researchers and entities participating in the research will use and share your protected health information. In addition, the Partners review board that oversees the research at Partners and its affiliated staff who have a need to access this information to carry out their responsibilities (for example, oversight, quality improvement, and billing) will be able to use and share your protected health information.

4. **With whom outside of Partners Healthcare System may my protected health information be shared?**
   All reasonable efforts will be made to protect the confidentiality of your protected health information, which may be shared with the following others for the reasons noted above:
   - Outside individuals or entities that have a need to access this information to perform functions on behalf of Partners and its affiliates (for example, data storage companies, insurers, or legal advisors).
   - ☐ The sponsor(s) of the study and/or its agents: ______
   - ☒ Other researchers and medical centers participating in this research, if applicable.

Page 3

Subject Identification

- ☐ Federal and state agencies (for example, the Department of Health and Human Services, the Food and Drug Administration, the National Institutes of Health, and/or the Office for Human Research Protections), or other domestic or foreign government bodies if required by law and/or necessary for oversight purposes.
- ☐ Hospital accrediting agencies.
- ☐ A data and safety monitoring board organized to oversee this research, if applicable.
- ☐ Other, specify: ______

We recognize that some of those who receive protected health information may not have to satisfy the privacy requirements that we do and may redisclose it, so we share your information only if necessary and we use all reasonable efforts to request that those who receive it take steps to protect your privacy.

5. **For how long will protected health information about me be used or shared with others?**
   - There is no scheduled date at which your protected health information that is being used or shared for this research will be destroyed, because research is an ongoing process. Research information may be analyzed and re-analyzed in light of scientific and medical advances, or reviewed for quality assurance, oversight, or other purposes.

6. **Statement of privacy rights:**
   - You have the right to withdraw your permission for the researchers and participating Partners entities to use or share your protected health information. We will not be able to withdraw all of the information that already has been used or shared with others to carry out the research or any information that has been used or shared with others to carry out related activities such as oversight, or that is needed to ensure the quality of the study. If you want to withdraw your permission, you must do so in writing by contacting the researcher listed as the Study Contact.
   - You have the right to choose not to sign this form. If you decide not to sign, you cannot participate in this research study. However, refusing to sign will not affect your present or future care and will not cause any penalty or loss of benefits to which you are otherwise entitled.
   - ☒ You have the right to request access to your protected health information that is used or shared during this research and that relates to your treatment or payment for your treatment, but you may access this information only after the study is completed. To request this information, please contact the researcher listed under Study Contacts on the consent form.

<u>**PUBLICATION OF RESULTS OR USE FOR TEACHING PURPOSES**</u>

The results of this study may be published in a medical book or journal or used for teaching purposes. However, your name or other identifiers will not be used in any publication or teaching materials without your specific permission.

Research Consent Form

Brigham and Women's Hospital
Massachusetts General Hospital
Newton-Wellesley Hospital
North Shore Medical Center
Spaulding Rehabilitation Hospital
Partners Community HealthCare, Inc.

Version 5.1: March 2003

Subject Identification

REQUEST FOR MORE INFORMATION

You may ask more questions about the study at any time. The investigator(s) will provide their telephone number so that they are available to answer your questions or concerns about the study. You will be informed of any significant new findings discovered during the course of this study that might influence your continued participation. A copy of this consent form will be given to you to keep.

If you want to speak with someone not directly involved in the study about your rights as a research subject, your participation in the study, any concerns you may have about the study, or a research-related injury, contact a representative of the Human Research Committee at (617) 573-2366. You can also contact them if you feel under any pressure to enroll or continue to participate in this study.

REFUSAL OR WITHDRAWAL OF PARTICIPATION

Participation in this study is voluntary. Refusal to participate or dropping out of the study at any time will involve no penalty or loss of benefits to which you are otherwise entitled or affect your present or future care by the doctors or the participating hospitals. In addition, the doctor in charge of this study may decide to end your participation in this study at any time after he/she has explained the reasons for doing so and has helped arrange for your continued care by your own doctor, if needed. Please also see the statement of privacy rights above if you wish to withdraw permission for your health information to be used and shared for study purposes.

