Timo Partala

Affective Information in Human—Computer Interaction

ACADEMIC DISSERTATION
To be presented with the permission of the Faculty of Information Sciences of the University of Tampere, for public discussion in the Pinni auditorium B1096 on October 29th, 2005, at noon.

Department of Computer Sciences
University of Tampere

Dissertations in Interactive Technology, Number 1
Tampere 2005

ACADEMIC DISSERTATION IN INTERACTIVE TECHNOLOGY

Supervisor:	Assistant professor Veikko Surakka, Ph.D., Department of Computer Sciences, University of Tampere, Finland
Opponent:	Professor Gary Bente, Ph.D., Institute of Psychology, University of Cologne, Germany
Reviewers:	Professor Heikki Mannila, Ph.D., Department of Computer Science, University of Helsinki, Finland Professor Göte Nyman, Ph.D., Department of Psychology, University of Helsinki, Finland

Electronic dissertation
Acta Electronica Universitatis Tamperensis 476
ISBN 951-44-6440 -0
ISSN 1456-954X
http://acta.uta.fi

Dissertations in Interactive Technology, Number 1

Department of Computer Sciences
FIN-33014 University of Tampere
FINLAND

ISBN 951-44-6428-1
ISSN 1795-9489

Tampere University Press
Tampere 2005

Contents

List of publications

This thesis consists of a summary and the following original publications, reproduced here by permission.

Acknowledgements

The current thesis is a result of many years of research work. When writing my thesis, I have been part of the TAUCHI research unit at the department of computer sciences. First, I want to thank Veikko Surakka, the supervisor of my thesis, who has definitely been the biggest motivator for the current work. His enthustiastic guidance and support were valuable when I was preparing this thesis. Professors Kari-Jouko Räihä and Roope Raisamo have done a substantial job in developing our research unit into a high-class unit and providing good conditions for the researchers. The helpful administrative and technical staff has also made my work much easier. I would like to thank all the colleagues and students that I have worked with during the years, especially Toni Vanhala, Maria Jokiniemi, and Jussi Lahti, who have directly contributed to the research presented in this thesis. Mika Käki helped me in the publication issues of this thesis. I would also like to thank professor Pertti Järvinen, who introduced me to research, when I was still a M. Sc. student.

Most funding for my thesis came from Tampere Graduate School in Information Sciences and Engineering (TISE). Thus, I would like to thank TISE coordinator Pertti Koivisto and chairman Markku Renfors for arranging me this opportunity. After TISE funding, the heads of the department, professors Seppo Visala and Jyrki Nummenmaa provided me with an opportunity to work for the department, while finishing my thesis. For one academic year I also received funding from the Academy of Finland. The external reviewers of this dissertation were academy professor Heikki Mannila and professor Göte Nyman. I would like to thank them for their comments on the manuscript of this thesis. I also thank professor Gary Bente, who was willing to travel from Germany to be the opponent of this dissertation.

Finally, I want to thank my parents and other relatives for supporting me throughout my career in the education system. I would also like to thank my friends and all the people near me during the years for their support.

Abstract

The aim of this thesis was to study the possibilities of utilizing affective information in human-computer interaction. Throughout the thesis, a dimensional theory of emotions has been used as a frame of reference. According to it, emotions can be defined as three dimensions, which are valence (negative-positive), arousal (calm–highly aroused), and dominance (controlled–in control). The research is organized around four research goals. The first goal was to study, whether pupil size variation can be used in computer input to give the computer information about the user's affective responses. The second goal was to study the effects of affective speech interventions to the user's subjective experiences, physiology, and behavior. Third, we studied, whether experienced affective valence can be estimated reliably in real time by measuring the activations of the facial muscles associated with smiling and frowning. Fourth, the effects of the simulated proximity of a conversational character to the subjects' affective experiences were studied.

The results concerning pupil size variation showed generally that pupil size was bigger in response to positive and negative highly arousing stimuli, as compared to neutral stimuli. Thus, the results suggest that the magnitude of the pupillary response has a linear relationship with experienced affective arousal and a curvilinear relationship with experienced affective valence. The results concerning the experiment using affective interventions suggested that the users respond emotionally to human-computer interaction events and that by using affective interventions in the context of a negative event, positive affect-related changes can be induced in the user in terms of subjective experiences, facial activity, and task-related behavior.

The results for the estimation of emotions based on the activations of facial muscles were approximately on the same level for both adaptive and person-independent estimation. The results supported the idea that affective valence can be estimated relatively reliably by measuring the activations of the muscles activated in smiling and frowning. The best estimation accuracies, discriminating between positive and negative responses at the accuracy of over 80%, were obtained by measuring frowning activity, and a difference score of the activations of the two muscles, when the subjects watched emotionally arousing videos. The results of the experiment, in which the proximity of a conversational agent was simulated by the size of the agent on the screen, suggested that the

subjects' experienced dominance – or the feeling of control – can be significantly influenced by varying the size of the agent on the screen.

As a whole, the results suggest that pupil size variation and facial expressions can give the computer information about the user's affective responses, and they could be potential input signals for human-computer interaction in the future. In addition, the results suggest that the wording of synthetic speech interventions and messages can be effectively used in computer output to influence the user's affective physiological responses, behavior, and experienced valence. By varying the simulated proximity of a conversational agent, the users' experienced affective dominance can also be influenced.

1 Introduction

Computer use has often been regarded as purely rational activity, in which emotions are secondary, or even can get in the way of successful computer use. This traditional view has fortunately begun to change. We all know intuitively that emotions are involved in our everyday activity and guide our decisions and behavior. Lately, the importance of emotion-related issues in human thinking and behavior has also been emphasized by the scientific research community. As I started the research work for this thesis near the change of millennium, the research in the field of affective human-computer interaction was still in its infancy. As it is now, affective human-computer interaction research has evolved into perhaps the most significant new trend in human-computer interaction research. The technical and ideological developments, which have led to the worldwide interest in affective human-computer interaction, can be best understood by examining the long history of human-computer interaction.

Computers and the ways people use them has changed dramatically in the past few decades. The first computers could only be fully used by the engineers who designed them. The first user interfaces were command-based interfaces, in which the user wrote commands to the computer using the keyboard. Since then, a number of different methods for the interaction between the human and the computer have been developed. An important step forward was taken in the 1980's, when graphical user interfaces (GUIs) became the dominant user interfaces. Consequently, the mouse became the dominant input device along the keyboard. The need for more advanced interaction techniques produced a new research area, human-computer interaction, which has existed in large since the 1980's. This research area has constantly become larger in terms of the number of research forums and published research articles.

During the 1990's human-computer interaction developed into a noteworthy worldwide research area. The rapid development of computing technology enabled a variety of new technologies for computer input and output. User interface design and evaluation was also recognized as a critical factor in computer software development. As it is now, a large proportion of research on computers has connections to the human-computer interaction research area. The mouse and the keyboard are still the most usual methods for controlling the computer, but many other techniques are also being developed and used. It is possible to communicate with computers using, for example, speech, gestures, or gaze direction. The computer may also present information to the user using different media, for example, using text, pictures, video, sounds, speech and haptic feedback. Different input and output techniques can also be combined to gain additional benefits. Multimodal interfaces (e.g. speech and touch) and multimedia interfaces (including e.g. pictures and speech) have become more common (Oviatt, 2002; Sutcliffe, 2002).

Significant changes have also happened in the research methods used. Traditionally, human-computer interaction has been studied using cognitive and behavioral methods. In the design of computer systems, cognitive task analysis has been emphasized. Similarly, evaluation methods in the field of human-computer interaction have concentrated on studying the user's cognition of the user interface. For example, the GOMS (Goals, Operators, Methods and Selection rules) method (Card *et al*., 1983), and the different cognitive walkthrough methods (Butler, 1996) are based on cognitive psychology. Computing systems have also been evaluated using different behavioral performance metrics, for example, task completion times, error rates and the quality of output (Frøkjær *et al.*, 2000). In usability evaluation, there has also been a widely used measure called subjective satisfaction. Subjective satisfaction can be defined as the level of the user's positive attitudes towards the system or as the user's preference from different alternative solutions (Frøkjær *et al.*, 2000).

The change from the cognitive paradigm took place in large during the past few decades when the results of many psychological experiments stressed the importance of emotions in all human thinking and behavior, and emotion studies gained increased attention in the field of psychology. In 1997, the first book on the use of emotions in human-computer interaction, 'Affective computing' by Rosalind Picard (1997) was published. She defined affective computing as 'computing that relates to, arises from, or deliberately influences emotions'. The goal is to enable the computer to respond to the user's emotions in order to improve human-computer interaction on an affective level. Picard visioned that emotions are not only useful, but they are required for building truly intelligent computers. Bates (1994) suggested that emotions are one of the primary means for building believable agents. Brave and Nass (2002) suggested that any interface that ignores a user's emotional state or fails to manifest the appropriate emotion can dramatically impede performance and risks of being perceived as cold, socially inept, untrustworthy, and incompetent.

Other researchers have taken a more conservative view. For example, Sloman (1999) maintains that emotion-like mechanisms will probably be needed as a side product, when building intelligent systems, but they should not be integrated into systems for their own sake. Shneiderman (2002) takes a historical view and notes that early attempts at incorporating human-human interaction styles into computer software (e.g. Microsoft’s Bob and Clippit) have failed. According to him, it is important to note that users want to be in rapid control of their computers and information when designing software. He sees the mimicry of human form and behavior suitable mainly for electronic toys and animation productions.

While the question of what will be the role of emotions and affective technology in human-computer interaction in the long run is still unanswered, it seems safe to say that the initial results in this field have been promising. These results have been obtained in practical experiments in emotion recognition or use in interactive systems. During the new millennium, the area has gained more and more academic and commercial interest. For example, many reputable human-computer interaction journals have dedicated special issues to affect-related topics, and affective computing research is now invited to many worldwide human-computer interaction conferences. Recently, the first international conference specializing in affective computing (ACII, 2005) has emerged. It seems that affective computing has established its place in the scientific community, and it offers a lot of challenging research opportunities. First commercial companies around the theme of affective computing have emerged (e.g. Affective Media, 2005), and some major companies have developed both prototypes and widely marketed products utilizing affective technology.

The current thesis aims at contributing to this development by giving new information on the different emotion-related input and output channels that could be useful during human-computer interaction in practice. Specifically, the current thesis concentrates on pupil size variation and facial expressions as sources of affective information and the use of synthetic speech and the simulated proximity of human-like conversational agents in computer output.

2 Emotions

2.1 Introduction to emotions

While it is widely recognized that emotions are central in human behavior and communication, there is no universally agreed unambiguous definition of emotions available. In the beginning of the 1980's, Kleinginna and Kleinginna (1981) had already recorded nearly a hundred different definitions presented in the scientific literature. Most definitions describe emotion from only one aspect or only one subset of what is generally considered as emotion. A rather comprehensive definition for emotions was later suggested by Oatley and Johnson-Laird (1987):

"Emotions are part of a management system to co-ordinate each individual's multiple plans and goals under constraints of time and other resources. Emotions are part of the biological solution to the problem of how to plan and to carry out action aimed at satisfying multiple goals in environments, which are not perfectly predictable. Emotions are based on non-propositional communications which we will call 'emotion signals'. They function both to set the whole system suddenly into a particular mode, and to maintain it tonically in that mode. Emotion signals provide a specific communication system which can invoke the actions of some processors and switch others off."

Despite its relative obscurity, the term emotion has established its place as the primary scientific construct for describing the psychological and physiological phenomena described, for example, by Oatley and Johnson-Laird above. However, there are also other concepts that are clearly related to emotions, but have a slightly different meaning. An example of such a concept is mood. Emotions are typically regarded as short-term reactions. This distinguishes them from moods, which are longer lasting than emotions, as they can last for weeks or even months. Furthermore, emotions are often responses to emotion-triggering external and internal events, while moods lack a specific target, and are often temporally and situationally separated from the events that elicited the mood. Furthermore, emotions should be distinguished from the concept of feeling. For example, Damasio (1999) refers to a feeling as the private, mental experience of an emotion. According to this definition, feelings do not include bodily emotional responses, but merely mental perceptions of the state of the body.

Another concept frequently applied to emotion-related phenomena is affect. In early affective human-computer interaction research the words emotion and affect have been used somewhat interchangeably, which I will also do in this thesis. Similarly, both the words affective and emotional could be used to refer to the phenomena involving emotions, although I prefer affective, since it avoids the negative associations sometimes linked with the word emotional. Thus, I follow in this thesis the conceptual practices used, for example, by Picard (1997).

As it was already learnt from the definition of emotion presented above, emotion is a multifaceted construct. Öhman (1987) has presented a model, which describes the structure of emotion (Figure 1). In this model, emotional phenomena are caused by an emotionally meaningful situation. Emotional phenomena can be manifested in three different ways: verbal reports, physiological responses and behavior. These three ways of manifestation are also central in emotion research. Lang (1993) suggested that the behaviors related to each manifestation are partly independent and consequently the covariation of the different emotion measures can be poor. This implies that it is important to study emotions on all three levels.

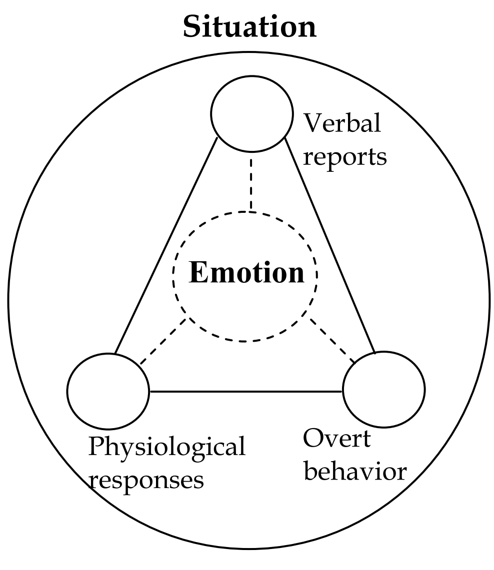

Figure 1. *The structure of emotion as suggested by Öhman (1987).*

In concordance with this, Izard (1977) has suggested that emotional reactions usually involve the following components:

- a change in the neurophysiological functions of different brain areas
- a change in the neuromuscular activity
- a change in behavior (e.g. avoidance vs. approach)
- a subjective emotional experience

Emotions are typically studied by carrying out practical experiments measuring the different manifestations or consequences of emotional phenomena, for example, those suggested by Öhman (1987) and Izard (1977). The lack of an unambiguous definition does not consequently prevent from studying the related phenomena successfully. In contrast, by studying emotional phenomena scientifically, it is possible to obtain a realistic idea of what emotions are and how can they be studied.

Emotions are generally considered as products of evolution. This idea was first expressed by Charles Darwin in his classic book Expression of Emotion in Man and Animal (1872). Plutchik's (1980) psychoevolutionary theory of basic emotions states that the concept of emotion is applicable to all evolutionary levels and applies to animals as well as to humans. Emotions have an evolutionary history and have evolved in various forms of expression in different species. Importantly, emotions have served an adaptive role in helping organisms deal with key survival issues posed by the environment. For example, Frijda (1986) and Lang (1995) view emotions as action dispositions – or states of readiness that vary widely in reported affect, physiology and behavior. Cosmides and Tooby (2000) suggested that emotions are mechanisms, which are functionally specialized for solving different adaptive problems that arose during our ancestral hunter-gatherer history. While the meaning of emotions in the evolutionary sense (e.g. adaptation for survival) has diminished, emotions are still important as they give meaning to everyday life (Surakka, 1998), and are important in all human thinking and behavior as suggested in section 2.4.

2.2 Discrete and dimensional emotions

Emotion theories can be divided to two different groups on a high level of abstraction: discrete and dimensional emotions. The former group defines emotions as a set of discrete categories, while the latter group assumes that emotions can be defined as continuous dimensions. As it is now, most scientists consider the two groups of theories as compatible and complementary to each other. Assumed that the discrete categories of emotion can be defined as subspaces in the space formed by the emotion dimensions, the two theories indeed seem complementary rather than exclusionary. In practical studies, however, usually only either of the two approaches is used. In the following chapters, the basic principles of the discrete and dimensional approaches are explained.

2.2.1 DISCRETE EMOTIONS

Different discrete emotion approaches share the idea of dividing the emotional space into discrete categories. Each category represents one emotion. The labels of these categories have mostly settled to everyday spoken language much before systematic emotion research began. Ortony and Turner (1990) have reviewed 14 different theories of discrete emotions. One of the most popular theories of discrete emotions is the theory of six basic emotions (i.e. anger, disgust, fear, joy, sadness, and surprise) suggested by Paul Ekman and his colleagues (Ekman and Friesen, 1971; Ekman, 1982; Ekman, 1992). The basis for selection of the six basic emotions was that each emotion is associated with a universally recognizable facial expression. In some contexts, Ekman lists contempt as a seventh basic emotion.

Like Ekman's six basic emotions, many discrete emotions theories share the idea of a set of emotions, which are more basic (or primary, fundamental) than the other emotions. This division to basic and secondary emotions is usually explained based on physiological (basic emotions are hard-wired to the human physiology) or on behavioral basis (basic emotion are related to action tendencies or action readiness). Secondary emotions can be seen as combinations of basic emotions or social constructions within the basic emotions categories. For example, the physiological reactions associated with grief can only exist in association with the knowledge of a negative event meaningful to the person (e.g. a death of a close person).

Over the years, Ekman has developed a theory of characteristics, which define basic emotions. In Ekman (1999) he distinguished eleven such characteristics. The first three characteristics distinguish one basic emotion from another and the last eight characteristics distinguish emotions from moods, emotional traits, and emotional attitudes:

1) distinctive universal signals
2) distinctive physiology
3) distinctive universals in antecedent events
4) presence in other primates
5) coherence among emotional response
6) quick onset

7) brief duration

8) automatic appraisal

9) unbidden occurrence

10) distinctive thoughts, memories, images, and

11) distinctive subjective experience.

2.2.2 DIMENSIONAL EMOTIONS

In the dimensional theory of emotions, it is assumed that emotions and emotional experiences can be organized on a number of dimensions. The dimensional emotions viewpoint avoids some problems inherent to the discrete emotions viewpoint. For example, using the dimensional emotions viewpoint, emotional experiences can be evaluated without obeying the boundaries of certain discrete emotions. The dimensional emotions viewpoint is closely connected to a research method called the semantic differential first proposed by Osgood (1952). Using the semantic differential method, the subjects rate different stimuli, for example, words or events on bipolar scales, which consist of two opposite adjective pairs.

Empirical research has shown that people's experiential ratings about differences in affective meaning among stimuli like words, objects, and events can be succinctly described by only three basic dimensions. The notion of three basic dimensions originates in the work of Wundt (1896), it was empirically validated by Osgood (1952) using factor analyses, and further evidence was obtained by Mehrabian and Russell (1974). Their method consisted of 18 bipolar adjective pairs, six for each dimension, to assess the three-dimensional structure of objects, events, and situations. For example, the valence scale contained the adjective pair unhappy – happy, while the arousal contained the adjective pair calm – excited. However, factor analyses confirmed that it is sufficient to use only three dimensions, which account for much of the co-variance in the ratings.

As it is now, the most often used dimensional emotions framework has been presented by Lang and his co-workers (e.g. Bradley and Lang, 1994). That framework defines three different affective dimensions:

- *valence*, which ranges from negative to positive emotion, and it is neutral at the center of the dimension,
- *arousal*, which ranges from calm to highly aroused emotion, and it is also neutral at the center of the dimension, and
- *dominance*, which is related to the degree in which a person feels she/he dominates the current situation. This scale ranges from 'in control of the situation' to 'controlled by the situation', and it is also neutral at the center of the dimension.

In contemporary research, it is usual to use only the fundamental three dimensions in subjective evaluations of experienced emotion. Out of these three dimensions, valence and arousal are the most frequently used dimensions to capture the nature of emotional information. Lang and his co-workers have suggested that the valence dimension reflects the presence of an appetitive or aversive motive systems linked to behavioral tendencies of approach and withdrawal, respectively. The arousal dimension reflects the intensity of either the appetitive or the aversive system (Lang *et al.*, 1993; Lang, 1995; Bradley and Lang, 2000).