INJURY STATEMENT

If you are injured during the course of the study and as a direct result of this study, you should contact the investigator at the number provided under the Study Contacts section in this form. You will be offered the necessary care to treat that injury. This care does not imply any fault or wrong-doing on the part of the Partners institutions participating in this study or the doctor(s) involved. Where applicable, the appropriate Partners institution participating in this study reserves the right to bill third party payers for services you receive for the injury and to make other decisions concerning payment in such instances. The Hospitals will not provide you with any additional compensation for such injuries.

CONSENT TO PARTICIPATE IN RESEARCH AND AUTHORIZATION TO USE OR RELEASE INDIVIDUAL HEALTH INFORMATION FOR RESEARCH

I confirm that the purpose of the research, the study procedures, the possible risks and discomforts and potential benefits that I may experience have been explained to me. Alternatives to my participation in this research study also have been discussed. All my questions have been answered. I have read this consent form. My signature below indicates my willingness to participate in this research study and my authorization to use and share with others my “protected health information” as described in the preceding paragraphs.

Research Consent Form

Brigham and Women's Hospital
Massachusetts General Hospital
Newton-Wellesley Hospital
North Shore Medical Center
Spaulding Rehabilitation Hospital
Partners Community HealthCare, Inc.

Version 5.1: March 2003

Subject Identification

SIGNATURES:

____ ________________
Subject or Parent(s), if minor child Date/Time

OR, if applicable, individual authorized by subject to make health care decisions

____ ________________
Court-appointed Guardian/Health Care Proxy Date/Time

OR

____ ________________
Family Member/Next-of-Kin Date/Time

Identify relationship to subject:____

Subject's preferred contact information during course of study:___________________________

____

I have explained the purpose of the research, the study procedures, identifying those that are investigational, the possible risks and discomforts and potential benefits. I have answered any questions regarding the research study to the best of my ability.

____ ________________
Investigator/Individual Obtaining Consent Date/Time

In certain situations, the Human Research Committee will require the use of a subject advocate in the consent process. The subject advocate is an individual who has no vested interest in the research and who agrees to act as an impartial third party in the consent process.

____ ________________
Subject Advocate (if required by the HRC for this study) Date/Time

Page 6

APPENDIX II

Questionnaire Study #1

The following questionnaire was presented to all the participants of the study discussed in Chapter 4, after the testing session.

EMG-Based Methods for Design of Human-Computer Interface

Subject Ref # ___________

In a scale from 1 (lowest) to 5 (highest) how would you classify your experience with the Word application? 1 2 3 4 5

Please indicate bellow all the incidents and difficulties you found while performing the word processing task you been asked for. Use the space to state briefly what you were trying to do at the time and what happened, and circle the most appropriate answer for the questions bellow. Use as many pages as you need.

___

___

___

___

Did you consider this event an obstacle for the task you were trying to do? Yes No
Did you face a similar situation before? Yes No
Did you know what to do to solve the problem? Yes No
Did you consider the situation caused by you
or by the computer/software? You Computer/software

___

___

___

___

Did you consider this event an obstacle for the task you were trying to do? Yes No
Did you face a similar situation before? Yes No
Did you know what to do to solve the problem? Yes No
Did you consider the situation caused by you
or by the computer/software? You Computer/software

more space on other side

APPENDIX III

Research Consent Form #2

Follows the consent form for research, given to all participants prior to the experiment discussed in Section 6.2. This form was approved by University of Rhode Island's Internal Review Board on Human Subjects on June 13, 2005. The study was conducted at IMEDIA, Providence RI, USA.

IMEDIA Academy
400 Westminster Street
Providence RI 02903

The University of Rhode Island
Department of Computer Science
Tyler Hall
Kingston RI 02881

Title of Study: Assessing User Response to Virtual Humans with Social Intelligence

CONSENT FORM FOR RESEARCH

Purpose:
You have been asked to take part in a research project described below. The researcher will explain the project to you in detail. You should feel free to ask questions. If you have more questions later, Ms. Nada Jaksic (401) 996-0662 and Mr. Pedro Branco, (401) 383-1900 ext 118., will discuss them with you.

Description of the project:
You have been asked to take part in the study that examines how computer users react to on-screen virtual assistants, particularly those capable of recognizing and addressing user emotions by monitoring their facial expressions.