Examples of where some emotion categories might be positioned in the dimensional emotional space are shown in Figure 2 (Lang, 1995). The suggested positions were obtained by displaying affective pictures to subjects and asking their subjective emotional experiences after each picture. For example, after a picture with fear content, a snake in the grass, the subjects rated their subjective experiences as rather negative and rather highly aroused on average.

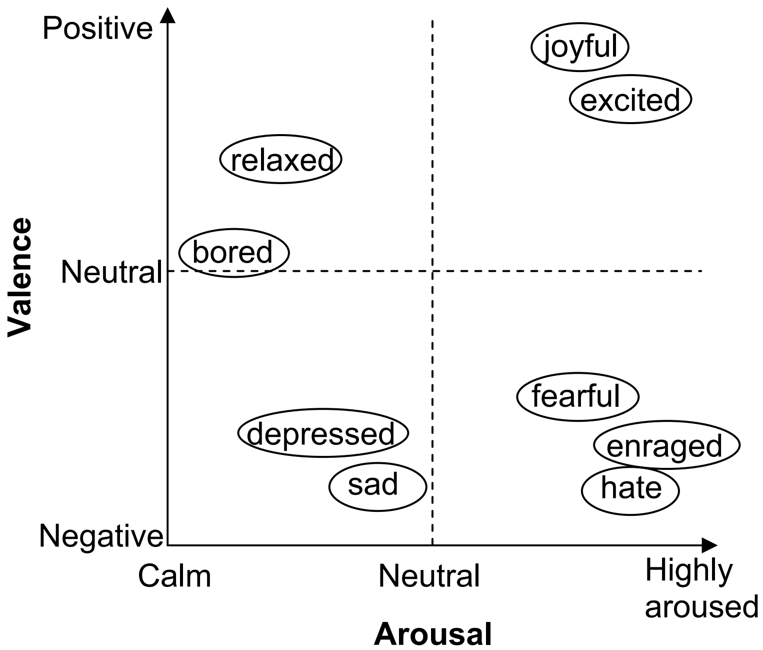

Figure 2. *Some discrete emotional categories in the dimensional space constituted by emotional valence and arousal as suggested by Lang (1995).*

Based on the semantic differential method, Bradley and Lang (1994) have developed a culture and language independent pictorial instrument called the Self-Assessment Manikin (SAM), in which a picture represents each point in the evaluation dimensions. For example, a smiling manikin represents a very positive experience, a manikin with a represents a very aroused experience, and a large sized manikin represents an experience high in dominance. This method has been used in many studies by Peter Lang and his colleagues (e.g. Lang *et al.*, 1993; Bradley and Lang, 2000).

Typically, an odd-pointed scale is used, so that it is also possible to give a neutral evaluation by using the middle point of the dimension.

In the research of the current thesis, the dimensional representation was chosen as the basis for experimental design. The dimensional emotions framework has some advantages over the discrete emotions framework in the studies presented in this thesis. When people are asked to report emotions using a discrete emotions approach, a forced-choice response format is often used, which oversimplifies the subjects' experiences. Alternatively, if the subjects are allowed to describe their emotions freely, they often report experiences of many different emotions (Picard, 1997). In contrast, by using a dimensional approach, a more unambiguous estimate of the subjects' subjective emotions can be obtained as a point in the space constituted by the affective dimensions. Furthermore, collections of auditory and pictorial stimuli were available (Lang *et al.*, 1995; Bradley and Lang, 1999), which were well studied on the experiential dimensions of valence, arousal, and dominance. By using these stimuli, it was possible to construct studies with stimuli systematically varying on the emotional dimensions.

As seen before, the semantic differential method is highly suitable for studying emotional experiences. In the experiments described in this thesis, the subjects rated their subjective affective experiences on bipolar scales of valence and arousal (and dominance in paper 6). Thus, our system for studying experiential ratings was similar to that of SAM, but instead of pictorial representations we used ordinary bipolar scales from 1 to 9, with adjectives related to the affective dimensions on each pole and the middle point (e.g. "negative", "neutral", and "positive"). These scales were shown to the subjects on the computer screen, and the subjects gave the ratings using the computer mouse or keyboard.

2.2.3 THE PSYCHOPHYSIOLOGY OF EMOTIONS

The human psychophysiological system for processing emotional information involves the entire human nervous system. It consists of the central nervous system (the brain and the spinal cord) and the peripheral nervous system, which consists of the sensory-somatic nervous system and the autonomic nervous system. The sensory-somatic nervous system moves our limbs by controlling the muscular system and receives information from our senses (e.g via the skin). However, the most important parts of the nervous system related to emotional processing are the central nervous system and the autonomic nervous system.

In the central nervous system, sensory input from the environment is received and processed by the thalamus, which passes the information on to the limbic system and the cortex. The limbic system, which is often described as the ‘seat’ of emotions, consists of the hypothalamus, the hippocampus, and the amygdala. There is evidence from studies of fear and fear processing that the amygdala is in a critical role in the processing of emotional information (LeDoux, 2000).

The autonomic nervous system plays an important role in regulating emotions in the body. The autonomic nervous system transmits the impulses from the central nervous system (especially the limbic system) to the peripheral organs. It is mostly involuntary. It controls, among other things, the heart rate, dilations and constrictions of blood vessels, dilations and constrictions of the pupil, and air flow in the lungs. The two main parts of the autonomic nervous system are the sympathetic nervous system and the parasympathetic nervous system. The sympathetic system enables the body to be prepared for fear, flight or fight. Sympathetic responses include, for example, an increase in heart rate, blood pressure and cardiac output, and increased pupil size. The other part of the autonomic nervous system is called the parasympathetic nervous system. It has its roots in the brainstem and in the spinal cord of the lower back. Its function is to bring the body back from the emergency status that the sympathetic nervous system puts it into. In physiological terms, the parasympathetic system is concerned with conservation and restoration of energy, as it causes, for example, a reduction in heart rate and blood pressure.

According to Lang (1995), emotions are driven by two opposite motive systems in the brain, the appetitive system, which is prototypically expressed by behavioral approach, and the aversive system, which is prototypically expressed by behavioral escape and avoidance. The activations of these motive systems are associated with positive (the appetitive system) and negative (the aversive system) experiences of affective valence. This is the basis for the primacy of the valence dimension in organizing emotion-related physiology and experiences. In this view, arousal reflects variation in the metabolic and neural activation of either or both systems.

An important question in understanding emotions is, whether the emotion-related physiological reactions are different for each emotion. According to the James-Lange theory (James, 1884), an event causes physiological arousal first and then we interpret this arousal. Only after our interpretation of the arousal can we experience emotion. This requires that each emotion has a unique autonomic signature. According to the Cannon-Bard theory (e.g. Cannon, 1927), physiological arousal and emotional experience occur simultaneously. Cannon and Bard suggested that the emotion-related physiological arousal is similar for many different emotions. There is, however, currently evidence from empirical research that autonomic differences among emotions exist. For example, anger is associated with higher heart rate acceleration than happiness, and with higher finger temperature than fear (Levenson, 1992; Levenson *et al.*, 1990).

2.2.4 THE ROLE OF EMOTIONS IN HUMAN THINKING AND BEHAVIOR

Modern research stresses the importance of emotions in human thinking and behavior. Emotions have been found to be in central role in rational decision making, perception, learning and other cognitive functions (Simon, 1967; Izard, 1993). There is evidence that emotion can have primacy over cognition (Zajonc, 1980; Murphy and Zajonc, 1993). In their studies, Murphy and Zajonc showed that emotional responses could occur without awareness so that the subject cannot even remember the stimulus that caused the emotional response.

It has been found that emotions greatly influence rational decision-making. At low and moderate levels of intensity, emotions can enter into decision processes and play an advisory role in decision-making. For example, if the present affective valence is positive, the decision maker tends to evaluate the decision-related options relatively positively (Loewenstein and Lerner, 2003).

According to Isen (2000) even mild positive emotion can lead to improved decision making and problem solving performance. However, too much or too little emotion can impair decision-making. When emotions are very intense, they can overwhelm cognitive processing and decision making altogether (Loewenstein and Lerner, 2003). Too little emotion can also cause severe problems. For example, Damasio (1994) found that patients with damages in the emotion processing areas of the brain had significant problems in rational decision-making. Limited emotional expressivity or expression recognition can also be a problem. Ekman (1999) wrote that his patients with facial paralysis (inability to produce facial expressions) had great difficulties in developing and maintaining even casual relationships. Similarly, Ross (1981) reported that stroke patients with specific problems in either expressing or recognizing emotional speech prosody, have severe interpersonal difficulties.

There is also evidence that emotion influences memory. Bradley *et al.* (1992) showed that long-term memory performance is affected by the level of emotional arousal. In their experiment, either positive or negative highly arousing pictures were remembered better than low-arousing ones. Emotions also drive attention (Öhman *et al.*, 2001) and play a significant role in learning (Sylwester, 1994).

As seen above, there is evidence that emotions play a central role in human rational thinking. Thus, it is very likely that emotions are also central in human-computer interaction. When the users are interacting with computers, emotions modulate the important cognitive functions during computer use: perception, attention, problem solving, memory, and learning. Perhaps even more importantly, however, the users react affectively to events in human-computer interaction. In the next chapter, a review will be presented on the ways of how emotion can be used in human-computer interaction in order to enhance the quality of interaction on an affective level.

3 Affective human-computer interaction

The interest in studying human-computer interaction on an affective level began during the 1990's, as computers became fast enough to possess the processing capabilities required by the affective technology. Most of the early practical research was carried out at Massachusetts Institute of Technology (MIT) by Rosalind Picard. In her influential book, 'Affective computing', Picard (1997) defined the field of affective computing very broadly as 'computing that relates to, arises from, or deliberately influences emotions'. Picard and Klein (2002) have suggested that affective computing includes taking the users' emotional needs and emotional skill needs into account. The definition of affective computing has also raised some criticism. Hollnagel (2003) pointed out some problems in the concept of affective computing. His main criticism was that computing by its very definition cannot be affective. Indeed, it seems that the theoretical foundations of affective computing are not as developed as affective computing research and applications in practice. Therefore, I refer to affective human-computer interaction in this thesis, when I discuss the research area, which develops methods and studies the phenomena related to emotions and computers.

This chapter presents an overview of the existing work in the field of affective human-computer interaction. First, a review of different affective measures and communication channels is presented. The emphasis is on pupil size variation, facial expressions, and speech, which are studied in the empirical experiments of the current thesis. Second, a look at the current state of the technology for affective human-computer interaction is presented. Finally, the Computers Are Social Actors (CASA) research paradigm is introduced and regarded as useful in the design of affective human-computer interaction. Two topics especially related to the current thesis, machine emotional intelligence and proxemics, are also discussed.

3.1 AFFECTIVE MEASURES AND COMMUNICATION CHANNELS

In order to develop affective human-computer interaction, the computer has to be able to get information about the user's affective state. To accomplish this, there are two possibilities. Either the user has to explicitly and voluntarily express information about his/her emotions to the computer, for example, using natural language or the computer has to be able to recognize the user's affective state by automatically measuring emotion-related physiological and behavioral changes. In human-human communication, affect is communicated both voluntarily and involuntarily. On one hand, people can communicate affect explicitly, for example, using the wordings and/or prosody of the speech, or by using deliberate facial expressions and gestures. On the other hand, people also communicate affect involuntarily in speech prosody and facial expressions, and, for example, in body postures.

The affective channels that people use in everyday human-human communication form the basis for affective human-computer interaction. In typical human-human communication, people use, for example, other peoples' facial expressions, speech prosody, and the words in the spoken speech as cues for emotional information. A computer can be designed to imitate human perception of emotions from other humans. However, it is also possible that computer recognition of emotions could be different from that of humans. Importantly, a computer can also potentially utilize a wide range of affective information channels that are unperceivable to humans, or humans can perceive information from these channels only with a limited accuracy, in limited situations or the information has only little affective meaning in communication. The measures reviewed in this thesis include pupil size, respiration, heart rate, skin conductance, skin temperature, blood volume pressure, and brain activity. Measures, which have a greater impact in human-human communication, include facial expressions, speech prosody, speech wording, eye movements, postures, and gestures. In Figure 3, these measures have been organized by two factors: significance in affective human-human communication and voluntariness. The basis for this organization has been a review of existing research on affective measures, which is presented in more detail in subsections 3.1.1 – 3.1.4.

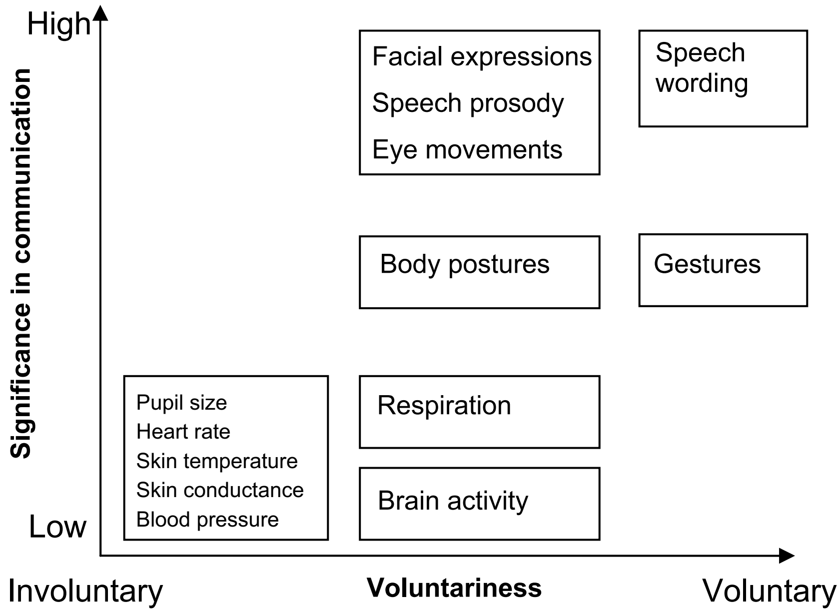

Figure 3. *Affective measures arranged by the degree of voluntary control over them and their significance in human-human communication.*

Even though affective computing studies are still relatively sparse, the relationship of various physiological measurements and emotions has been studied extensively in the field of psychophysiology. These studies, together with the findings that emotions are central in human thinking, form the scientific basis for the affective human-computer interaction discipline. What affective signals are potentially useful in computer input and what signals are potentially useful in output, then? Of the measures presented in Figure 3, the communication channels that are of central importance of affective human-human communication are likely to be well suited for both affective input and output, since humans have an innate ability to express emotions using those channels and perceive emotions of others expressed using those channels. Importantly, however, the computer could also utilize signals that are not of central importance in human-human communication. These signals offer interesting unexplored possibilities for human-computer interaction.

Pantic and Rothcrantz (2003) have created a taxonomy of the automatic human affect analysis problem domain. They identified four important questions in automatic analysis of human emotions:

1. Which channels of information, corresponding to which human communicative channels, should be integrated into an automatic affect analyzer?
2. How should the data carried by multiple channels be fused to achieve a human-like performance in recognizing affective states?

3. How should temporal aspects of the information carried by multiple channels be handled?
4. How can automatic human affect analyzers be made more sensitive to the context in which they operate (i.e., to the current situation and user)?

The current thesis focuses mainly on two affective input signals: pupil size variation, and the activations of facial muscles. The focus is to give new information on how robustly user emotions can be estimated based on these signals. Consequently, the main contribution of this thesis is related to question number one above. An adaptive and a person-independent affect estimator are compared in papers four and five, which will also give new information related to question four above. In addition, issues related to the temporal aspects of signals and context sensitivity will be discussed. In computer output, this thesis focuses especially on the affective effects of synthetic speech. Thus, the most important modalities from the viewpoint of this thesis, pupil size, facial expressions and speech, are introduced in their own subchapters below. Other potential affective computer input signals are presented after them in their own subchapter.

3.1.1 PUPIL SIZE VARIATION

The pupil is the opening at the center of the iris of the eye. Changes in pupil size are controlled by two smooth muscles in the iris: *sphincter pupillae* and *dilator pupillae*, which control pupillary constrictions and dilations, respectively. It is known that the autonomic nervous system controls the pupil dilations and constrictions, for example, to adjust the amount of light entering the eye (Andreassi, 1995). Previous studies have suggested that variations in pupil size are related to both cognitive and affective information processing. For example, it has been shown that the average pupil size increases with higher cognitive loads (Kahneman and Beatty, 1966; Hyönä *et al.*, 1995). In spite of considerable development in eye-tracking technology (e.g. ease of use, improved accuracy, and enhanced sampling rate), recent studies have been infrequent. It has been proposed that pupil size variation could be a potential computer input signal (Picard, 1997). However, most of the existing studies on the relation of pupil size and emotions are from the 1960's or 1970's, which highlights the need for new research with modern equipment.

While the pupil size is not in a central role in human-human communication, there is some evidence that pupil size is involved. It has been shown that pupil size affects males' judgments of females on a negative-positive scale. Females with larger pupils evoked more positive feelings in males (Hess and Petrovitch, 1987). It has also been found that pupil size variation is especially related to communication between the different sexes. For example, when subjects were shown facial stimuli of the opposite sex, the pupil sizes of the subjects increased with bigger pupil sizes in the stimuli (Simms, 1967). Early studies also focused on sexual arousal indicating pupil dilation to sexually interesting pictures (Hess and Petrovich, 1987).

The existing research on the relation of affective processing and pupil size variation has been somewhat controversial. Hess (1972) suggested that the pupillary responses range from extreme dilations to interesting and pleasing stimuli to extreme constrictions to unpleasant or distasteful material. Janisse (1974) challenged this model and argued that there is no pupil constriction to negative stimuli, or the constriction response may be limited to a few individuals and a small range of stimuli. Earlier Loewenfeld (1966) had also studied the effects of various sensory and psychological stimuli to pupil size variation and argued that none of them caused pupil constriction except increased light intensities. Janisse (1974) suggested that pupil size is linearly related to the intensity dimension of the stimuli. According to him, pupil size variation behaves curvilinearly on the valence scale. This means that it is largest at the negative and positive ends of the continuum and smallest at the center, which represents neutral affect.

When evaluating the previous results on pupil size and emotions, one has to notice that the stimulus materials used in the different studies have varied considerably. Mostly, they have been limited sets of pictures varying in content. The materials used have also suffered from methodological problems with color, luminance, and contrast (Hess and Petrovich, 1987). This might partly explain the controversial results and theories. Based on the earlier results, it seems that there is a need for new studies on pupil size variation and emotions in order to understand the use potentials of pupil size variation in affective human-computer interaction. Furthermore, controlled stimulus sets also seem to be a necessary precondition for the systematic study on the effects of emotions to pupil size variation.

3.1.2 FACIAL EXPRESSIONS

Facial expressions are the result of movements of facial skin and connective tissue (i.e., fascia) caused by the contraction of one or more of the 44 bilaterally symmetrical facial muscles (Rinn, 1984). It has been suggested and widely accepted that facial expressions are evolutionary adaptations (Schmidt and Cohn, 2001). Consequently, facial expressions have also been found to have an important meaning in human-human interaction (e.g. Surakka and Hietanen, 1998). Facial expressions can mediate emotions. For example, the six basic emotions (anger, disgust, fear, joy, sadness, and surprise) suggested by Ekman (1992) are related to distinct facial expressions, which are universally recognized. However, it is also known that facial expressions can be produced voluntarily and thus they can be masked and faked in certain circumstances (Ekman and Friesen, 1982; Ekman, 1985; Surakka and Hietanen, 1998). Facial expressions are also used for conveying certain nonemotional messages, for example, emblems such as a wink, or illustrators such as a raised brow during speech (Ekman and Friesen, 1969). It is also known that facial expressions are contagious, i.e. people show tendency to both experience and express other person's emotions inferred from facial expression (Lundqvist and Dimberg, 1995; Hietanen *et al.*, 1998; Surakka and Hietanen, 1998).

Many facial expressions, for example, the expressions associated with the basic emotions suggested by Ekman (1992) involve the movement of many facial muscles. The most usual standard for describing facial expressions is the Facial Action Coding System (FACS) developed by Ekman and Friesen (1978). It divides the human face currently into 46 facial actions based on visually observable individual changes in the face, and head and eye movements. In the light of affective human-computer interaction, however, instead of tracking the facial muscle activations in the whole face, it would be easier and more convenient to start by tracking movements of only a few muscles, which are strongly related with emotional experience. There is substantial evidence that two facial muscles relatively strongly related to the experience of affective valence are *zygomaticus major* (the muscle that draws the lip corners up producing a smile) and *corrugator supercilii* (the muscle that knits and lowers the brows producing a frown). The activations of these muscles can be studied, for example, using electromyography (EMG), which is a method for studying the electrical activity of muscles typically using surface electrode technology. Dimberg (1990) found that *corrugator supercilii* EMG activity increased in response to angry facial stimuli, and *zygomaticus major* EMG activity increased in response to happy facial stimuli. Greenwald *et al.* (1989) used affective pictures as stimuli and found a significant negative linear covariation of *corrugator supercilii* EMG activity and ratings of experienced valence: negative experiences were associated with high *corrugator supercilii* activity and positive experiences were associated with low *corrugator supercilii* activity. Their data also showed the presence of a linear relation between *zygomaticus major* activity and ratings of valence, even though a quadratic relationship was stronger for this particular muscle. Specifically, *zygomaticus major* activity increased, when experienced valence got more positive, with the exception that in response to very negative experiences, the *zygomaticus major* activity was higher than for neutral and subtly negative and positive experiences. In addition, a score based on subtracting the *corrugator supercilii* EMG activity from *zygomaticus major* EMG activity showed a positive linear correlation and an even stronger quadratic trend.

A study, which also used affective pictures as stimuli, was presented by Lang *et al.* (1993). The pictures were taken from the International Affective Picture System (IAPS), which is a collection of universally recognized affective pictures categorized by their effects on experienced valence, arousal, and dominance (Lang *et al.*, 1995; Ito *et al.*, 1998). Their results showed a negative linear trend between *corrugator supercilii* EMG activity and ratings of experienced valence, and a positive linear trend between *zygomaticus major* EMG activity and ratings of experienced valence. A negative linear relationship between *corrugator supercilii* EMG activity and ratings of experienced valence was also found using affective sounds as stimuli (Bradley and Lang, 2000). They used sounds from the International Affective Digitized Sounds (IADS), which is a collection of universally recognized affective sounds well studied on the experiential dimensions of valence, arousal, and dominance (Bradley and Lang, 1999). Recently, Larsen *et al.* (2003) found a negative linear correlation between *corrugator supecilii* EMG activity and valence ratings for affective IAPS pictures and IADS sounds for female subjects. They also found a weaker linear correlation between *zygomaticus major* EMG activity and valence ratings. Taken together, these studies give convincing evidence for the association of emotions and these two muscles. Based on the information provided by these studies, it seems that these muscles could be useful in affective human-computer interaction.

Although there is evidence for universal facial expressions of certain emotions, it is important to realize that there are also differences in the facial behavior of different people. For example, Ekman (1985) stated that accurate interpretation of facial expression benefits from the knowledge of what is normative for the specific individual. In concordance with this, Cohn *et al.* (2002) found evidence for relatively stable individual differences in facial expressions. There was a correlation of .58 in *zygomaticus major* EMG responses to the same emotionally positive film clips between two sessions with a 12 months in-between measurement interval.

As stated above, it is known that under certain condition, facial expressions can be masked and faked by using voluntary control over them (Ekman and Friesen, 1969; Ekman, 1985; Surakka and Hietanen, 1998) according to display rules appropriate for each situation. Voluntary facial expressions have also been used as pointing methods in human-computer interaction (Partala *et al.*, 2001; Surakka *et al.*, 2004). However, facial expressions in *zygomaticus major* and *corrugator supercilii* regions are not totally under voluntary control either. For example, Kappas *et al.* (2000) have shown that when the subjects were shown humorous positive movie clips and tried to inhibit any facial expressions, they still showed significantly more *zygomaticus major EMG* activity and less *corrugator supercilii* EMG activity as compared to a neutral control task. Their results suggest that voluntary control cannot entirely mask the spontaneous facial activation to positive stimuli. When these results are reflected upon in the light of human-computer interaction, they imply that in addition to voluntary affect expressions, spontaneous affect could also be measured from activations of facial muscles in affective human-computer interaction.

While *zygomaticus major* and *corrugator supercilii* seem to the most prominent facial muscles for affective computing in the light of previous research, and especially from the viewpoint of dimensional emotions, there are also other alternatives. Healey (2000) has also suggested the *masseter* muscle, which is activated in jaw clenching, as a potential affective input signal. In her experiment among eight self-induced emotions, anger expressions resulted in clearly elevated levels of the *masseter* muscle activity, as compared to the other categories. Other emotion-related facial muscles include, for example, the *levator* muscle, which is activated in response to disgust (Vrana, 1993; Andreassi, 1995). In addition, based on existing research (e.g. Ekman, 1985; Surakka and Hietanen, 1998) genuine (felt) smiles could be distinguished from faked (unfelt) smiles by tracking the activations of the *orbicularis oculi* muscle (the periocular muscle that pulls the outer corners of the eyebrows slightly downwards) in conjunction with the *zygomaticus major* associated with smiling.

Thus, there are many possible muscles and combination of muscles, whose activations can potentially provide useful information about the user's emotions during human-computer interaction. Because of the substantial evidence that *zygomaticus major* and the *corrugator supercilii* muscles are correlated with affective valence, we decided to measure the activations of those muscles in our early experiments on facial expression in the context of affective human-computer interaction presented in this thesis.

3.1.3 SPEECH

In human communication, the speech system is specialized for the rapid transfer of information (Mattingly and Liberman, 1988). Speech has a tremendous meaning in human-human communication, and it has also been widely studied in human-computer interaction. Speech is also an important channel for emotional communication between humans, and based on that information alone it is very conceivable that speech modality could also be of major significance in affective human-computer interaction.

Speech can be used in both computer input and output in a variety of ways. In computer input, the computer can recognize three kinds of information based on the acoustic pattern of the speech: who the speaker is, what the speaker said, and how the speaker said it (Breazeal and Aryananda, 2002). It is possible to recognize the identity of the speaker based on automatic analysis of speech signals (Campbell and Reynolds, 1999). Furthermore, in the field of affective computing, speaker emotions can be estimated based on the affective wordings of the spoken message or the prosodic features of the speech. Prosody means the non-linguistic aspects of speech, and prosodic features of the speech include, for example, pitch, stress, rhythm, articulation and speech rate. Speech recognition systems can recognize the wording of spoken messages at rates of about 90% in ideal conditions. However, in speech recognition with current systems the recognition rates can drop with emotional speech as compared to neutrally spoken speech (Picard *et al.*, 2001).

Emotions can be automatically recognized from the prosody of the speech with accuracy comparable to that of humans. For example, Banse and Scherer (1996) were able to automatically distinguish between 14 different emotion-related categories (hot anger, cold anger, panic fear, anxiety, despair, sadness, elation, happiness, interest, boredom, shame, pride, disgust, contempt) based on a variety of acoustic speech parameters in acted speech with average hit rates of 40-53%. Their automated methods approached or in some cases even outperformed the ability of human judges, which was slightly less than 50% on average. Recently, Oudeyer (2003) showed that with a large speech database and advanced data mining and learning algorithms, it was even possible to achieve recognition accuracies of more than 95% between four categories: joy, sadness, anger, and neutrality.

Speech can also be used effectively in computer output. There is evidence that by varying certain speech characteristics, easily distinguishable computer personalities can be created solely based on speech (Nass and Lee, 2001). What the computer says also has important effects. For example, Aula and Surakka (2002) found that using negative, neutral, and positive worded synthetic speech feedback to simple mathematical tasks it was possible to evoke different emotional experiences in the subjects. The feedbacks also produced differential autonomic nervous system activity as measured in terms of pupil size. Positive feedback also provoked shorter response times than negative feedback.

The prosody of computer-generated speech can also be used to mediate emotion-related information to the user. Among the first, Cahn (1990) generated distinct synthetic speech prosodies for the six basic emotions (anger, disgust, fear, happiness, sadness and surprise) suggested by Ekman (1992). She created the different prosodies by altering four prosodic features of speech: pitch, timing, voice quality and articulation. In her study, the subjects recognized the intended emotion among the six basic emotions based on only prosodic information at a rate of slightly over 50% on average. In a recent study by Breazeal (2003a), the subjects were able to distinguish between the six basic emotions in lip-synchronized synthetic robot speech at average rates of about 60%. These results were quite promising considering that in a similar study with real human voices expressing the six basic emotions, the subjects were only about 60% accurate in recognizing the correct emotion (Scherer *et al.*, 1991).

In sum, there are two main affective communication channels in computer-generated speech, the wording of the messages and the prosody of the speech. These two have been sometimes studied together in the same experiments. For example, Scherer *et al.* (1984) conclude that emotions can be recognized from the contents (i.e. wording) of spoken messages, from the emotion-related prosodic cues or by combining the two. In addition, Nass *et al.* (2001) found that emotion conveyed in synthetic and recorded speech affects perceptions of emotional valence, suitability, liking, and credibility of the content of the messages. In practice, however, it is often not desirable or possible to study both affective speech prosody and the semantic contents of the message in the same study. Consequently, it is typical that either speech prosody or the affective contents of the speech is kept as neutral as possible, while the other factor is systematically varied as an independent variable by the researcher.

3.1.4 OTHER MEASURES

In addition to the three affective measures discussed above, there are also several other means both for measuring affect-related information from the user and for presenting affective information from the computer to the user. An overview of measures, which have been suggested to be related to emotions, is presented in the next few pages in order to give the reader an understanding of all communication channels applicable to affective human-computer interaction.

Blood pressure. The blood pressure signal is an indicator of blood flow. Blood pressure increases with negative emotions such as fear and anxiety, and decreases with relaxation (Picard, 1997; Healey, 2000). In an early experiment, Fernandez and Picard (1998) were able to develop methods based on blood volume pressure and galvanic skin response to achieve better than random recognition of frustration during a game-playing situation.

Brain activity. The electrical activity of the brain can be measured, for example, with electroencephalography (EEG) using electrodes attached to the scalp. It has been found, for example, that EEG asymmetries over the frontal cortex during emotions related to the behavioral tendency of approach (e.g. joy, interest, and anger) are relatively greater in the left prefrontal cortex than the right prefrontal cortex. Correspondingly, it has been suggested that during emotions related to the behavioral tendencies of withdrawal (e.g. sadness, fear and disgust) EEG asymmetries are relatively greater in the right prefrontal cortex, even though the effect is somewhat less clear (Coan *et al.*, 2001). EEG has also been used in computerized emotion recognition. Takahashi (2004) found that using EEG measures from the frontal lobe, it was possible to distinguish between five self-induced emotions with rates over 40%, which is much better than chance level (20%). In that study, the recognition rates for EEG also outperformed those of pulse and skin conductance.

Galvanic skin response. Galvanic skin response is an indicator of skin conductance. Skin conductance measures changes in the resistance of the skin caused when glands on the skin produce ionic sweat (Healey, 2000). Skin conductance is one common measure in psychophysiology and reflects the ability of the skin to conduct electrical current. Skin conductance has been shown to correlate with autonomic nervous system arousal so that an increase in affective arousal causes an increase in skin conductance (Lang, 1995; Codispoti *et al.*, 2001). Importantly, skin conductance has also been chosen to be applied in practice in some early prototypes of affective systems, described later in chapter 3.2 of this thesis.

Gaze direction. Fukayama *et al.* (2002) studied systematically the effects of the gaze direction of an embodied interface agent to the users' impressions by varying the gaze of an agent in gaze direction and amount of eye contact. They used the semantic differential method to assess the users' impressions of the agent on two main factors, friendliness (like/dislike) and dominance. Their results showed that the amount of eye contact between the user and the agent and the agent's gaze direction when not in eye contact significantly affected the users' ratings related to like/dislike and dominance of the agent. Calvo and Lang (2004) studied gaze patterns when their subjects were looking at emotionally neutral, positive, and negative pictures paired with simultaneously presented nonemotional control pictures. They found out that emotionally positive or negative pictures were fixated significantly longer and more often with the first fixation, as compared to emotionally neutral pictures. On the basis of these two studies, eye movements could be tracked from the user during human-computer interaction, and they could be useful in signaling affect-related attentive selection and/or communicative aspects in both input and output.

Heart rate. The results on the relationship of heart rate and emotions are not unanimous. It has been suggested that heart rate is especially sensitive to experimental task parameters (Lang *et al.*, 1990). However, it has been found in many studies that heart rate can differentiate between positive and negative emotional responses (Lang *et al.*, 1993; Anttonen and Surakka, 2005) or between specific emotions (Levenson *et al.*, 1990). Some studies have found differences as a function of affective arousal instead of valence (Bradley and Lang, 2000). In some studies average heart rate acceleration responses to affective stimuli have been reported (Lang *et al.*, 1993), and in some studies average decelerations have been reported (Bradley and Lang, 2000; Anttonen and Surakka, 2005). Levenson *et al.* (1990) reported average heart rate accelerations for anger, fear and sadness and decelerations for disgust. Heart rate has been used in practice, for example, by Healey and Picard (2005), who found that heart rate can be used as a relatively good indicator of drivers' stress in traffic.

Muscle activations. It is also possible to use electromyography (EMG) to measure the emotion-related activations of other than facial muscles. For example, Healey (2000) recognized car drivers' stress by measuring upper back tension from the *trapezius* muscle. Her results showed that it was possible to distinguish between rest, low stress highway driving and stressful city driving at an average rate of almost 70% based on *trapezius* EMG alone.

Posture. Postural information could be used in affective computing both in input and in output. Mota (2002) could distinguish between nine naturally occurring user body postures with a rate of 87.6%. Her results suggest that five affect-related states can be found in these postures: high interest, interest, low interest, boredom, and taking a break. In computer output, Coulson (2004) studied how well his subjects could recognize the six basic emotions from static body postures of a computer-generated agent viewed from three different angles. The results indicated that the subjects could recognize affect from the agent's postures even with rates of over 90% for some anger, happiness, and sadness postures. In contrast, the recognition rates for disgust, fear and surprise were below 50%.

Respiration. Emotional arousal is associated with faster and deeper respiration as compared to rest and relaxation, which are associated with slower and shallower respiration. Negative emotions can also cause irregularities in respiration (Frijda, 1986). Healey (2000) used six derivatives of respiration signals successfully in an experiment to distinguish between seven self-induced emotions (anger, hate, grief, love, romantic love, joy, reverence) and no emotion. Together with five other measures derived from blood volume pulse, skin conductance, and *masseter* EMG, she could recognize the correct emotion among the eight emotions with an accuracy of 81%.

Skin temperature. There is some evidence that negative emotionally aroused states (e.g. anger) are related to elevated skin temperature, as compared to calmer and more positive emotional states (McFarland, 1985).

In addition to this, Levenson *et al.* (1990) found that anger produces significantly larger finger temperature responses than fear.

Tactile information. There are many interaction techniques, which use the tactile modality in human-computer interaction in both input and output (e.g. touch screens and force-feedback output devices). However, the number of tactile techniques designed especially for affective interaction is small. Qi *et al.* (2001) have developed PressureMouse, which is a mouse with eight tactile sensors, which measure the pressure, with which the user is touching the mouse. Based on this information they could recognize between user frustration and non-frustration during computer use at a rate of 88%. Brave and Dahley (1997) described the inTouch device, which allows the remote users to communicate physically using a shared physical object consisting of cylindrical rollers. Physical contact is a basic means through which people achieve a sense of connection, indicate intention, and express emotion (Brave and Dahley, 1997).

Above I have listed affective measures, which seem to be potential candidates for practical affective human-computer interaction. It is likely that by combining different affective measures, enhanced reliability in estimating the user's emotional state could be obtained. In human-human communication, people use information from many previously mentioned communication channels (especially facial expressions and voice prosody) to infer the emotional state of other people. According to Pantic and Rothkrantz (2003) the computerized fusion of emotion-related signals similar to that of humans can be seen as the ultimate goal for a multimodal analyzer of human affective states. Such fusion can be accomplished at three levels: at data level, at feature level, or at decision level.

Some of the existing studies on affective signals encourage the fusion of multiple affective signals in affect analysis. For example, Collet *et al.* (1997) developed methods for differentiating emotions based on different measurements of autonomic nervous system correlates evoked by pictures of the six basic emotions suggested by Ekman (1992). Based on the recordings of six autonomic nervous system parameters - skin conductance, skin potential, skin resistance, skin blood flow, skin temperature and instantaneous respiratory frequency – they were able to find pairwise differences between all 15 pairs of basic emotions. Electrodermal responses alone differentiated between 14 out of 15 pairs, and thermo-vascular and respiratory responses alone differentiated between 13 out of 15 pairs of basic emotions.

Good results were also obtained recently by Lisetti and Nasoz (2004), who measured galvanic skin response, heart rate and skin temperature, while their subjects watched movie clips aimed at eliciting six target emotions: sadness, anger, surprise, fear, frustration, and amusement. They found out that by analyzing the signals with Marquardt backpropagation machine learning algorithms, their fusion system found the targeted emotion with an average accuracy of as high as 84.6%.

Based on the studies described above, it seems that there are many alternative measures, which might be used successfully in the future. However, it has to be noticed that there are also deficiencies associated with some of the early studies described above. Many of these problems are related to the set of emotions chosen to be used in the empirical studies, or to the ways that these emotions are induced. Frustration, interest, romantic love, and stress are examples of phenomena that are related to emotions, but are not regarded as primary emotions by most theorists. Many studies used voluntarily expressed or self-induced affective states, which can significantly differ from spontaneously felt emotions. In some cases, the results obtained have not been very good and in some cases they have been contradictory. Despite these problems with some studies, the results have been generally promising, and provide a basis for further research on affective human-computer interaction.

3.2 AFFECTIVE SYSTEMS

The ultimate goal of affective computing is creating computer systems, which recognize the user's emotions, and are capable of reacting in a way that would enhance the overall quality of interaction. Different researchers (e.g. Lisetti and Schiano, 2000; Picard *et al.*, 2001; Picard and Klein, 2002; McNeese, 2003) have suggested a variety of different application areas, which include:

- computer-assisted learning, distance learning, tele-teaching
- healthcare, tele-healthcare
- artificial intelligence, robotics
- computer-mediated human-human communication
- affective mirror, emotional skills trainer, human-human communication trainer
- automobile drivers' alertness/stress/emotions monitoring
- usability/product testing
- stress detectors, lie detectors
- ubiquitous computing
- wearable computing
- virtual environments and avatars, and
- entertainment and computer games.

As it is now, there are already some existing systems recognizing and utilizing affective information measured from the user. Examples of the current affect-sensing systems and devices include the affective learning companion (Burleson, 2004), which senses different affective signals of the learner including facial expression and eye movement information measured by a video camera, posture detection from seat pressure pads, and galvanic skin response. Based on the knowledge inferred from the affective signals, the system changes the behavior of the affective learning companion – a virtual character - in a way that is meaningful according to the learning goals.

Another class of existing affective system consists of innovative affective input devices, which visualize the user's affective responses to other people. An example of such a system are expression glasses (Scheirer *et al.*, 1999), which are a wearable device capable of recognizing and visualizing the user's interest and confusion level based on the activations of two facial muscles. In a study, the system was able to distinguish between interest and confusion with an accuracy of 74%. Using the device, the confusion and interest levels of the wearer can also be visualized using red and green lights, respectively. Another innovative input device developed at MIT is the Galvactivator (Picard and Scheirer, 2001). It is a glove-like hand-worn device that measures the wearer's skin conductance, which is linked to physiological arousal. To visualize elevated levels of arousal, the device starts glowing brightly. Another early input device measuring affective information was the emotion mouse, which measures the user's skin temperature, galvanic skin response, and heart rate (Ark *et al.*, 1999). In their design the affect recognition sensors are embedded on a mouse, which enables unobtrusive measurement.

As seen above, the first efforts for affective systems were quite input-oriented with advanced methods for sensing affective input and limited methods for affective feedback. This is, of course, understandable, because well-working sensing of affective signals is a prerequisite for affective computing in full. Studies in which the effects of real-time affective feedback given by the computer are studied are still sparse.