Procedures:
If you decide to take part in this study you will be situated at a computer station equipped with a camera that will monitor your face. After answering a few short questions, you will begin interacting with an online shopping website and given a list of specific tasks that you are expected to complete. You will then have the opportunity to continue interacting freely with the application, for up to 5 minutes. When finished, you will be given another questionnaire regarding your experience with the application. The investigators will then debrief you on the how the experiment was conducted and the procedures used. The overall procedure is expected to take approximately 30 minutes. For your participation in the study you will receive $15.00 for your time and transportation costs.

Risks or discomfort:
This experiment does not imply any serious risk of injury. You may experience some annoyance or frustration while using the online shopping website.

Benefits of this study:
You are not expected to personally benefit from your participation in this study. The information collected in this study will be useful in developing more effective ways to interact with computers.

Confidentiality:
Your part in this study is confidential within legal limits. The researchers and IMEDIA and the University of Rhode Island will protect your privacy, unless they are required by law to report information to city, state or federal authorities, or to give information to a

court of law. Otherwise, none of the information will identify you by name. All records will be identified by a number.

Decision to quit at any time:
The decision to take part in this study is yours only. You do not have to participate. If you decide to take part in the study you may quit at any time. However your decision might be, you will not be penalized in any way.   If you wish to quit, you simply inform the person conducting the study.

Rights and Complaints:
If you are not satisfied with the way this study is performed, you may discuss your complaints with Ms. Nada Jaksic (401) 996-0662, Mr. Pedro Branco at (401) 383-1900x118 or with Dr. Peter Stephenson at (401) 383-1900x201, anonymously, if you choose. In addition, you may contact the office of the Vice Provost for Graduate Studies, Research and Outreach, 70 Lower College Road, Suite 2, University of Rhode Island, Kingston, Rhode Island, telephone: (401) 874-4328.

You have read the Consent Form. Your questions have been answered. Your signature on this form means that you understand the information and you agree to participate in this study.

______________________	______________________
Signature of Participant	Signature of Researcher
______________________	______________________
Typed/printed Name	Typed/printed name
______________________	______________________
Date	Date

Please sign both consent forms, keeping one for yourself.

APPENDIX IV

Pre-Experiment Questionnaire Study #2

The following questionnaire was presented to the participants of the study discussed in Section 6.2, before the testing session.

Date: ___/___/___
Time: ___:____

Participant ID#_________

PRE-EXPERIMENT QUESTIONNAIRE

1. What is your gender? F M

2. What is your age group?

☐ 19 or under
☐ 20-24
☐ 25-29
☐ 30-34
☐ 35-39
☐ 40-44
☐ 45-49
☐ 50-59
☐ 60+

3. How often do you shop online?

☐ 0 (Never)
☐ 1 (Rarely -- 1 or 2 times a year)
☐ 2 (Occasionally -- 3 – 5 times a year)
☐ 3 (Frequently -- 6-10 times a year)
☐ 4 (Very frequently -- more than 10 times a year)

4. Please rate, in the order of importance, the biggest disadvantages of online shopping, with 1 being the most important and 5 the least.

Security _______

Technical annoyances _______

Lack of human interaction _______

Boring _______

Inability to see/feel the items _______

Too complicated ________

5. Have you ever interacted with a virtual assistant (an animated on-screen character that offers assistance with your computer application)?

Yes ☐
No ☐

1

Continued on next pa

If you answered yes to question 5, please answer questions 6-8.

6. Which virtual assistant(s) have you interacted with (name or brief description of application(s) and assistant(s)) ?

___

___

7. Please check off any characteristics below that describe your experience with the virtual assistant.

Entertaining	☐
Helpful	☐
Pleasing	☐
Calming	☐
Engaging	☐
Annoying	☐
Distracting	☐
Useless	☐

8. How would you rate your experience with the assistant? (check one)

Positive	☐
Negative	☐
Neutral	☐

___

Thank you for your time. Please submit this survey to the investigator, who will give you further instructions.

APPENDIX V

Post-Experiment Questionnaire Study #2

The following questionnaire was presented to the participants of the study discussed in Section 6.2, after the testing session.

Date: ___/___/___
Time: ___:____

Participant ID#_________

POST-EXPERIMENT QUESTIONNAIRE

1. If you've shopped online before, how would you compare this shopping application with ones without a shopping assistant?