Klein *et al.* (2002) conducted an experiment in which the user was deliberately frustrated during a game-playing situation, and the user was given affective support using a text-based agent. Their results showed that the users chose to interact significantly longer with the computer when they were given textual affective support by the agent as compared to the condition in which no support was given. The use of textual affective support also led to significantly longer computer usage as compared to an opportunity to vent feelings using a text box. Based on these findings Klein *et al.* (2002) suggested that "computers are capable of alleviating strong, negative emotions, even when they are the source of these negative emotions".

Computer-mediated communication (CMC) is also a very potential application area of affective computing. In fact, there has been affective computer-mediated communication for quite a long time now, as the users worldwide have adopted a system of using smileys to communicate emotion and intent in text based media (e-mail, text chat, web pages etc.). The methods studied in the field of affective computing, however, offer a whole new repertoire of possibilities related to computer-mediated communication. An example of an actual system, which uses physiological information to enrich computer-mediated communication is the system described by Wang *et al.* (2004). Their system measures the users' galvanic skin response during online chatting and the computer interprets the user's autonomic arousal based on this signal. In addition, the chatters can label the messages with tags (e.g. "<happy>"), which indicate affective valence. Based on the valence and arousal information obtained this way, the computer uses animated text varying in size, speed and color to show the actual message in different affectively meaningful ways to the other chatter. In their study they found out that as compared to a chat with no emotional feedback the chatters became more involved in the conversation and got to know the emotional states of their partners better using the affective chat system.

Another interesting emotion-related application area is human-robot interaction. Robots are by definition usually anthropomorphic or animalmorphic, they usually can move around and possess some expressive functions such as talking. For them, traditional computer input methods (e.g. mouse) are not typically applicable and the human-robot interaction must essentially use methods from human-human communication. One of the most advanced affective robots to date called Kismet has been developed by Breazeal (2003b). It engages people in natural and expressive face-to-face interaction, as it can control its gaze and other facial features (e.g. eyelids, brows, lips, and ears). It has an emotion model, which combines the discrete and dimensional emotions theories so that its emotional state is represented by three dimensions (i.e. arousal, valence, stance). When the emotional state clearly enough reaches the region of a specific emotion, the corresponding emotion expression is activated. It can express nine specific emotions (content, unhappy, anger, fear, accepting, surprise, disgust, content, stern) by adjusting the facial features and the postures of the robot.

Another recent innovative affective system is SenToy, which is a tangible doll with sensors that allows a user to influence the emotions of a character in a game (Paiva *et al.*, 2003). SenToy is an input device that can be used to communicate anger, fear, surprise, sadness, gloating and happiness with gestures that the sensors inside the doll pick up. The sensors pick up acceleration and limb position information. The user's gestures using the doll affect the emotional state of a character in a 3-D game. Displaying the right emotion with the doll is a prerequisite for progress in the game. In an empirical study with child, teenage, and adult subjects the subjects did fairly well especially in expressing happiness, anger and sadness with the doll. The ratings for subjective satisfaction indicated that the design was successful, especially for children.

At the time of writing this thesis (spring 2005), first commercial affective applications are hitting the market. For example, the recently launched Siemens CX70 Emoty (2005) mobile phone is equipped with a keypad with ten emotional categories and integrated stroke, press and shake sensors. Using these sensors the users' emotions can be conveyed to animated 3D-characters, sharing affective multimedia messages with friends. It seems that pet robots are going to be the next applications systematically utilizing emotions, which are going to be commercially available and targeted to a large audience worldwide. The toys available in market seem to be like limited versions of Kismet by Breazeal (2003b). For example, Hasbro's Furby (2005) can express five emotions (happy, sad, surprised, scared, and sleepy) by movements of its flexible beak, expressive eyes, and movable ears and eyebrows.

An even wilder concept has been implemented by Toyota in co-operation with Sony, when they designed Toyota POD (2005), which is a car that can measure and respond to the driver's emotions and express its own emotions. This concept was presented for the first time as early as 2001. This prototype can measure the driver's pulse and perspiration, and try to relax the driver when needed by warning, playing soft music, or blowing cool air. Four emotions can be expressed using LEDs in the front panel, and the appearance of front lights, which look like eyes.

When designing affective applications, games and interactive entertainment are a potential application area. Ravaja *et al.* (2004) found that different games evoke differences in physiological activity and in terms of subjective self-reported emotion. Gilleade and Allanson (2003) contributed to the field by introducing a toolkit for affectively adaptive games. A prototype of an actual affective game has been presented by Sakurazawa *et al.* (2004). Their game used the skin conductance response as an input method to the game. Skin conductance determined the difficulty level of the game. By staying calm, the player got significantly fewer obstacles as compared to being agitated. About two thirds of the subjects found the game augmented with skin conductance more interesting than the version without it.

As seen above, the first commercial applications have been entertainment-centric, such as toys. The reason for this is probably that the state of the art in machine emotional intelligence is not yet advanced enough to fully simulate believable intelligent human-like characters. Nevertheless, affective human-computer interaction as a research paradigm ultimately aims at something more serious than entertainment. Many researchers agree with the optimistic vision of Picard (1997), who sees that emotions are a necessity for computers to become truly intelligent. Furthermore, the ability to respond to the user's needs on the affective level is seen as a relevant goal by many researchers (e.g. Picard and Klein, 2002). Understanding the affective component of human-computer interaction, together with the traditionally studied cognitive and behavioral aspects, also seems to be a necessity for scientists to obtain a holistic picture of interaction.

3.3 COMPUTERS ARE SOCIAL ACTORS

Computers are social actors (CASA) is a research and design paradigm developed by Clifford Nass and his colleagues (Nass *et al.*, 1994; Reeves and Nass, 1996). The main finding from the studies using this paradigm is that the users' interactions with computers are fundamentally social. Even though the users do not think that computer are humans or human-like, or that they are communicating with the programmer, they respond to socially to computer-generated probes. According to these writers, social responses to computers are commonplace and easy to generate. They have found many aspects, in which human-computer interaction closely resembles human-human communication. For example, the users express politeness towards computers (Nass, 2004), attribute personality to computers using minimal cues (Nass *et al.*, 1995; Nass and Lee, 2001), and collaborate with computers as teammates (Nass *et al.*, 1996).

The CASA paradigm seems also valuable in the design and research of affective human-computer interactions. In concordance with the CASA paradigm, Picard and Klein (2002) have suggested that human-human interaction should be a guide for building affective systems. According to them, thinking about the equivalent human-human interaction is a critical part of designing a successful system. Many products have failed to take this view into consideration. For example, a car that talks and commands the driver, but is not able to listen is analogous to a person that talks and commands, but never listens to others.

The current thesis is related to the CASA paradigm in two ways. First, in the experiment involving affective interventions (Paper III), the computerized interventions were designed according to what could be a useful human intervention in that situation. Thus, in that experiment the computer shows emotional intelligence similar to that of humans. We also used two different emotional intelligence strategies for human-computer interaction. Second, in paper VI, the affective effects of communicating with a conversational agent at different simulated proximities are studied. It is known that in human-human communication proximity has an important role in regulating a variety of different factors in communication. The experiment described in paper VI studied whether proximity can be simulated in human-computer interaction by varying the size of the conversational agent on the screen.

3.3.1 Machine emotional intelligence

Creating affective human-computer interaction is essentially about giving computers emotional intelligence skills similar to those of humans. Emotional intelligence has been defined as the ability to recognize, express and have emotions, to regulate these emotions, harness them for constructive purposes, and skillfully handle the emotions of others (Picard *et al.*, 2001). This definition also includes the main tasks in the process of building affective computers. The natural starting point is to build computerized emotion recognition as discussed in the last chapter, but an equally important question is how the computer should express emotion in order to be able to skillfully handle the user's emotions.

Picard and Klein (2002) suggested an approach, in which the computer should acknowledge and support the user's emotional needs, which can be divided into skill needs and experiential needs. Emotional skill needs include emotional self-awareness, managing emotions, self-motivation, affect perception, and empathy. Experiential needs include the need for attention, companionship, security, and the need to feel that one's emotional state and emotional responses are understood and accepted by others, and appropriate for the situation. In order to accomplish these goals, they suggested that a computer can act, for example, as an experiential emotional aid (try to affect the user's emotional experiences in a positive way), as a pre-emptive tool (try to prevent situations, which presumably cause frustration for the particular user), or as an emotional skill-building 'mirror' (try to provide feedback, which enhances the user's emotional skills, self-awareness, or management).

Empirical experiments in which different affective strategies have been compared during human-computer interaction are still rare. Brave *et al.* (2005) studied the effects of anthropomorphic agents, represented by photos with speech bubbles, during a computerized blackjack game. After each round, an agent expressed no emotion, self-oriented emotion (e.g. "I'm happy. I won this time."), empathic emotion (e.g. "I'm glad you won."), or both self-oriented and empathic emotion related to the outcome of the round (won/loss). Their subjects responded positively to the agents with empathic emotion, including greater likeability, trustworthiness, perceived caring, and felt support. On likeability, trustworthiness, and perceived caring, the presence of both self-oriented and emphatic emotion was rated even a little bit higher than emphatic emotion alone, but the difference was not significant. The results also indicated that an agent described as submissive received more positive ratings. Another experiment, which aims at affecting the user's experiences in a positive way during human-computer interaction, is described in paper III of this thesis.

3.3.2 PROXEMICS

The study of the way humans' personal space is structured is referred to as proxemics. It is a research area, which has been largely underestimated in modern psychological and human-computer interaction studies. Most of the groundbreaking work related to the meaning of proximity in communication has been done in the 1960's or 70's. Nevertheless, proximity – or interpersonal distance – is an important factor affecting human-human communications. To give some examples how proximity affects communication between two persons, it affects the duration of the interaction, as well as the emotional tenor of the interaction (Mehrabian, 1968a; Konecni *et al.*, 1975).

In a classic book, *The Hidden Dimension* (1966), Edward T. Hall illustrated how physical distances between people in face-to-face interaction reflected degrees of formality. He defined four specific proximity levels:

- *Intimate*: up to 18 inches (about 46cm),
- *Personal*: 18 inches to 4 feet (about 46cm - 122cm),
- *Social*: 4 to 12 feet (about 122cm - 366 cm), and
- *Public*: more than 12 feet (about 366cm)

The different proximity levels have a clear influence on the nature of communication. On the intimate level people often communicate with their intimates and closest friends and communication is often nonverbal. The personal level is the most usual level for one-to-one communication between individuals. Speech is the most commonly used method of communication on the personal proximity level. On the social level, people often communicate with many persons at the same time, for example, in a meeting. On the social level communication is more formal than on the intimate and personal levels. On the public level, only a little detail and nuances can be communicated. The communication is typically one-way (e.g. a person lecturing to the audience).

Reeves and Nass (1996) have dedicated a whole chapter of their award-winning book 'The Media Equation' to the issue of interpersonal distance. They describe an experiment, in which their subjects were shown pictures of people in different sizes, from different distances, and both close shots and large shots. They found that these factors related to impression of proximity affected the subjects' ratings of the pictures: the closer the proximity, the more intense the evaluation. The use of close proximities also led to higher attention and recall. As Reeves and Nass (1996) noticed, people respond to all types of media (e.g. picture, TV, computers) in a somewhat similar way. This suggests that proximity has an important role in human-computer interaction as well. In the context of human-computer interaction, however, research on proxemics has been very infrequent. Jeffrey (1998) showed that in virtual worlds, the users consciously control the proximity of their avatars. Close proximities are avoided. In the paper VI of the current thesis, we experiment with the possibility of simulating the proximity between the user and an anthropomorphic agent by varying the size of the agent.

4 Experiments

This thesis consists of six publications, all of which are written based on empirical experiments. Papers I and II (chapters 4.1 and 4.2) present the results of two empirical experiments, which studied pupillary responses to emotionally provocative sound stimuli. Paper III (chapter 4.3) describes an experiment on the experiential, facial, and behavioral effects of affective speech interventions during human-computer interaction. In Papers IV and V (chapters 4.4 and 4.5) it was studied, whether emotional experiences can be estimated based on facial expressions with a reasonable accuracy in real time. Finally, paper VI (chapter 4.6) presents an experiment on the affective effects of agent proximity in conversational systems.

4.1 PUPILLARY RESPONSES TO EMOTIONALLY PROVOCATIVE STIMULI

The first paper describes two experiments. The purpose of the experiments was to study pupillary responses while the subjects were listening to emotionally provocative sound stimuli. By using sound stimuli, we avoided the problem of different luminances associated with visual stimuli. In both experiments, the stimuli were taken from IADS and they were systematically varied on the valence and arousal dimensions. The subjects listened to the stimuli, while their pupil size was measured unobtrusively by using a floor-mounted eye tracker. After the measurements, the subjects heard the same stimuli again and rated their emotional experiences on the valence and arousal dimensions.

In the first experiment, 30 subjects (15 females) listened to ten negative and highly arousing stimuli (e.g. a baby crying), ten neutral stimuli (e.g. office background noise), and ten positive and highly arousing stimuli (e.g. a baby laughing), while their pupil diameter was measured. The results indicated that during the period of two seconds after the stimulus offsets, the pupil diameter was significantly larger after both negative and positive stimuli, as compared to neutral stimuli. The results also indicated that on average, the subjects of this experiment experienced the stimuli as intended (i.e. as positive/highly arousing, neutral and negative/highly arousing) on average.

In the second experiment, 22 subjects' pupil diameter was measured while they listened to four negative and highly arousing sounds, four neutral sounds, and four positive and highly arousing sounds. The stimuli were selected so that their contents were similar within each category (i.e. laughters, background noise, fighting and screaming). We assumed that this would reduce the possible effects caused by the differences in the amount of cognitive processing required by the different kinds of stimuli. The results showed that, the subjects' pupil diameter was significantly larger during the negative stimuli, as compared to the positive stimuli. During the period of two seconds after the stimulus offsets, the pupil diameter was also larger after the negative stimuli than after the neutral stimuli and the positive stimuli. The results also indicated that on average, the subjects of this experiment experienced the stimuli as intended on the valence scale, but on the arousal scale, there were no significant differences between the ratings of the highly arousing positive stimuli and the neutral stimuli, while the positive stimuli should have been more arousing according to IADS ratings.

4.2 PUPIL SIZE VARIATION AS AN INDICATION OF AFFECTIVE PROCESSING

Paper II presents the results of further analyses for experiment one described in paper I. The data of experiment one was baseline corrected in order to reduce the effects of an overall decrease in pupil size during the experiment observed in the data. We used a one-second prestimulus baseline. Specifically, the averaged pupil diameter during the one-second period before each stimulus onset was subtracted from the averaged pupil diameter during each stimulus and also from the average pupil diameter during the period of two seconds after the stimulus offsets. The results were also analyzed by using gender as a between-subjects factor.

The results indicated that during the stimuli the pupil dilation was larger during both highly arousing negative and positive stimuli, as compared to the neutral stimuli. A similar result was also found for the data of the two-second period after the offsets of the stimuli. A gender difference was also found; the pupil dilations were significantly larger for female than for male subjects during the neutral stimuli. Gender did not have significant effects on the ratings of experienced emotions.

4.3 The effects of affective interventions in human-computer interaction

The third paper investigated the effects of affective intervention during human-computer interaction. The aim was to study, whether computer-generated affective interventions can have beneficial effects on the user, when the user is experiencing negative emotions during computer use. 18 subjects (9 females) played an interactive problem-solving game, during which they encountered deliberately programmed system delays. After the delays, the subjects either got positive worded intervention, negative worded intervention, or no intervention. The positive and negative interventions were given by an agent via speech synthesis, and the speech prosody was kept as neutral as possible. The positive worded interventions described the continuation in a positive way (e.g. "The program functions were suspended. Great that they will soon continue working."), whereas the negative worded interventions deplored the negative event (e.g. "The program stopped. Now this is frustrating."). During the experiment, the EMG activity of two facial muscles was measured: *zygomaticus major* (activated in smiling) and *corrugator supercilii* (activated in frowning). Two time windows were used to analyze the EMG data: during the interventions and during the period of two seconds after the offsets of the interventions. The effects of the interventions to task performance were measured in terms of correct moves in the task interface during the period of first ten seconds as the interface was working again. After the experiment, the subjects also rated their subjective emotional experiences related to the system delay event and both the intervention categories in terms of affective valence and arousal.

The results showed that both types of intervention had desirable effects over ignoring the user. During the positive interventions, smiling activity was bigger as compared to negative intervention and the no intervention condition. After the positive interventions, smiling activity was larger than after the negative interventions and the no intervention condition. After the positive interventions, the frowning activity attenuated more than after the other conditions. Furthermore, after the positive interventions, the subjects' problem solving performance in the first ten seconds of the next task was significantly better than after no intervention. The subjects also rated their subjective valence as more positive after the positively worded interventions than after the negatively worded interventions and the no intervention condition. Subjectively experienced arousal was not significantly influenced by the different intervention conditions.

4.4 REAL-TIME ESTIMATION OF EMOTIONAL EXPERIENCES FROM FACIAL EXPRESSIONS

The purpose of the experiment was to study, how accurately the subjects' emotional experiences could be automatically estimated based on their facial expressions in real time. A person-adaptive system was constructed, with a total of 70 preprogrammed estimation models, which were implemented into the system for the current experiment. Three signals were measured: the activity two facial muscles, *zygomaticus major* (activated in smiling) activity and *corrugator supercilii* (activated in frowning) activity, and also an EMG difference score (smiling activity minus frowning activity). All these signals were analyzed in five different ways:

1) averaged EMG data without a baseline correction
2) averaged EMG data with a baseline correction
3) using peak values of the EMG data
4) using peak values baseline corrected with average activity during the baseline period
5) using peak values baseline corrected with peak activity during the baseline period.

A 0.5 s pre-stimulus baseline was used in all the models involving baseline correction. The system was tested as ten test subjects watched series of emotionally arousing negative, neutral, and positive pictures and videos. The pictures were taken from the IAPS system, and they varied systematically in affective valence. After seeing each picture or video, the subject rated it on a 1-9 affective valence scale. The system was first calibrated using 24 (8 negative, 8 neutral, and 8 positive) emotional picture stimuli, during which the subjects' *zygomaticus major* and *corrugator supercilii* EMG responses were registered.

After this, the system formed estimation models based on the subject's facial emotional responses and the ratings of experienced emotion. The categorical estimation models were formed so that between two categories the median value was used as a limit value determining the output of the model in the testing phase. For example, if *zygomaticus major* activity (increases with positive experienced valence) for a stimulus in the testing phase was higher than the median value in the calibration phase, the subject's response was judged as positive by that model. In the case of classifying into three categories, the same procedure was used except that 1/3 percentile and 2/3 percentile values were used as the limiting values between the three categories. The linear estimation models were formed using linear regression of each EMG signal and the valence ratings.

Then the estimation models were tested as the subject saw and rated 28 further pictures with varying emotional valences. During this phase, each model in the system estimated the subjects' emotions either categorically using two (negative and positive) or three (negative, neutral, and positive) emotion categories, or dimensionally using linear regression. The models were further tested, as the subjects saw six (two negative, two neutral, and two positive) emotionally provocative videos.

The results showed that for the static pictures, the best models were able to estimate between positive and negative experiences with an accuracy of 70.1% (difference score peak values baseline corrected with peak values). For videos, the best models distinguished between positive and negative experiences with an accuracy of 80.8% (*corrugator supercilii* baseline corrected). When estimating into three categories, negative, neutral, and positive, the best estimation accuracies were 46.0% for pictures, and 47.5% for videos. In both cases, difference score peak values baseline corrected with peak values was the best model. When estimating on the valence dimension with linear regression, the best obtained correlations were .91 for pictures (*corrugator supercilii* peak values baseline corrected with peak values) and .93 for videos (baseline corrected EMG difference score). Considering all the analyses, the best models were *corrugator supecilii* baseline corrected, the EMG difference score baseline corrected, *corrugator supercilii* peak values, and *corrugator supercilii* peak values baseline corrected with peak values.

4.5 PERSON-INDEPENDENT ESTIMATION OF EMOTIONAL EXPERIENCES FROM FACIAL EXPRESSIONS

The purpose of this study was to analyze the results of paper IV using person-independent methods. Three measures were used: *zygomaticus major* (smiling) activity, *corrugator supercilii* (frowning) activity, and a difference score (*zygomaticus major* activity minus *corrugator supercilii* activity). A method was used, in which the system estimated the response as either negative or positive by comparing the averaged EMG responses to the 0.5s baseline. The three models estimated the emotional experiences by the following rules:

1) If there was an increase in smiling activity, the experience was estimated as positive, otherwise as negative.
2) If there was an increase in frowning activity, the experience was estimated as negative, otherwise as positive.
3) If there was an increase in the difference score, the experience was estimated as positive, otherwise as negative.

The results showed that for using picture stimuli, the correct estimation accuracies were between 60 and 70% for all the models. Using video stimuli, the correct estimation accuracies were about 60% by analyzing smiling activity only (model 1), over 80% by looking at frowning activity only (model 2), and also over 80% by analyzing the difference score based on both facial muscles (model 3).

4.6 Affective effects of agent proximity in conversational systems

The aim of this experiment was to study, if the subjects' affective experiences could be influenced by varying the size of a conversational agent character giving the messages using speech synthesis. The four proximity levels – intimate, personal, social, and public - suggested by Hall (1966) were simulated by displaying an anthropomorphic agent character on a screen in different sizes. The agent gave 24 speech messages to the subject, which consisted of two positive, two neutral, and two negative messages for each of the four proximity levels. After each speech message, the subjects rated their experienced emotions using three dimensions: valence, arousal, and dominance. They also rated the perceived intimacy of the messages. At the end of the experiment, the subjects also chose the preferred proximity level for an anthropomorphic conversational agent.

The results showed that by manipulating the simulated proximity level of an agent character, the subjects' felt affective dominance could be significantly influenced. The two farthest proximities evoked significantly more experienced dominance in the user than the two closest proximities. Within the personal proximity level, the negative messages were experienced as more arousing than the positive messages. By manipulating the affective wording in the message, the subjects' experienced valence could be significantly influenced. Five subjects chose the personal proximity level as the preferred proximity level, while three subjects chose the social level.

5 Discussion

5.1 PUPIL SIZE VARIATION

The first two papers of this thesis presented the results of two experiments on pupillary responses to emotionally provocative stimuli. Based on the results, the magnitude of the pupillary response to emotional stimuli is mostly determined by the amount of emotional arousal. This was evidenced by the larger pupillary dilations after the stimuli experienced as highly arousing, as compared to the stimuli experienced as neutral. In experiment one, both positive and negative stimuli were experienced as highly arousing, and they evoked significantly larger pupillary responses, as compared to the neutral stimuli. When the results of experiment one were baseline corrected (Paper II), the results showed significantly larger pupillary dilations both during and after the positive and negative highly arousing stimuli as compared to neutral stimuli. In experiment 2 of Paper I, during the negative stimuli there were significantly larger pupillary dilations than during the positive stimuli, which were experienced as significantly less arousing. After the negative stimuli, the pupil diameter was also significantly larger than after the positive stimuli, and the neutral stimuli, which were both experienced as less arousing as the positive stimuli. In all, these results give strong evidence for the connection of pupil size variation and emotional arousal.

In experiment one, there were a variety of different sounds. Thus, it is possible that the pupillary effects of cognitive processing load affected the results. For each sound, the subject had to try to understand the contents of the sound first. In experiment two, the contents of the stimuli were more controlled than in experiment one in order to possibly reduce the amount of cognitive processing required. This change in experimental design may indeed have caused the fact that there were significant differences during the stimuli in experiment two, but not in experiment one. In paper II, a baseline correction was employed to the data of experiment one. By using a baseline correction, it was possible to avoid the effects of possible autonomic drifts in pupil size during the experiment, as well as the effect of an overall decrease in pupil size during the experiment (e.g. Hess and Petrovich, 1987; Hyönä *et al.*, 1995). In comparison to the analysis of the raw data the baseline corrected data analysis showed reliable differences in pupil size even during the auditory stimuli.

We observed a linear connection between pupil diameter and emotional arousal. However, it is important to notice that the overall emotional experience is constituted by valence and arousal together, and cannot be examined fully separately. For example, high emotional arousal typically occurs only in conjunction with highly negative or positive experiences, and it is unlikely that subjects would report a neutral and very aroused emotional rating. Thus, the current results suggest that highly positive or negative stimuli, which are also experienced as highly arousing, evoke larger pupillary responses than stimuli experienced as neutral. In experiment one, both negative and positive stimuli were experienced as highly arousing. In experiment two, only the negative stimuli evoked sufficiently high arousal in order to activate the pronounced pupillary responses.

Because the pupil is governed by the autonomic nervous system, the results of the experiments described in Papers I and II suggest that the autonomic nervous system responds differently to emotionally arousing than to emotionally neutral stimuli. Studies by other researchers have shown similar results, for example, Bradley and Lang, (2000) showed higher autonomic nervous system activity, in terms of skin conductance responses, to both emotionally positive and negative than to neutral auditory IADS stimuli. In the same study, the effects of affective IADS sound stimuli to heart rate were studied. The results showed that valence and arousal had an interaction effect on heart rate so that the observed heart rate decelerated significantly more in response to negatively highly arousing sounds, as compared to negative low arousing sounds and neutral sounds. The difference between the positive and the neutral sounds were not significant. The subjective ratings indicated that negative sounds were rated as more arousing as positive and neutral sounds, and positive sounds were rated as more arousing than neutral sounds. These results further stress that the autonomic nervous system is differently activated when emotional arousal is sufficiently high.

Our results are also in line with earlier findings suggesting that emotional responses to discrete categorical stimuli associate with differential autonomic nervous system activity. For example, anger, fear and sadness stimuli have produced larger heart rate acceleration than happiness, surprise, and disgust. Furthermore, anger has been associated with higher finger temperature than the other five basic emotions, and sadness has been associated with larger skin conductance decrease than fear, anger, and disgust (Ekman *et al.*, 1983). In a further study (Levenson *et al.*, 1990) it was found that anger and fear produced larger heart rate acceleration than disgust and happiness, and anger produced larger finger temperature increase than fear. Further, fear and disgust produced larger skin conductance increases than happiness. Thus, these studies suggest differences in autonomic nervous system correlates both within negative emotions and between negative and positive emotions. Levenson (1992) has stated that autonomic differences appear to be more robust in relation to negative emotions. In the results presented in papers I and II of this thesis, however, the positive highly arousing sound stimuli also caused clear autonomic differences.

When the results of Bradley and Lang (2000) and Levenson *et al.*, (1990) are compared to our results, it can be concluded that pupil size is a relatively sensitive index of autonomic nervous system activation, as compared to other autonomic nervous system correlates. In experiment I, significantly elevated pupil dilations were caused by both positive and negative highly arousing IADS stimuli, while Bradley and Lang (2000) reported significantly more heart rate deceleration for only negatively highly arousing IADS stimuli. Levenson *et al.*, (1990) showed biggest heart rate accelerations for acted negative differential emotions such as anger, fear and sadness. Thus, it seems that the effects of emotional stimulation on heart rate change (acceleration vs. deceleration) are not clear or depend on the type of manipulation (e.g. sound stimuli vs. acted expressions). On the other hand, in Bradley and Lang (2000), significant differences in skin conductance were found for both negative and positive stimuli. This suggests that skin conductance is also a somewhat sensitive indicator of autonomic nervous system activity.

In the older studies of pupillary responses to emotional stimuli, the temporal resolution of the measurement technology was not sufficient enough to examine small variations in the responses temporally. Thus, in order to examine the subjects' pupillary responses as a function of time, we also reported some descriptive results in both papers I and II. Averaged timeline curves were drawn for all the stimulus categories for both experiments and both analyses. The visual inspection of the timeline curves revealed that the pupil dilations followed on average a similar type of curve. A sudden increase in pupil size at about 400 ms from the stimulus onset was followed by a slower decrease. In experiments one and two without a baseline correction, the pupil diameter lines for positive, neutral, and negative stimuli were clearly separated from each other after the initial increase, and the differences remained until the end of our analysis window. In the baseline corrected data of experiment one, the subjects' averaged pupil responses were separated from each other after the initial increase at about one second from the stimulus onset so that emotional negative and positive stimuli evoked clearly larger dilations than neutral stimuli. The lines of pupillary responses to negative and positive stimuli were approximately on the same level throughout the analysis window.

Drawing averaged timeline curves verified that our two chosen time windows – during the 6-s stimuli and during the period of 2s following stimulus offsets – were appropriate in the sense that there were significant differences between averaged pupil diameters. In future studies, the period of one second from the stimulus onset could be left out of analysis, because the responses to the different categories did not yet differ from each other to a great extent at this period. In our experiments, the differences were greatest right before the end of the stimuli, after which they started to decrease. The 2-s analysis window after the stimulus offsets seemed to be sufficient, because the lines for the different affective categories already started to approach each other before the end of this 2-s window.

Our data showed relative pupil diameter changes ranging approximately from five to nine percent from the baseline at largest. In some earlier studies, the variation has been smaller. For example, in Bernhardt (1996) the peak points of the responses were approximately from one to five percent of the baseline value, while the subjects viewed peaceful and fighting pictures. The larger dilations in our experiments suggest that the sound stimuli used in the experiments had relatively strong autonomic effects. In experiment two the lines were more varying than in experiment one. Because there were fewer different sounds than in experiment one, it is probable that this variation was caused by the temporal characteristics of individual stimuli.

By looking at the timelines of single responses we observed a constantly occurring initial increase in pupil diameter shortly after the stimulus onsets. However, there was some variation in the magnitude of the initial increase, and in some cases, the pupil diameter decreased below the baseline level before the stimulus offset. Our findings of pupillary dilations are in contrast to some earlier findings that indicated pupil size constrictions to negative affective stimuli (e.g. Hess, 1972). Thus, our results support the results that have found pupil dilations to affective stimuli (Loewenfeld, 1966; Janisse, 1974). Similarly to the results presented in this thesis, Janisse (1974) suggested that pupillary dilations are linearly related to affect intensity and curvilinearly related to affective valence so that positive and negative affective states produce largest pupillary dilations.

Pupillary dilations are associated with the sympathetic nervous system activation, and pupillary constrictions are associated with the parasympathetic nervous system activation. In our experiments, the typical pupillary responses to the emotional stimuli were initially relatively fast dilations, followed by slower constrictions towards the baseline. These results are in line with the notion that the sympathetic nervous system puts the body quickly into a sort of "alert state", from which the parasympathetic nervous system pulls it back.

In this respect, emotion-related autonomic nervous system changes are rather rapidly manifested as pupil size variations. These notions have some implications for the design of affective computing based on autonomic nervous system correlates. In the current thesis, we analyzed averaged data over the whole stimulus interval or averaged data from a two-second post-stimulus interval. In practical recognition of affective information from pupil size variation, it could be sufficient to concentrate on the magnitude of the initial pupillary dilation (sympathetic nervous system response). The findings that the sympathetic nervous system activates the pupil relatively rapidly suggest that a computer could possibly be able to respond to user emotion inferred from pupil size within an acceptable delay (e.g. 2-3s from the start of the emotional stimulus assuming minimal computer processing time). Using some another autonomic nervous system correlates as input methods the corresponding delay is longer. For example, skin conductance responses to pictures have become visible at about 2.5s after stimuli onsets and been at their peaks at about 4.5s after stimulus onsets (Codispoti *et al.*, 2001). Furthermore, heart rate changes to pictures and sounds have been at their peaks at 4 to 6s after stimulus onsets, and the differences between different affective stimulus categories (i.e. negative, neutral, and positive) have also been at their largest during that time period (Bradley and Lang, 2000; Codispoti *et al.*, 2001; Anttonen and Surakka, 2005). Based on these results, the pupil size differentiates between affective stimulus categories faster than heart rate or skin conductance. However, using the pupil the differences were also at their largest at 4-6s after stimulus onsets. Consequently, the computer could also infer emotions from the parasympathetic nervous system activity (pupil size decrease after initial increase). Then the computer response time would be longer, but still well comparable to that of, for example, heart rate and skin conductance.

In paper II, the baseline corrected results of experiment one were also analyzed using gender as a between-subjects factor. By looking at the descriptive results, one easily notices that the female subjects' pupillary dilations were largest for the positive stimuli, while the male subjects' pupillary dilations were largest for the negative stimuli. However, the gender differences were only statistically significant during the listening of neutral stimuli. Earlier, Bernhardt *et al.* (1996) found a somewhat similar gender difference, as they studied pupillary responses to peaceful and fighting pictures. In their study, female subjects' pupils dilated significantly more to peaceful pictures and male subjects' pupils dilated significantly more to fighting pictures. These results together suggest that emotion-related gender differences in pupillary behavior are worth studying further and that emotion recognition based on pupil size variations might benefit from knowing the user's sex.

In both experiments, the stimuli used were selected from the IADS system based on the results obtained by Bradley and Lang (1999). It was assumed that the subjects of these experiments would experience the stimuli in a similar way. Indeed, the subjective ratings in both experiments were generally similar to the ratings obtained by Bradley and Lang (1999). In experiment one, the division of the independent variable into the three stimulus categories was quite similar as in Bradley and Lang (1999) and thus successful. However, in experiment two, the positive stimuli were rated as less arousing than the neutral stimuli, while according to Bradley and Lang (1999) they were rated as more arousing than the neutral stimuli. As our results generally suggest that the magnitude of pupillary dilations is connected to affective arousal, this subjective finding might also explain the relatively small pupillary dilations during and after the positive stimuli as compared to experiment one: our subjects did not experience the positive stimuli as especially arousing in experiment two.

In earlier studies of pupil size and affect, the stimulus materials have varied greatly. Mostly they have been pictorial stimuli, which had not been studied previously, and thus their impact on the subjects was unsure. In the experiments of papers I and II, we controlled as many variables as possible. Hess and Petrovich (1987) have listed several different sources of pupil size variation, including, for example, the light reflex, different stimulus parameters (e.g. visual and chemical), and information-processing load. We utilized well-studied auditory stimuli, thus, avoiding the possible problems caused by variations in stimulus luminance while using visual stimuli. The stimuli were of equal length and they were played back at equal volume. There were also an equal number of female and male subjects.

The motivation for the research of papers I and II was the idea that pupil size could possibly be a practical input signal in the field of affective computing (e.g. Picard, 1997). For affective computing, measuring pupil size has important advantages. It can be measured unobtrusively with video cameras, and no sensors have to be attached to the user. New techniques are being developed, which can measure pupil size reliably and accurately with ordinary inexpensive video cameras. Furthermore, an important advantage of pupil size measurement is that pupil size variation is an involuntary index of autonomic nervous system activity.

This means that pupil size variation is not easy to control voluntarily and it indexes real spontaneous activity. Thus, pupil size measurement can avoid problems that are inherently related, for example, to the automatic analysis of facial expressions, because it is known that facial emotional behavior can be and are masked, inhibited, exaggerated, and faked (e.g. Ekman and Friesen, 1982; Ekman, 1985; Surakka and Hietanen, 1998).

The practical success of pupil size variation as an input signal will depend on the effectiveness of associated digital signal processing methods. In paper II, it was suggested that baseline correction is a useful procedure in pupil size research and applications. The averaged pupillary responses discriminated between the different categories on an average level. In order to use pupil size in human-computer interaction, however, the computer has to try to infer the user's emotions based on single pupillary responses. To achieve improved recognition accuracy, for example, between aroused and calm emotional responses, a more detailed focus or improved analysis methods are needed instead of just averaging the responses over the whole stimulus time or two seconds after it. For example, it could be possible to recognize emotionally aroused responses based on the magnitude or slope of the sudden increases in pupil size that occurred within the first two seconds in the experiments presented in this thesis. Another challenge is how to compensate for the effects of other factors affecting pupil size. During human-computer interaction, two of the other most important factors affecting pupil size are probably variations in cognitive load and lighting conditions. It has been known for a long time that cognitive load is one factor significantly affecting pupil size (Kahneman and Beatty, 1966; Hyönä *et al.*, 1995). For example, in the pupil size data, it seems to be relatively hard to compensate for the cognitive effects. Because the pupil size grew as the stimulus started, and it also grew a little when the stimulus ended, it seems that cognition of a sensory change causes an increase in pupil size, which is sustained if the stimulus is perceived as affectively meaningful. Furthermore, the pupil has a very strong light reflex. In varying lighting conditions or when using programs with varying screen luminances, a compensation for the light reflex is also likely to be obligatory to be able to use pupil size variations in practical human-computer interaction. The implementation for such compensation could be realistically achievable, because the pupil light reflex is very well studied, and there are accurate mathematical models describing it (e.g. Bressloff *et al.*, 1996). Integrating the models on cognitive pupil responses and pupil light reflex to emotion recognition could possibly enable the distinction of emotional responses from responses to cognitive load and lighting.

What kind of emotion recognition seems possible based on the current results then? The timeline analyses and statistical results showed that pupil dilation behaves coherently on an average level and it was significantly related to emotional stimulation. It seems possible to develop signal analysis methods for detecting significant deviations of pupil size from baseline activity. These deviations could be used as indicators that the user is either emotionally or cognitively involved in human-computer interaction. With more advanced methods it might even be possible to distinguish between different kinds of reactions, for example, emotional vs. cognitive, or negative vs. positive.

5.2 THE EFFECTS OF AFFECTIVE INTERVENTIONS

In paper III, the effects of affective speech interventions were studied during human-computer interaction. The interventions were delivered as the subjects were exposed to series of deliberately caused mouse dysfunctions during a problem-solving task. The results suggest that in the context of software delays, which prevents the user from accomplishing her/his task, computer generated affective interventions can be used to alleviate the effects of the system problems on an affective level. This was evidenced by the results, which showed that as compared to the condition, in which the subjects did not get any intervention, both positively and negatively worded affective interventions caused positive changes on subjective, physiological, and behavioral levels. Specifically, the changes were accomplished in terms of subjective emotional experiences, EMG activity of facial muscles associated with smiling and frowning, and task performance using a ten second analysis window in the beginning of the next task.

Even though there are currently no experiments wholly similar to the experiment described in paper III, the results are generally in line with many findings from previous research. Previous research has suggested an inversely linear relationship between frowning activity and ratings of valence. Decreased frowning activity has been found to be related to emotionally positive material, while negative material has caused increased frowning activity (Lang *et al.*, 1993; Bradley and Lang, 2000). In the current experiment, the results for the *corrugator supercilii* showed a decrease in the activity of that muscle for the both negatively and positively worded interventions. The reason for that was probably that in the context of the delay event, the subjects appreciated both types of interventions as more positive than the delay event, which can also be seen in the subjective ratings. In the current experiment, the differences in frowning activity between the different categories were significant only after the positive interventions, as compared to no intervention. However, the other pairwise differences after the interventions were also very near statistical significance.

The current results for the *zygomaticus major* muscle site EMG activity showed an increase in response to positively worded interventions and a smaller increase in response to negatively worded intervention. A similar trend has been reported earlier for emotionally positive and negative visual material (Lang *et al.*, 1993). Bradley and Lang (2000) studied the facial EMG responses to affective sounds. Their results showed a significant linear dimensional correlation between smiling activity and ratings of pleasure, but the differences between different categories (negative, neutral, and positive) were not significant. In the current experiment, the mouse delays and the affective interventions seemed to elicit more variation in smiling activity than the emotional stimulus categories in Bradley and Lang (2000), and most of the differences in smiling activity between the different intervention categories were also statistically significant.

The behavioral effects of the interventions are also supported by the previous research. Isen (2000) stated that a growing body of research indicates that even mild positive affective states can markedly influence everyday thought processes leading to improved decision-making and problem-solving. This conclusion was supported by the current results, in which the use of positive interventions resulted in improved problem-solving performance in terms of the number of successful moves shortly after the interventions. A similar result was obtained by Aula and Surakka (2002), who found that emotionally positive feedback during human-computer interaction provoked shorter response times to simple mathematical problems as compared to negative feedback. In the current study we showed that this effect could be achieved with interventions not related to the actual task.