☐ 0 (N/A -- Have not shopped online before)
☐ 1 (A lot less interesting and fun)
☐ 2 (Less interesting and fun)
☐ 3 (About the same)
☐ 4 (More interesting and fun)
☐ 5 (Much more interesting and fun)

2. How friendly/approachable did you find the shopping assistant?

☐ 1 (Unfriendly/Unapproachable)
☐ 2 (Somewhat unfriendly/Unapproachable)
☐ 3 (Neutral)
☐ 4 (Friendly/Approachable)
☐ 5 (Very friendly/Approachable)

3. How intelligent did the shopping assistant seem?

☐ 1 (Not intelligent)
☐ 2 (Some basic intelligence)
☐ 3 (Somewhat Intelligent)
☐ 4 (Intelligent)
☐ 5 (Very Intelligent)

4. Please rate your level of agreement with the following characteristics that may describe the shopping assistant.

	Strongly disagree	Somewhat Disagree	Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree
Annoying	☐	☐	☐	☐	☐	☐	☐
Distracting	☐	☐	☐	☐	☐	☐	☐
Useless	☐	☐	☐	☐	☐	☐	☐
Entertaining	☐	☐	☐	☐	☐	☐	☐
Helpful	☐	☐	☐	☐	☐	☐	☐
Pleasing	☐	☐	☐	☐	☐	☐	☐
Calming	☐	☐	☐	☐	☐	☐	☐
Motivating/Engaging	☐	☐	☐	☐	☐	☐	☐

5. Did it seem like the shopping assistant was sympathetic to any feelings of frustration that you may have felt during the interaction?
□1 (Not at all)
□ 2 (At times)
□ 3 (Most of the time)
□ 4 (Always)
□ 5 (Not sure)

6. During the most frustrating part of the shopping experience today, what was the level of frustration you experienced?
□ 1 (Not frustrated)
□ 2 (Slightly annoyed)
□ 3 (Somewhat frustrated)
□ 4 (Frustrated)
□ 5 (Very frustrated)

7. How did the assistant's behavior at times of frustration influence your mood?
□ 0 (Did not have an influence on mood)
□ 1 (Slightly reduced frustration)
□ 2 (Reduced frustration)
□ 3 (Slightly added to the frustration)
□ 4 (Added to the frustration)
□ 5 (Not Sure)

8. What about the times when you were not frustrated; i.e. the interaction with the application was going smoothly?
□ 0 (No influence)
□ 1 (Somewhat elevated mood/ enhanced experience)
□ 2 (Elevated mood/ enhanced experience)
□ 3 (Significantly elevated mood / enhanced experience)
□ 4 (Slightly annoyed/distracted me)
□ 5 (Annoyed/distracted me)
□ 6 (Significantly annoyed/distracted me)
□ 7 (Not sure)

9. Overall, how would you rate your online shopping experience today?
□ 1 (Annoying/Frustrating)
□ 2 (Not very enjoyable)
□ 3 (Neutral)
□ 4 (Enjoyable)
□ 5 (Very enjoyable)

10. Based on what you saw today, has your willingness to do online shopping (circle one) increased/decreased/stayed the same

11. In the future, would you like to see shopping assistants when you do your online shopping?

☐ 1 (Definitely Not)
☐ 2 (Probably Not)
☐ 3 (Not Sure / Neutral)
☐ 4 (Probably)
☐ 5 (Definitely)

12. Please rate how much you would like to see a virtual assistant in the following types of software or online applications:

	Definitely Not	Probably Not	Not Sure	Probably	Definitely
Office Applications	☐	☐	☐	☐	☐
Job Training	☐	☐	☐	☐	☐
Online courses	☐	☐	☐	☐	☐
Online banking	☐	☐	☐	☐	☐
Language learning partner	☐	☐	☐	☐	☐
Tax preparation software	☐	☐	☐	☐	☐
Technical support	☐	☐	☐	☐	☐

13. How would you feel about having your facial expressions monitored by the computer during use of an application, for the purpose of detecting and addressing possible frustrations?

☐ 1 (Don't mind at all)
☐ 2 (Don't care)
☐ 3 (Should be fine in some situations)
☐ 4 (Prefer to never be monitored)

Thank you for participating in our study! We appreciate your time. Please submit the survey to the investigators and collect your reward.

For investigator use only: Additional time spent shopping: ____ minutes.