The valence ratings showed that the interventions were successful in evoking emotional experiences. The positive worded interventions were successful in evoking significantly higher subjective ratings than negatively worded interventions and no intervention. In contrast to valence ratings, there were no significant differences in the ratings of arousal for the different conditions. Thus, it was possible to manipulate experienced valence without significantly affecting experienced arousal. It is possible that the arousal associated with the frustration remaining after the mouse jam period was on average approximately on the same level as the level of arousal generated by the interventions.

The results of the current study showed that synthetic speech with emotional content could be efficiently used in affective interventions to achieve positive effects. In line with this, Aula and Surakka (2002) found that synthetic speech feedback could be used to evoke different autonomic nervous system responses in the user, measured in terms of pupil size variation, and subjective ratings of affective state. In a recent study, Ilves and Surakka (2004) showed that emotionally worded synthetic speech could evoke differential *corrugator supercilii* activity. In their study, frowning activity attenuated significantly more during and after positive worded speech than after neutral speech. Both these studies also found differences in subjective ratings of different affective stimuli or events in terms of affective valence and arousal. Thus, it seems that using affective speech, the subject's emotions can be regulated in terms of both physiological activity and subjective experiences. In addition, it seems possible to evoke significant improvements in the user's task performance using affective speech.

Recently, another study closely related to our experiment on affective interventions was published by Prendinger *et al.* (2005). Their subjects played a mathematical quiz game with an anthropomorphic figure, "Shima", acting as a quizmaster. Similarly to our experiment, the user experienced system delays during the critical moments in the competitive quiz. There were two conditions. First, there was an affective condition, in which the agent gave textual "happy-for" or "sorry-for" affective feedback after each question depending on whether the answer was right or not, as well as emphatic feedback after the delay. Second, there was a non-affective condition, in which the agent gave no feedback after the delays and only a very brief feedback ("correct"/"incorrect") after each question. The authors concluded that the agent expressing empathy after the delays was helpful as it elicited significantly lower stress, measured in terms of skin conductivity responses, than the non-emphatic agent. These results are in line with our conclusions on affective interventions.

An inevitable question following the analysis of all these results is, whether the results can be generalized to support the notion that the use of affective interventions is generally useful during human-computer interaction. Before making such an interpretation, one has to consider a few characteristics of the experimental setup of the experiment described in paper III. The interventions were given in the context of an emotional event, during which the subjects averagely reported negative experienced emotions. The system was also unusable at the time when the interventions were given, and the interventions did not have the possibility of interrupting the user. Furthermore, the interventions were designed with care, and they were especially tailored for the events deliberately caused in the experiment (system problems). Consequently, the current results suggest that affective interventions can be useful, considering that they are 1) well designed, 2) given at a right time, and 3) in an emotional context, in which they are useful. It is known from studies using the CASA paradigm that people generally respond to computers essentially in the same way as to other people (Reeves and Nass, 1996). It seems very likely that the effects of computer-generated affective interventions closely resemble the effects of similar interventions given by other humans. It has been suggested that strategies from human-human communication can be applied in human-computer interaction (e.g. Picard and Klein, 2002). The above-mentioned three principles certainly apply in human-human communication, and they seem to be prerequisites for effective affective interventions in human-computer interaction as well.

In paper III, we introduced the concept of affective interventions. This was due to the fact that in that experiment the interventions were not related to the task, but to the problematic use situation. Then the computer took the initiative to influence on the user's affects. In human-computer interaction, the transition from affective feedback to affective interventions could open new possibilities. The majority of the current systems do not use any affective interventions. In the current experiment, the no intervention condition represented the state of the art in most current systems. Thus, the results suggest that current systems would benefit from incorporated affective interventions. An example of a system that could be constructed based on the results this experiment would be one that notice system difficulties that are not under user control (e.g. a failure or delays in software execution), and responds to the difficulties in an affectively supportive way. Another application of affective interventions could be in situations, in which the user has to wait for something to be processed (e.g. exceptionally long application loading times or long download times from the Internet).

Affective interventions could also be pre-emptive (Picard and Klein, 2002), for example, they may be given before the user has encountered a certain problem in her/his work. There are some application fields, which could especially benefit from affective interventions. In fields in which motivation is a crucial factor, such as computer-assisted learning, affective interventions could be assumed to have beneficial effects. In addition to affective interventions, there is much room for improvement in the error messages, which are frequently used in different applications in ordinary human-computer interaction. They could show emotional intelligence strategies similar to our affective interventions by being affectively supportive in a constructive way.

Affective interventions could also be used as computer responses to significant changes in the user's physiology. Following the knowledge on how to recognize emotional states from affective signals, the question of how to react to the changes in the user's affective state becomes important. Physiological signals can be monitored in real time by the computer, and the computer can also monitor the physiological changes caused by its affective interventions, and adjust its behavior accordingly. Consequently, this kind of advanced adaptive technology could even show emotional intelligence similar to that of humans. For example, the computer could rule out the use of interventions that cause negative user responses in the context of a particular situation or a particular user. Consequently, in further research, it would be worth studying, in which cases affective intervention could be useful, and also whether they would also be useful within human-computer interaction without any special system problems. This would require sensing the user's emotional state using some of the methods discussed in Chapter 3. It would be important to know, when to give affective interventions to the user, and how the interventions should be formed to achieve the desired effects. Different output channels could be used in affective interventions. Some forms of interventions such as text (Prendinger *et al.*, 2005) would often mean interrupting the user, while, for example, speech could often be given synchronously with human-computer interaction. Further, the effects of animated agents with different gestures and facial expressions to convey additional affective information could be studied. When using natural language, it would be also worth studying how the intervention should be worded to achieve the desired effects. In speech interventions, the effects of different speech prosodies could also be studied. By also manipulating the prosody of the speech, it could be possible to achieve even stronger effects than reported in paper III.

In analyzing facial expression, the cognitive and social facial expressions need to be distinguished from emotional facial expressions. For example, smiling or frowning may be a response to a story told by someone else in the room, which signals that the user has understood the positive or negative message of the story, but does not necessarily experience the respective emotion. In addition to negative emotion, *corrugator supercilii* is also activated in knowledge of goal obstacles (Pope and Smith, 1994). In our study, the frustrating event was also a goal obstacle, as the users were instructed to complete the task as fast as possible. In some cases, however, there could be frowns associated with momentary goal obstacles, not necessarily accompanied by any emotion.

But what will be the role of facial expressions in human-computer interaction in the future? The findings of paper III encourage the measurement of facial expressions during human-computer interaction. The subjects' emotional reactions to affective interventions were verified in the activity of two facial muscles. This suggests that the measurement of facial expression could be useful in evaluating human-computer interaction on an affective level. In addition, because it seems possible to measure the user's facial expression during human-computer interaction with enough accuracy, the measurement of facial expressions could be used as an affective computer input signal, and the computer could interactively react to event changes from facial expressions. Thus, based on our results we share the view of, for example, Lisetti and Schiano (2000), who consider the measurement of facial expressions as useful computer input. Hazlett (2003) found practical evidence that measuring *corrugator supercilii* could reveal user frustration during human-computer interaction. Ward (2004) studied facial expressions during human-computer interaction and found that sudden events during computer use caused significant visible changes in facial expressions in many different parts of the face. These and our findings suggest that facial expressions are sensitive to events in human-computer interaction. Thus, they could be useful in both system evaluation and as an input signal in affective human-computer interaction.

5.3 ESTIMATION OF EMOTIONAL EXPERIENCES FROM FACIAL EXPRESSIONS

The results presented in papers IV and V gave support for the argument that emotional responses are manifested in facial expressions, and suggest that facial expressions are potentially useful as affective computer input signals. This was evidenced by the results that by measuring the electromyographic activity of two facial muscles, *zygomaticus major* and *corrugator supercilii*, it was possible to estimate the subjects' affective experiences reasonably well. In paper IV, using both picture and video stimuli, most of the 70 models used in the experiment distinguished between positive and negative affective responses at a better than chance estimation rate. When using video stimuli, the best estimation rate between positive and negative responses was over 80%. The models that classified estimated experiences into three affective categories, positive, neutral, and negative, also performed better than chance for both pictures and videos. There was also a relatively good correlation between the estimations of emotional experiences and the actual ratings of experiences given by the subjects themselves. The analysis of the subjective ratings confirmed that the subjects experienced emotions as intended in the experimental design.

In paper V, the data presented in paper IV were analyzed using person-independent methods, that is, methods that could be used for any person without a person-adaptive calibration. We used very simple models (see section 4.5) to distinguish between positive and negative emotional responses based on facial expressions. The results showed that by measuring the changes in the EMG activity of two facial muscles, *zygomaticus major* and *corrugator supercilii*, it was possible to distinguish between positive and negative experiences with reasonable recognition rates (60% - over 80%), which were also clearly better than chance for most of the models. Thus, the results of this analysis showed that it was possible to obtain person-independently results that were about on the same level as in paper IV using a person-adaptive system.

In the experiment of papers IV and V, the subjects had increased *zygomaticus major* activity mostly in response to positive stimuli, and increased *corrugator supercilii* activity mostly in response to negative stimuli. Thus, the subjects' EMG responses were generally in line with those found in previous research using emotional pictures (Greenwald *et* *al.*, 1989; Dimberg, 1990; Lang *et al.*, 1993; Larsen *et al.*, 2003). Previous research suggests that baseline corrected *corrugator supercilii* activity has relatively high correlations with experiences of affective valence. Overall, the models based on the activity of that muscle were also among the most accurate models in the current estimation experiment. However, our results also suggest alternative models that might be used with similar or possibly even better results. Models based on *corrugator supercilii* were most successful with the dynamic video stimuli. However, in some analyses with the picture stimuli models based on *zygomaticus major* performed even better. In addition, the difference scores, which were calculated by subtracting the *corrugator supercilii* activity from the *zygomaticus major* activity, had relatively high estimation rates in many analyses. These results suggest that in some cases it could be useful to combine information from both muscles, as compared to just measuring the *corrugator supercilii*.

Overall, however, the current and the previous research suggests that *corrugator supercilii* is the single muscle that is most closely linked with the affective valence dimension. The superiority of *corrugator supercilii* is also in line with the results by Costantini *et al.* (2005) who found recently that the upper part of the face (the eye region) is more important in human recognition of emotions from facial expressions (of other humans and synthetic characters) than the lower part of the face (the mouth region). The results presented in papers IV and V suggest that the case is similar in the computer recognition of human emotions: *corrugator supercilii* provides especially much affective information.

Of the different ways used to analyze the data from the two muscles, the estimation rates were highest either using a baseline correction or using difference score peak values that were baseline corrected by peak values. The baseline correction is a commonly used method in psychophysiological studies. In a baseline correction, activity during a baseline measurement preceding a stimulus is subtracted from the physiological activity caused by the stimulus. The current results suggest that a baseline correction can improve results in emotion estimation based on single responses. The analysis of peak responses is less often reported in psychophysiological literature, but the current results indicate that it could be a noteworthy possibility in emotion estimation, especially when baseline corrected with peak responses during a baseline period. In all, the findings using different analysis methods indicate that models estimating psycho-emotional experiences on the basis of facial activity can be created in several ways and they are worth studying further in future research.

It has been noted in some earlier studies (e.g. Ekman, 1985; Cohn *et al.*, 2002) that the individual differences in facial expressions are quite large. Individual differences were also quite large in the current experiment.

This suggests that the success of emotion estimation from facial expressions in practice will depend on the particular user's facial behavior. In our experiment some of the subjects showed only minor activations of the *zygomaticus major* and *corrugator supercilii* muscles. It seems that for these persons emotion estimation is only possible in response to very strong affective stimuli or events. On the other hand, some subjects were quite expressive, and their facial responses were quite clear using both the *zygomaticus major* and *corrugator supercilii* muscles, and the estimation worked quite robustly even with picture stimuli and simple analysis methods.

Based on previous research, which has emphasized the role of individual differences in facial emotion expressions (e.g. Cohn *et al.*, 2002), it would have been reasonable to presume that person-adaptive estimation would have been more accurate than person-independent estimation in the current experiment. One explanation why the person-adaptive system was not more accurate in our experiments seemed to be that there was relatively much variation in the responses even within individual subjects' expressions during the experiment. In the regression analysis, the regression lines became rather flat for some subjects, because the facial responses were subtle, and because of the variation was still rather big. The current results were, however, good for the majority of the subjects. More advanced adaptive systems are very likely to be more efficient than person-independent systems in the future, because they can estimate what is normative for the specific user. In the future, the system should also be also able to differentiate between short emotional reactions and longer-lasting affective contexts. For example, if the user is in a very positive emotional state using a computer and encounters a small problem, the facial and experiential response is likely to be negative compared to the state before the problem, but the resulting emotional state could still be fairly positive, and there could still be a fair amount of smiling activity left. In this case, adaptive methods could in the future recognize the overall valence (fairly positive), while the person-independent methods of paper V would only recognize the direction of change (negative). However, in analyzing momentary emotional responses, the methods of paper V are likely to be valuable. From the direction of changes in facial responses, the system could learn, how the user affectively responds to different small events in human-computer interaction. In the future those methods could be extended with methods for assessing the affective context, in which the responses occur. An advanced system could even use a solution, which alternates between adaptive and non-adaptive methods like the methods used in papers IV and V.

An important advantage of the methods of paper V for emotion estimation is that the estimation can be done in real time based on even small single emotional responses. In the analyses of paper V, the analysis window was

6s, but it could be smaller just enough to capture the first affective reaction. Thus, the methods presented have some practical importance, as they enable interactive affective feedback from the system in human-computer interaction almost at the same time as the emotion-related changes in the user's facial muscle occur. In the current experiment, even though our system detected the subjects' facial activations in real time, we did not incorporate any system responses to the facial activations, because they would have influenced our results on emotion estimation. However, in future studies it seems especially important to study the effects of a system that reacts to the user's affective facial responses in real time.

In experiments IV and V, we used EMG measurement technology, which is an accurate means for studying muscle activations. EMG technology is advancing rapidly, and it is a noteworthy measurement technology for human-computer interaction in the future. Good results in facial expression recognition can be, however, also achieved unobtrusively by using machine vision techniques to track the lip corners and the location of the brows. For example, an automatic machine vision system developed by Tian *et al.* (2001), achieved an average recognition rate of over 96% for activations of all upper and lower face action units, and a perfect rate of 100% for both the activations of the action unit associated with frowning (AU4) and the action unit associated with smiling (AU12), when they appeared independently of other facial muscle activation. Even though these activations were larger than the subtle changes studied in the current paper, the results by Tian *et al.* (2001) suggest that in addition to using EMG, facial expressions can also be somewhat reliably detected using machine vision methods. The EMG measurements of facial muscles are correlated with changes in visually detectable features in the face. For example, Cohn and Schmidt (2004) have found that in smiles with detectable *zygomaticus major* EMG onset, 72% of the responses detectable using EMG were in agreement with visible onsets of the lip-corner coordinate. Furthermore, there was a .95 temporal correlation between detectable *zygomaticus major* EMG onset and detectable visible onset of the lip-corner coordinate. In emotion estimation using machine vision methods, there are still challenges, for example, in compensation for changes in illumination (Adini *et al.*, 1997) and head rotations (Oliver *et al.*, 2000; Bartlett *et al.*, 2003). However, these results suggest that in future human-computer interaction, changes in facial expressions could also be successfully tracked using machine vision methods, in addition to the still more accurate EMG-based methods employed in the current thesis.

We estimated the subjects' experienced emotions from their facial expressions using the dimensional emotions framework. Many other researchers' early visual systems have taken another viewpoint. They have aimed at recognizing emotions categorically, most often by using visual methods to discriminate between the six basic emotions proposed by

Ekman (1992). In many cases, other systems are only successful in recognizing emotions from acted expressions, presented as still pictures. Examples of such systems include, for example, the systems by Pantic and Rothcrantz (2000) and Dailey *et al.* (2002). However, the basic emotions by Ekman (1992) are associated with large prototypical facial expressions, which occur quite rarely. There is some empirical evidence that facial expressions during computer use are generally not very large (Ward, 2004). Spontaneous facial expressions also differ from posed expressions in many ways (Ekman, 1985). These findings highlight the need for real-time emotion estimation based on subtle spontaneous facial expressions, as seen in the current approach.

On the other hand, there are even now some systems, which can recognize even subtle spontaneous facial expressions and estimate the intensity of the expressions fairly well (e.g. Lien *et al.*, 1998; Oliver *et al.*, 2000; Bartlett *et al.*, 2003). A holistic spatio-temporal model of facial expressions has been developed by Essa and Pentland (1997), and timing information has also been also successfully utilized in emotion estimation by Cohen *et al.* (2003). In order to develop methods presented in Papers IV and V further, one promising direction could be incorporating advanced models of timing information into our emotion estimation system. The recent results by Cohn and Schmidt (2004) gave further evidence that timing information is important in the analysis of facial expressions. They could discriminate between posed and spontaneous smiles with an accuracy of 93% by using duration, amplitude, and ratio of duration-to-amplitude, and with an accuracy of 89% by using only duration-to-amplitude. In line with this, incorporating the more advanced use of timing information to our system could enhance estimation results in future studies.

The methods presented in this paper offer a noteworthy extension to other currently existing methods for estimating emotion from physiological measures. In line with this we suggest that taken together the findings on the association of emotions and facial muscle activity it is very likely that measuring facial EMG, especially *corrugator supercilii* and *zygomaticus major* activity, can objectively reveal both usability problems and moments of user satisfaction. Measuring *corrugator supercilii* activity could also reveal car drivers' stress as in Healey (2000) or, for example, stress during computer use.

In sum, the current results showed that it was possible to estimate the user's subjective affective experience based on the activations of the user's facial muscles with reasonable accuracy. Especially the categorical estimation to positive and negative responses worked rather well and could be useful in practical systems. Using our system, the estimation can be done in real time, which means that in subsequent studies the computer could also change its real time behavior in a socially meaningful way according to the estimation results. We believe that these results are a step toward the development of computers, which have human-like social and emotional capabilities, such as the ability to infer affective experiences from facial expressions.

5.4 AFFECTIVE EFFECTS OF AGENT PROXIMITY

The results of paper VI suggest that the size of a conversational agent significantly affects the subjects on the affective level. This was evidenced by the results that the subjects' felt affective dominance could be significantly influenced by manipulating the simulated proximity level of an agent character. The results also give further support to the finding also discussed based on the results of paper III that by manipulating the affective wording of the speech messages significant differences in experienced affective valence can be found. Furthermore, in the experiment presented in paper VI significant effects of affective contents on perceived intimacy were found. An interesting interaction effect was found when the effects of affective contents of the speech on subjective arousal were analyzed as a function of proximity level: only within the personal proximity level the affective content significantly affected experienced arousal. The results also showed that the subjects would like to interact with a conversational agent on the personal or social simulated proximity level, if they were given a free choice.

Simulated proximity level did not have a significant effect on experienced intimacy in the present experiment. Based on previous research (e.g. Hall, 1966; Reeves and Nass, 1996) it is likely that in human-human communication the corresponding effect exists. In the current experiment, the possibility exists that the affective contents of the speech overrode the simulated proximity levels in the subjects' evaluations of intimacy so that they primarily made the evaluations based on the affective contents of the speech. Another possibility is that there really is a fundamental difference between experiences of proximity in real and (non-immersive) virtual human-human interaction situations. For example, in human-computer interaction, there is usually no possibility of physical threat, which exists in human-human interaction situations. In evolution, humans have evolved to avoid unnecessary physical threats. In addition, despite the advances in affective computing, the vast majority of current computers are not capable of sensing the user very closely (e.g. using machine vision), what could also cause a mismatch between human-human and human-computer interaction. Thus, the experiences of proximity could also be stronger if the user knows that the computer possesses some human-like sensing and reasoning skills, and is capable of reacting in real-time according to information perceived of the user. In other CASA research, social phenomena from human-human communication have been also observed in human-computer interaction, often using minimal cues, such as text boxes (e.g. Nass *et al.*, 1995). The results of paper VI suggest that the case might be somewhat different for experiences of proximity. The sense of intimacy is related to the way people use space and thus it may be hard to communicate it using visual representations on the computer, at least on two-dimensional displays. It is possible that communicating proximity-related intimacy requires other or stronger cues than those used in human-human interaction. On the other hand, advanced perceptual and expressive computer intelligence could be sufficient to change the users' impressions of the intimacy of the interaction.

The results of paper VI have some clear implications for designing interactive conversational systems. It was shown that felt dominance could be significantly affected by using different simulated proximity levels. In most studies on emotions and human-computer interaction the third major affective dimension, dominance, has not been studied. By broadening the view and also taking felt dominance (i.e. the feeling of control) into account in human-computer interaction, the overall user experience could be improved in the future. Felt dominance could also be an important factor in other systems than those with conversation agents. When interacting with a conversational agent, a reasonable amount of user dominance is likely to be desirable. This can be inferred from the results that the subjects of this experiment chose the personal or the social proximity level as the preferred level for everyday communication with a conversational agent. In the previous results presented in this thesis, it was shown that the subject's emotional state could be systematically manipulated. The current results imply that felt dominance could be systematically manipulated by the system by adjusting the simulated proximity level during interaction. The agent could appear to be in control of the interaction, for example, when the agent has to command the user to do something. Then a close agent proximity could help in achieving the desired effect. In conjunction with manipulating dominance using proximity cues, the user's experienced valence and intimacy could also be consciously affected by the designer by altering the affective content of the speech.

Further studies could use other possible cues than agent size for communicating proximity to the user. For example, the agent figure could appear black-and-white, dimmed or less detailed, when it is simulated to be at a far proximity. In addition, the volume or emphasis (e.g. shouting vs. whispering) of the speech could be different at different levels of desired proximity. As suggested by Dehn and van Mulken (2000), there are also a lot of parameters not related to proximity (e.g. gender, appearance, realism), whose effects when used in interactive characters are still not studied thoroughly, not to mention, for example, the use of different facial expressions and gestures in conversational agents (Cassell, 2000).

5.5 GENERAL DISCUSSION

In this thesis, I have presented and discussed the results of six research publications. But how do the results relate to the developments in the field of affective human-computer interaction? There were three major themes in the contributions of this thesis. The first theme was presenting the results on two human physiological signals as affective input methods. New information on pupil size and emotions was provided with significantly better measurement technology and stimuli than in the earlier studies, which were published over 30 years ago. Facial expressions, namely the activations of the two muscles activated in smiling and frowning, their computerized analysis, and relationship in experienced affect were studied in detail in papers IV and V. It was found that the activations of these muscles have systematic relations to experiential affective valence, and they could be used in real time to mediate information about the user's emotions to the computer during human-computer interaction.

The second theme was studying the effects of emotion during interactive human-computer interaction. It was found that the users respond affectively to events in human-computer interaction on three levels: the experiential, the physiological level, and the behavioral level. It was also suggested that using affective interventions during system problems in interactive problem solving, the users' emotions and the associated task behavior can be changed in the positive direction by using affective interventions.

The third theme was the manipulation of the subjects' emotions. Using systematically selected stimuli we showed that people respond on experiential affective dimensions (valence, arousal, dominance) mostly in a predictable way. It was also shown that using affective words in synthetic speech is an effective and predictable way of inducing positive or negative emotion in the user. Further, it was shown that positive effects could be achieved with well-designed affective interventions during human-computer interaction. In computer output, it was showed that the subjects' affective experiences and physiological responses could be systematically manipulated using affective sounds and pictures, as well as speech interventions and the simulated proximity of a conversational agent.

How do the choices in experimental methodology made in the current research compare to other research published in this area? In conducting experiments, especially experiments involving emotions, the methodological choices are important in determining the validity of the research. For studying emotions in the context of utilizing them in human-computer interaction, Picard *et al.* (2001) have listed five important decisions in the design of affective human-computer interaction experiments:

1) Subject-elicited vs. event-elicited emotion,
2) Lab setting vs. real world,
3) Emotion expression vs. feeling,
4) Open recording vs. hidden recording, and
5) Emotion purpose vs. other purpose.

In the experiments described in this thesis, attention has been paid to experimental design. First, we decided to use event-elicited and stimulus-elicited emotion. Subject-elicited emotion can be a problematic approach: the researcher instructs the subject to pose an emotion, but in many cases it has been proven to be difficult to produce felt emotions. Instead, we used stimulus materials, which had been thoroughly studied previously in many experiments, an affective event - mouse delays - which was successful based on previous research (Klein *et al.*, 2002), and carefully designed affective speech messages.

Second, in our experiments, we used a laboratory with controlled conditions. Especially in the physiological recordings, a controlled environment and a controlled experimental setup are essential. However, in evaluating the possibilities of using affective measurements as a part of human-computer interaction it has to be realized that the stimuli used in this experiment were prepared affective sounds, and the subjects listened to them in an isolated laboratory environment. By using controlled experimental conditions, we adopted a basic research approach in order to study the phenomena related to affective human-computer interaction precisely. The findings from basic research form a steady basis for more applied research in the future. In human-computer interaction with real tasks outside the laboratory, an irritating or pleasant computer use situation could provoke even stronger emotions than observed in our experiments, because the situation would probably be meaningful to the person. It is also very likely that with audiovisually integrated stimulus materials, for example, video clips with sound, the emotional responses would be enhanced during human-computer interaction and could also result in improved estimation results.

Third, in our papers I-V, we studied both emotion expression (facial expressions) or physiological responses (the pupillary response) and subjectively felt emotions. In Paper III, affect-related behavior was also studied in terms of problem-solving speed after affective events. Thus, our experiments are designed according to the ideas of, for example, Öhman and Birbaumer (1993) who emphasize that the different manifestations of emotions (i.e. experiential, physiological, and behavioral) need to be taken into account to obtain a realistic picture of affect-related phenomena. Fourth, in the experiments with physiological measurements, for practical reasons, the subjects were aware that they were being measured. However, the subjects were told cover stories about the measurements in order to avoid changes in the subjects' affective behavior due to the knowledge that they can affect the results with voluntary expressions. After the EMG experiments, many subjects told that they forgot about the facial electrodes as soon as the experiment started. Fifth, the subjects knew that the subjective experiential data was used for analyzing their emotion-related experiences. However, the emotion purpose was not emphasized, and in the experiments described in papers I-III, the experiential ratings were given after the physiological measurement phase, and the subjects were unaware of the emotion-purpose until then. In papers IV-VI, it was essential that the subjects rated their subjective experiences during the experiment. In these experiments the affective rating dimensions were explained to the subjects before the experiment.

In order to compare the affective input signals studied in this thesis, to other affective measures available, there is no clear answer, which measures would be ideal for affect analysis in future human-computer interaction. Pantic and Rothcrantz (2003) describe an ideal affect analyzer as multimodal, robust and accurate, generic, sensitive to the dynamics of affect expression, and context-sensitive. In the current thesis, signals were studied in a controlled manner to gain information, which will form the basis for fusing different signals in the future to accomplish multimodal affect estimation. Context sensitivity was also studied in paper III, in which the system gave affective interventions in the context of device problems.

In the current state of affective computing research, it would be especially important to find robust, accurate, and generic (i.e. independent of variability between subjects) methods for emotion estimation. It was already suggested that there are future challenges associated with both the signals studied in this thesis (e.g. compensating for the light reflex and cognitive effects in pupil size variation measurements, and masking and faking of facial expressions) and the other possible measures for human-computer interaction (e.g. the diversity in the results regarding heart rate and emotions). It seems that there is no affective input signal clearly superior to the other methods. In order to recognize affective valence the possibilities in the light of this thesis and the current research are facial expressions, speech prosody, heart rate, and measurements of brain activity (e.g. using EEG). Of these facial expressions, and speech prosody are channels, which people naturally use in everyday human-human communication. In a classic study, Mehrabian (1968b) found that in human-human communication, valence judgments are formed mostly based on facial expressions (55%) and vocal features (38%), and only little based on spoken word (7%). These results highlight the importance of facial expressions in human-human communication. On the other hand, some potential methods do not have a central meaning in human-human communication, but could be well suited to human-computer interaction. For example, heart rate can also be measured unobtrusively during computer use by using a special chair (Anttonen and Surakka, 2005).

In order to measure arousal, there is a wider repertoire of methods available. Because affective arousal is manifested in the autonomic nervous system, all the autonomic nervous system correlates (e.g. pupil size, respiration, skin conductance, skin temperature, blood volume pressure) are potential indicators of affective arousal. Other alternatives include again, for example, speech prosody, brain activity, and facial expressions. Many of these methods also have their own problems. Facial expressions are more directly related to changes in valence than changes in arousal (e.g. Greenwald *et al.*, 1989; Ward, 2004). Speech prosody analyses can only be continuously used in conversational systems, in which speech is the main input modality. Most of the methods associated with autonomic nervous system activity, as well as brain activity measurements, require physical contact with the user using sensors. Pupil size is the only signal that indexes physiological autonomic nervous system arousal and can be measured unobtrusively using visual methods. Based on the results of papers I and II, the future success of pupil size will largely depend on the development of the associated signal analysis methods.

Also, in a more general context it seems that the future success in developing affect analysis methods are affected by both the robustness of the physiological phenomena underlying the methods (how robustly emotions are manifested in the measured signals) and the development of the associated digital signal processing methods. In this thesis, it was shown that both facial expressions and pupil size as physiological signals have clear connections to human affective information processing. Based on this, facial expressions and pupil size have potential to become robust and widely used input signals in human computer interaction.

A very important factor to consider in the development of future affective systems is acceptability. According to Nielsen (1993) important factors contributing to system acceptability include social acceptability, cost, reliability, usability and utility. Social acceptability is especially crucial, because emotions have traditionally been regarded as part of people's personal lives, which they can choose to share or not to share with other people. In affective systems, a feature, which gives the users a possibility not to share their affects with the computer, is probably required. Another crucial question is how people will react to the devices needed in the affect measurements. Probably the most readily acceptable will be devices, which do not invade the user's personal space physically (e.g. video cameras). However, acceptable systems could also be created using sensor technology, which requires physical contact with the user. Electrode technology has advanced rapidly and electrodes are getting smaller, thinner and wireless. Furthermore, electrodes can be mounted on other computer input devices, for example, mice and keyboards (Takahashi, 2004), which are in contact with the user in any case. The price of the technology needed for the affective technology has gone down, and it is possible to achieve affective human-computer interaction, for example, with ordinary desktop video cameras.

Another important aspect will be reliability. Good results in emotion recognition have been obtained by fusing different signals (Collet *et al.*, 1997; Lisetti and Nasoz, 2004) and also some of the experimental results of papers IV and V of this thesis based on measurements of facial muscle activations. The methods for emotion recognition are developed in many projects worldwide, and the reliability of emotion recognition is very likely to be better in the future. In the case of affective human-computer interaction, usability is also very much linked to the reliability of recognition, and the appropriateness of the emotional expressions by the computer. Maybe the most important acceptability criterion for affective technology is, however, the level of utility that the users experience from using the technology. Traditionally, utility has been often thought to be the functionality provided by a piece of software. In the case of affective technology, the utility could be, for example, more natural communication with the computer, increased motivation for learning, or enhanced engagement in a game. Without a clear utility perceivable by the users, affective technologies will not be adopted by the users.

An even bigger question is, whether human-computer interaction should resemble human-human interaction in the future. It has also been suggested that early attempts at integrating human-like interaction styles (e.g. Bob and Clippit) into computing systems have failed because they have not possessed the capabilities the users would expect from an intelligent agent to have (Dehn and van Mulken, 2000). A more intelligent agent has been created by, for example, Cassell (2000) in form of Rea, an embodied conversational real estate agent, which has many of the capabilities that are typically used in human-human conversations, for example, turn-taking and conversation regulation, gestures, postures and, for example, different facial displays related to the dynamics of the conversation. Similarly, a more realistic model of affective behavior could make agents far more realistic and believable. So what will be the role of human-like characters and communication methods in the future? It is almost definitely sure that human-like interfaces or communication channels will not be the only interaction methods with computers. Even Picard (1997) acknowledges that there is no point in bringing emotion or anthropomorphism to each and every application. Instead, a more realistic view would be that in the future there would be a wide repertoire of communication channels with computers, ranging from traditional control with mice and keyboards to affective conversations with human-like characters and everything in between. Which method is used depends on the task to be carried out, and it is also possible that the user can choose a preferred communication method appropriate for each situation, or even use many methods synchronously.

6 Conclusion

This thesis presented research, which was aimed at enhancing the current knowledge on affective human-computer interaction. In addition, I presented a survey on potential input and output methods for affective information, as well as a survey on existing affective systems. In Chapter 5, I discussed the results in more detail, including the advantages and problems related to the chosen methods. The current results were also proportioned to other recent developments in the research field of affective human-computer interaction and some visions for the future were presented.

In all, the studies reported in this thesis suggest that emotions have an important role in human-computer interaction. It was shown that facial expressions can potentially convey information about the user's affective valence, and pupil size can give the computer information about the user's experienced affective arousal. Thus, these two signals can complement each other and they are potential candidates for human-computer interaction input methods. It was also shown that different events during computer use significantly affect the users on the affective level. Further, the results indicated that with carefully studied stimuli or interventions, the user's emotional state could be manipulated in a systematic and meaningful way during human-computer interaction. Even now, the users are emotionally involved and frequently encounter problems - and also moments of satisfaction - when using a computer. Affective information measured from the user could also be interactively utilized to improve the quality of human-computer interaction on an affective level. In the future, by using affective methods in human-computer interaction in a systematic way, more enjoyable user experiences could be achieved, and the worst experiences could be avoided.

The field of affective human-computer interaction is still a young discipline, but it is advancing rapidly. The field offers many possibilities for future research, as there are still many unsolved challenges. Currently, most research work is being carried out in the areas of emotion recognition based on affective measurements and the effects of computerized affect expressions. Very recently more attention has been paid to studying emotions and their usage in real interactive applications. This trend is likely to be strengthened as the number of interactive affective system grows. In the future, however, an especially challenging project is building machine emotional intelligence similar to that of humans, because of the vast number of affective and environmental contexts, in which such systems should operate successfully. Even though the research associated with machine artificial intelligence is very challenging, it seems clear that machines are getting more intelligent all the time in terms of perceptual intelligence and also emotional intelligence. By time, it will be the user that benefits from these developments.

References

ACII (2005). http://www.affectivecomputing.org/2005/ (link checked 09/23/05).

Adini, Y., Moses, Y., and Ullman, S. (1997). Face recognition: the problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 721-732.

Affective Media (2005). http://www.affectivemedia.com/ (link checked 09/23/05).

Andreassi, J. L. (1995). *Psychophysiology: Human Behavior and Physiological Response*. Hillsdale, NJ: Erlbaum.

Anttonen, J. and Surakka, V. (2005). Heart rate and emotions while sitting on a chair. *Proceedings of CHI 2005, Conference on human factors in computing systems*, Portland, OR, ACM Press, 491-499.

Ark, W., Dryer, D., and Lu, D. (1999). The Emotion Mouse. In H. J. Bullinger. and J. Ziegler (Eds.) *Human-Computer Interaction: Ergonomics and User Interfaces*, Lawrence Erlbaum, 818-823.

Aula, A. and Surakka, V. (2002). Auditory emotional feedback facilitates human-computer interaction. *Proceedings of HCI 2002*, London, UK, Springer Verlag, 337-349.

Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*. 70(3), 614-636.

Bartlett, M. S., Littlewort, G., Braathen, B., Sejnowski, T. J., and Movellan, J. R. (2003). A prototype for automatic recognition of spontaneous facial actions. In S. Becker and K. Obermayer (Eds.), *Advances in Neural Information Processing Systems*, Vol. 15. Cambridge, MA: MIT Press.

Bates, J. (1994). The role of emotions in believable characters. *Communications of the ACM*, 37(7), 122–125.

Bernhardt, P. C., Dabbs, J. M., and Riad. J. K. (1996). Pupillometry System for Use in Social Psychology. *Behavior Research Methods, Instruments, & Computers*, 28(1), 61-66.

Bradley, M. M., Greenwald, M. K., Petry, M. C., and Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 18(2), 379-390.

Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavioral Therapy & Experimental Psychiatry*, 25, 49-59.

Bradley, M. M. and Lang, P. J. (1999). *International affective digitized sounds (IADS): Stimuli, Instruction Manual and Affective Ratings*. Technical Report B-2, University of Florida, FL: The Center for Research in Psychophysiology.

Bradley, M. M. and Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology*, 37, 204-215.

Brave, S. and Dahley, A. (1997). inTouch: A Medium for Haptic Interpersonal Communication. *Proceedings of CHI'97, Extended Abstracts*, 363-364.

Brave, S. and Nass, C. (2002). Emotion in human-computer interaction. In J. Jacko and A. Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Mahwah, NJ: Lawrence Erlbaum, 81 - 96.

Brave, S., Nass, C., and Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-computer Studies*, 62(2), 159-306.

Breazeal, C. (2003a). Emotive qualities in lip synchronized robot speech. *Advanced Robotics*, 17(2), 97-113.

Breazeal, C. (2003b). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1-2), 119-155.

Breazeal, C. and Aryananda, L. (2002). Recognizing affective intent in robot directed speech. *Autonomous Robots* 12(1), 83–104.

Bressloff, P.C., Wood, C.V., and Howarth, P.A. (1996). Nonlinear shunting model of the Pupil Light Reflex. *Proceedings of the Royal Society B*, vol. 263, 953-960.

Burleson, W. (2004). Affective Learning Companions. *Doctorial Consortium in conjunction with the 7th International Conference on Intelligent Tutoring Systems*, Maceio - Alagoas, Brasil. Available from http://affect.media.mit.edu/pdfs/04.burleson.pdf (link checked 09/23/05).

Butler, K. A. (1996). Usability Engineering Turns 10. *Interactions*, 3(1), 59-75.

Cahn, J. (1990). The generation of affect in synthesized speech. *Journal of the American Voice I/O Society*, 8, 1-19.

Calvo, M. G. and Lang, P. J. (2004). Gaze Patterns When Looking at Emotional Pictures: Motivationally Biased Attention. *Motivation and Emotion*, 28(3), 221–243.

Campbell, J. P. and Reynolds, D. A. (1999). *Corpora for the Evaluation of Speaker Recognition Systems*, Proceedings of International Conference on Acoustics, Speech, and Signal Processing, IEEE, 2247-2250.

Cannon, W. B. (1927). The James-Lange theory of emotions: A critical examination and an alternate theory. *American Journal of Psychology*, 39, 106-124.

Card, S. K., Moran, T. P., and Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cassell, J. (2000). Embodied Conversational Interface Agents. *Communications of the ACM* 43(4), 70-78.

Coan, J. A., Allen, J. J. B., and Harmon-Jones, E. (2001). Voluntary facial expression and hemispheric asymmetry over the frontal cortex. *Psychophysiology*, *38*, 912–925.

Codispoti, M., Bradley, M. M., Cuthbert, B. N., and Lang, P. J. (2001). Affective reactions to briefly presented pictures. *Psychophysiology*, 38, 474 – 478.

Cohen, I., Sebe, N., Chen, L., Garg, A. and Huang, T. S. (2003). Facial Expression Recognition from Video Sequences: Temporal and Static Modeling. *Computer vision and image understanding,* 91(1-2), 160-187.

Cohn, J. F. and Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2, 1-12.

Cohn, J. F., Schmidt, K., Gross, R., and Ekman, P. (2002). Individual differences in facial expression: stability over time, relation to self-reported emotion, and ability to inform person identification. *Proceedings of the International Conference on Multimodal User Interfaces 2002*.

Collet, C., Vernet-Maury, E., Delhomme, G., and Dittmar, A. (1997). Autonomic nervous system response patterns specificity to basic emotions. *Journal of the Autonomic Nervous System*, 62, 45–57.

Costantini, E., Pianesi, F., and Prete, M. (2005). Recognising emotions in human and synthetic faces: the role of the upper and lower parts of the face. *Proceedings of the 10th international conference on Intelligent user interfaces, IUI 2005*, San Diego, CA, January 2005, ACM Press, 20-27.

Cosmides, L. and J. Tooby (2000), Evolutionary psychology and the emotions. In Lewis, M. and J. M. Haviland-Jones (Eds.) *Handbook of Emotions* 2nd ed. New York: Guilford, 3-134.

Coulson, M. (2004). Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. *Journal of Nonverbal Behavior* 28(2), 117-139.

Dailey, M. N., Cottrell, G. W., Padgett, C., and Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience,* 14, 1158-1173.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Putnam.

Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* New York: Harcourt Brace.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. New York: D. Appleton and Company.

Dehn, D. M. and van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research, *International Journal of Human-Computer Studies,* 52, 1-22.

Dimberg, U, (1990). Facial electromyography and emotional reactions. *Psychophysiology*, 19, 643-647.

Ekman, P. (1982). *Emotion in the human face*. Cambridge: Cambridge University Press.

Ekman, P. (1985). *Telling Lies*. New York: W. W. Norton.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3/4), 169-200.

Ekman, P. (1999). Basic Emotions. In T. Dalgleish and T. Power (Eds.) *The Handbook of Cognition and Emotion*, Sussex, U.K.: John Wiley & Sons, 45-60.

Ekman, P. and Friesen, W. V. (1969). The Repertoire of Non-Verbal Behaviour: Categories, Origins, Usage and Codings. *Semiotics*, 1, 49-98.

Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17, 124-129.

Ekman, P. and Friesen, W. V. (1978). *Facial Action Coding System (FACS): A Technique for the Measurement of Facial Action*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P. and Friesen, W.V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, 6, 238-252.

Ekman, P., Levenson, R. W., and Friesen, W.V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221, 1208-1210.

Essa, I. A. and Pentland, A. P. (1997). Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 757-763.

Fernandez, R. and Picard, R. W. (1998). Signal Processing for Recognition of Human Frustration, *Proceedings of ICASSP98, International Conference on Acoustics, Speech and Signal Processing 1998*. Available from http://vismod.media.mit.edu/pub/tech-reports/TR-447.pdf (link checked 09/23/05).

Frijda, N. (1986). *The Emotions*. Cambridge: Cambridge University Press.

Frøkjær, E., Hertzum, M., and Hornbæk, K. (2000). Measuring usability: Are effectiveness, efficiency and satisfaction really correlated? *Proceedings of CHI 2000, Human Factors in Computing Systems*, ACM Press, 345-352.

Fukayama, A., Ohno, T., Mukawa, N., Sawaki, M., and Hagita, N. (2002). Messages embedded in gaze of interface agents - impression management with agent's gaze. *Proceedings of CHI 2002, Human Factors in Computing Systems*, ACM Press, 41-48.

Gilleade, K. and Allanson, J. (2003). A Toolkit for Exploring Affective Interface Adaptation in Videogames. *Proceeding of HCI International 2003*, volume 2. Mahwah, NJ: LEA, 370-374.

Greenwald, M. K., Cook, E. W. III, and Lang, P. J. (1989). Affective judgment and psychophysiological response: dimensional covariation in the evaluation of pictorial stimuli. *Journal of Psychophysiology*, 3, 51-64.

Hall, E. T. (1966). *The Hidden Dimension*, New York: Doubleday Press.

Hasbro Furby (2005). http://www.hasbro.com/furby/ (link checked 09/23/2005).

Hazlett, R. (2003). Measurement of User Frustation: A Biologic Approach. *Proceedings of CHI'03, Conference on human factors in computing systems*, *extended abstracts,* April 2003, Fort Lauderdale, FL, 734-735.

Healey, J. (2000). *Wearable and Automotive Systems for the Recognition of Affect from Physiology*. PhD Thesis – Massachusetts Institute of Technology, Electrical Engineering and Computer Science Department.

Healey, J. and Picard, R. (2005). Detecting Stress During Real-World Driving Tasks. Accepted for publication in *IEEE Transactions on Intelligent Transportation Systems.*

Hess, E. H. (1972). Pupillometrics. In N. S. Greenfield and R. A. Sternbach (Eds.) *Handbook of Psychophysiology*. New York, NY: Holt, Richard & Winston, 491-531.

Hess, E. H., and Petrovich, S. B. (1987). Pupillary behavior in communication. In A. W. Siegman and S. Feldstein (Eds.) *Nonverbal Behavior and Communication*. Hillsdale, NJ: Erlbaum, 327-348.

Hietanen, J. K, Surakka, V., and Linnankoski, I. (1998). Facial electromyographic responses to vocal affect expressions. *Psychophysiology*, 35(5), 530-536.

Hollnagel, E. (2003). Is affective computing an oxymoron? *International Journal of Human-Computer Studies*, 59(1-2), 65-70.

Hyönä, J., Tommola, J., and Alaja, A-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, 48A, 598-612.

Ilves, M. and Surakka, V. (2004). Subjective and physiological responses to emotional content of synthesized speech. In Magnenat-Thalmann, N., Joslin, C., and Kim, H. (Eds.) *Proceedings of CASA 2004, 17th International Conference on Computer Animation and Social Agents*, 19-26.

Isen, A. M. (2000). Positive affect and decision making. In Lewis, M. and Haviland, I. (Eds.) *Handbook of emotions*. New York: Guilford Press, 417-435.

Ito, T. A., Cacioppo, J. T., and Lang, P. J. (1998). Eliciting affect using the international affective picture system: trajectories through evaluative space. *Personality and social psychology bulletin*, 24(8), 855-875.

Izard, C. E. (1977). *Human Emotions*. New York: Plenum Press.

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review*, 100, 68-90.

James, W. (1884). What is an emotion? *Mind,* 9, 188-205.

Janisse, M. P. (1974). Pupil size, affect and exposure frequency. *Social Behavior and Personality*, 2, 125-146.

Jeffrey, P. (1998). Personal space in a virtual community, *Proceedings of CHI'98, Human factors in computing systems, conference summary*, ACM Press, 347-348.

Kahneman, D. and Beatty, J. (1966). Pupillary diameter and load on memory. *Science*, 154, 1583-1585.

Kappas, A., Bherer, F., and Thériault, M. (2000). Inhibiting facial expressions: Limitations to the voluntary control of facial expressions of emotion. *Motivation and Emotion*, 24, 259-270.

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

Kleinginna, P. R. and Kleinginna, A. M. (1981). A categorized list of emotion definitions with suggestions for a consensual definition. *Motivation and Emotion*, 5, 345-379.

Konecni, V. J., Libuser, L., Morton, H., and Ebbesen, E. B. (1975). Effects of a violation of personal space on escape and helping responses. *Journal of Experimental Social Psychology*, 11, 288-299.

Lang, P. J. (1993). The Three-System Approach to Emotion. In N. Birbaumer and A. Öhman (Eds.), *The Structure of Emotion*. Seattle, WA: Hogrefe and Huber Publishers, 3-17.

Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm A. O. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30, 261-273.

Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *American Psychologist*, 50, 372-385.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1990). Emotion, Attention, and the Startle Reflex. *Psychological Review*, 97, 377-395.

Lang, P.J., Bradley, M.M., and Cuthbert, B.N. (1995). *International affective picture system (IAPS): Technical Manual and Affective Ratings.* Gainesville, FL. The Center for Research in Psychophysiology, University of Florida.

Larsen, J. T., Norris, C. J., and Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over *zygomaticus major* and *corrugator supercilii*. *Psychophysiology*, 40, 776-785.

LeDoux, J. E. (2000). Emotion Circuits in the Brain. *Annual Review of Neuroscience, 23*, 155-184.

Levenson, R. W. (1992). Autonomic nervous system differences among emotions. *Psychological Science*, 3, 23-27.

Levenson, R. W., Ekman, P., and Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, 27(4), 363-384.

Lien, J. J.-J., Kanade, T., Cohn, J. F., and Li, C.-C. (1998). Subtly Different Facial Expression Recognition and Expression Intensity Estimation. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, June 1998, 853-859.

Lisetti, C. L. and Nasoz, F. (2004). Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Applied Signal Processing*, 11, 1672-1687.

Lisetti, C. L. and Schiano, D. J., (2000). Automatic facial expression interpretation: where human-computer interaction, artificial intelligence and cognitive sciences intersect. *Pragmatics and Cognition*, 8(1), 185-235.

Loewenfeld, I. E. (1966). Pupil size. *Survey of Opthalmology*, 11, 291-294.

Loewenstein, G. and Lerner J. S. (2003). The role of affect in decision making. In R. Davidson, H. Goldsmith, and K. Scherer (Eds.) *Handbook of Affective Science*, Oxford: Oxford University Press, 619-642.

Lundqvist, L.-O. and Dimberg, U. (1995). Facial expressions are contagious. *Journal of Psychophysiology,* 9, 203-211.

Mattingly, I. G. and Liberman, A. M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman (Ed.), *Auditory Function*, Chichester: Wiley, 775-793.

McFarland R. A. (1985). Relationship of skin temperature changes to the emotions accompanying music. *Biofeedback and selfregulation*, 10, 255-267.

McNeese, M. D. (2003). New visions of human-computer interaction: making affect compute, *International Journal of Human-Computer Studies*, 59(1-2), 33-53.

Mehrabian, A. (1968a). Inference of attitudes from the posture, orientation, and distance of a communicator. *Journal of Consulting and Clinical Psychology*, 32, 296-308.

Mehrabian, A. (1968b). Communication without words. *Psychology Today*, 2(9), 52-55.

Mehrabian, A. and Russell, J.A. (1974). *An approach to environmental psychology.* Cambridge, MA: M.I.T. Press.

Mota, S. (2002), *Automated Posture Analysis For Detecting Learner's Affective State*, M. Sc. Thesis, Massachusetts Institute of Technology, September 2002.

Murphy, S. T. and Zajonc, R. B. (1993). Affect, cognition, and awareness: affective priming with suboptimal and optimal stimulus. *Journal of Personality and Social Psychology*, 64, 723-739.

Nass, C. (2004). Etiquette equality: exhibitions and expectations of computer politeness. *Communications of the ACM*, 47(4), 35-37.

Nass, C., Fogg, B. J., and Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies,* 45(6), 669-678.

Nass, C., Foehr, U., Brave, S., and Somoza, M. (2001). The effects of emotion of voice in synthesized and recorded speech. *Proceedings of the AAAI Symposium Emotional and Intelligent II: The Tangled Knot of Social Cognition,* 91-96.

Nass, C. and Lee, K. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*(3), 171-181.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43, 223-239.

Nass, C., Steuer, J. S., and Tauber, E. R. (1994). Computers Are Social Actors. *Proceedings of CHI'94, Conference on Human Factors in Computing Systems*, ACM Press, 72-78.

Nielsen, J. (1993). *Usability Engineering*. Boston, MA: Academic Press.

Oatley, K. and Johnson-Laird, P. N. (1987). Towards a cognitive theory of the emotions. *Cognition and Emotion*, 1, 29-50.

Öhman, A. (1987) The psychophysiology of emotion: An evolutionary-cognitive perspective: In P.K. Ackles, J.R. Jennings, and M.G.H.Coles (Eds.) *Advances in Psychophysiology*. Vol. 2. Greenwich, CT: JAI Press. 79-127.

Öhman, A. and Birbaumer, N. (1993). Psychophysiological and Cognitive-Clinical Perspectives on Emotion: Introduction and Overview, In N. Birbaumer and A. Öhman (Eds.) *The Structure of Emotion*, Seattle: Hogrefe and Huber Publishers, 3-17.

Öhman A, Flykt A, and Esteves F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal Experimental Psychology: General,* 130, 466-478.

Oliver, N., Pentland, A., and Berard, F. (2000). LAFTER: a real-time face and lips tracker with facial expression recognition. *Pattern Recognition*, 33, 1369-1382.

Ortony, A. and Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, 97, 315-331.

Osgood, C.E. (1952). The Nature and Measurement of Meaning. *Psychological Bulletin*, 49, 197-237.

Oudeyer, P.-Y. (2003) The production and recognition of emotions in speech: features and algorithms. *International Journal in Human-Computer Studies*, 59(1-2), 157-183.

Oviatt, S. L. (2002). Multimodal interfaces. In J. Jacko and A. Sears (Eds.) *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, Mahwah, NJ: Lawrence Erlbaum, 286-304.

Paiva, A., Costa, M., Chaves, R., Piedade, M., Mourão, D., Sobral, D., Höök, K., Andersson, G., and Bullock, A. (2003). SenToy: an Affective Sympathetic Interface. *International Journal of Human Computer Studies*, 59(1-2), 227-235.

Pantic, M. and Rothkrantz, L. J. M. (2000). Expert system for automatic analysis of facial expressions. *Image and Vision Computing Journal*, 18(11), 881-905.

Pantic, M. and Rothkrantz, L. J. M. (2003). Towards an Affect-Sensitive Multimodal Human-Computer Interaction. *Proceedings of the IEEE*, 91(9), 1370-1390.

Partala, T., Aula, A. and Surakka, V. (2001). Combined voluntary gaze direction and facial muscle activity as a new pointing technique. *Proceedings of INTERACT 2001*, IOS Press, 100-107.

Picard, R.W. (1997). *Affective Computing*. Cambridge, MA: M.I.T. Press.

Picard, R. W. and Klein, J. (2002). Computers that Recognise and Respond to User Emotion: Theoretical and Practical Implications. *Interacting with Computers*, 14(2), 141-169.

Picard, R.W. and Scheirer, J. (2001). The Galvactivator: A glove that senses and communicates skin conductivity. *Proceedings of 9th International Conference on Human-computer interaction*, New Orleans, LA. Available from: http://vismod.media.mit.edu/pub/tech-reports/TR-542.pdf (link checked 09/23/05).

Picard, R. W., Vyzas, E., and Healey, J. (2001). Towards Machine Emotional Intelligence: Analysis of Affective Physiological State, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(*10*), 1175-1191.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik and H. Kellerman (Eds.) *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*. New York: Academic, 3-33.

Pope, L. K. and Smith, C. A. (1994). On the distinct meanings of smiles and frowns. *Cognition and Emotion*, 8, 65-72.

Prendinger, H., Mori, J., and Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human-Computer Studies*, 62 (2), 231-245.

Qi, Y., Reynolds, C., and Picard, R. W. (2001). The Bayes Point Machine for Computer-User Frustration Detection via PressureMouse. *Proceedings of PUI 2001, Workshop on Perceptive Interfaces*, ACM Digital Library (http://portal.acm.org/dl.cfm).

Ravaja, N., Saari, T., Salminen, M., Laarni, J., Holopainen, J., and Järvinen, A. (2004). Emotional response patterns and sense of presence during video games: potential criterion variables for game design. *Proceedings of NordiCHI'04, the third Nordic conference on Human-computer interaction*, 339 - 347.

Reeves, B. and Nass, C. (1996). *The Media Equation: How people treat computers, Television, and New Media Like People and Places*. Cambridge: CSLI Publications.

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

Ross, E.D. (1981). The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology,* 38(9), 561-569.

Sakurazawa, S., Yoshida, N., and Munekata, N. (2004). Entertainment Feature of a Game Using Skin Conductance Response. *Proceedings of ACE 2004, Advances in Computer Entertainment Technology*, ACM Press, 181-186.

Scheirer, J., Fernandez, R., Picard, R. W. (1999). Expression Glasses: A Wearable Device for Facial Expression Recognition. *Proceedings of CHI'99, Conference on Human Factors in Computing Systems, Extended Abstracts*, 262-263.

Scherer, K. R., Banse, R., Wallbott, H. G., and Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, 15(2), 123-148.

Scherer, K. R., Ladd, D. R., and Silverman, K. E. A. (1984). Vocal cues to speaker affect: testing two models. *Journal of Acoustical Society of America* 76(5), 1346-1356.

Schmidt, K. L. and Cohn, J. F. (2001). Human Facial Expressions as Adaptations: Evolutionary Questions in Facial Expression Research. *Yearbook of Physical Anthropology,* 44, 3-24.

Shneiderman, B. (2002). *Leonardo's laptop*. Cambridge, MA: MIT Press.

Siemens CX70 Emoty. (2005). http://www.siemensmobile.com/cx70emoty (link checked 09/23/2005).

Simms, T. M. (1967). Pupillary response of male and female subjects to pupillary differences in male and female picture stimuli. *Perception and Psychophysics*, 6, 115-121.

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review*, 74, 29-39.

Sloman, A. (1999). Review of Affective Computing. *AI Magazine*, 20(1), 127-137.

Surakka V. (1998). Contagion and Modulation of Human Emotions. *Acta Universitatis Tamperensis, 627*, Ph.D. Thesis, University of Tampere, 1998.

Surakka V. and Hietanen J. K. (1998). Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology*, 29, 23-33.

Surakka, V., Illi, M., and Isokoski, P. (2004). Gazing and frowning as a new technique for human-computer interaction. *ACM Transactions on Applied Perception,* 1(1), 40-56.

Sutcliffe, A. (2002). Multimedia User Interface Design. In J. Jacko and A. Sears (Eds.) *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications.* Mahwah, NJ: Lawrence Erlbaum, 245-262.

Sylwester, R. (1994). How Emotions Affect Learning. *Educational Leadership* 52(2), 60-65.

Takahashi, K. (2004). Remarks on Emotion Recognition from Bio-Potential Signals. *Proceedings of the 2nd International Conference on Autonomous Robots and Agents*, Palmerston North, New Zealand, December 2004, 186-191.

Tian, Y.-I., Kanade, T., and Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 97-115.

Toyota POD (2005). http://www.toyotaoffroad.com/Articles/Toyota/POD/toyota_pod.htm (link checked 09/23/2005).

Wang, H., Prendinger, H., and Igarashi, T. (2004). Communicating emotions in online chat using physiological sensors and animated text. *Proceedings of CHI'04: Conference on Human Factors in Computing Systems,* Vienna, Austria, 2004, 1171-1174.

Ward, R. (2004). An analysis of facial movement tracking in ordinary human–computer interaction. *Interacting with Computers*. 16(5), 879-896.

Wundt, W. (1896). *Grundriss der Psychologie*. Leipzig: Wilhelm Engelmann.

Vrana, S. (1993). The Psychophysiology of Disgust: Differentiating Negative Emotional Contexts with Facial EMG. *Psychophysiology*, 30, 279-286.

Zajonc, R. B. (1980). Feeling and thinking: preferences need no inferences. *American Psychologist*, 35, 151-175.

Paper I

Partala, T., Jokiniemi, M., and Surakka, V. (2000). Pupillary Responses to Emotionally Provocative Stimuli. In *Proceedings of ETRA 2000, Eye Tracking Research and Applications Symposium 2000*, Palm Beach Gardens, FL, November 2000, ACM Press, 123-129.

Available online at:
`http://doi.acm.org/10.1145/355017.355042`
(requires ACM Digital Library subscription)

Paper II

Partala, T. and Surakka, V. (2003). Pupil Size Variation as an Indication of Affective Processing. *International Journal of Human Computer Studies, 59, 1-2*, 185-198.

Available online at:
http://dx.doi.org/10.1016/S1071-5819(03)00017-X
(requires subscription)

Paper III

Partala, T. and Surakka, V. (2004). The effects of affective interventions in human-computer interaction. *Interacting with Computers, 16, 2,* 295-309.

Available online at:
http://dx.doi.org/10.1016/j.intcom.2003.12.001
(requires subscription)

Paper IV

Partala, T., Surakka, V., and Vanhala, T. (in press). Real-time estimation of emotional experiences from facial expressions. *Interacting with Computers*.

Paper V

Partala, T., Surakka, V., and Vanhala, T. (2005). Person-independent estimation of emotional experiences from facial expressions. In *Proceedings of the 10th International Conference on Intelligent User Interfaces, IUI 2005*, San Diego, CA, January 2005, ACM Press, 246-248.

Available online at:
`http://doi.acm.org/10.1145/1040830.1040883`
(requires ACM Digital Library subscription)

Paper VI

Partala, T., Surakka, V., and Lahti, J. (2004). Affective effects of agent proximity in conversational systems. In *Proceedings of NordiCHI 2004*, Tampere, Finland, October 2004, ACM Press, 353-356.

Available online at:
http://doi.acm.org/10.1145/1028014.1028070
(requires ACM Digital Library subscription)