PHYSIOLOGY-BASED AFFECT RECOGNITION AND ADAPTATION IN HUMAN-MACHINE INTERACTION

By

Changchun Liu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

May, 2009

Nashville, Tennessee

Approved:

Professor Nilanjan Sarkar

Professor George E. Cook

Professor Mitch Wilkes

Professor Richard Shiavi

Professor Zachary E. Warren

TALBE OF CONTENTS

ACKNOWLEDGEMENTS

To begin with I would like to express sincere appreciation to Dr. Nilanjan Sarkar - my academic advisor and the chair of my dissertation committee. I would like thank him for his invaluable support, encouragement, and advices throughout this study. His innovative ideas and thought-provoking discussions made research a great experience. I am grateful to the members of my dissertation committee, Dr. Richard Shiavi, Dr. George Cook, Dr. Mitch Wilkes, and Dr. Zachary Warren for their guidance. I appreciatively acknowledge NSF grant IIS-0107775, NASA grant NAS5-98051, the Vanderbilt University Discovery grant, and the MARI grant that supported the research described in this dissertation.

I would like to thank my teammates Pramila and Karla. They played the vital roles in this work and their critical thinking and immaculate approach were of great help. I also want to thank my colleagues Uttama, Yu, Bibhrajit, Furui, Naim, Jadav, and Vishnu at Robotics and Automations Systems Laboratory for their support.

Last but definitely not the least, my deep gratitude, appreciation, indebtedness and thanks go to my family: My mother, my father, my sisters, and my soul mate - Mingwei. Thanks for their belief in me all the time, their understanding and encouragement. They have been by my side at every crossroad of life helping me take the right decision. I would like to give my very special thanks to Mingwei. Her invaluable support and continual encouragement kept driving me till the end. Without her boundless love, this work could not be happened.

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION AND SUMMARY

1. Introduction

Recent advances in robotics and intelligent systems are expected to usher in a new era where smart autonomous systems will have significant impact on our daily lives. As machines and people begin to co-exist and cooperatively share a variety of tasks, the need for machines to "understand" humans becomes increasingly important. Human interactions are characterized by explicit as well as implicit channels of communication. While the explicit channel transmits overt messages, the other one transmits implicit messages about the communicator. Ensuring sensitivity to the other party's emotions is a significant part in human communication and is associated with the second, implicit channel (Cowie et al., 2001). "The latest scientific findings indicate that emotions play an essential role in rational decision-making, perception, learning, and a variety of other cognitive functions (Picard, 1997)." It is well documented in the literature of psychology that a person's affective state is very important in relationships (York, et al., 1984). Understanding others' affective states and behaving responsively are determining factors in human communication. It has also been shown that people's interactions with computers, TV and similar machines/media are fundamentally social and natural, just like interactions in real life (Reeves and Nass 1996). Therefore, endowing machines with a degree of emotional intelligence should permit smoother, natural, and more efficient human-machine interaction (HMI).

The motivation for this work stems from the observation that despite tremendous advancements in the field of human-machine interaction, truly affect-sensitive systems that can communicate implicitly and intuitively with humans have still remained largely unexplored. As "Sensibility" is typically defined as awareness of and responsiveness toward something (as emotion in another), the objective of this work is to investigate the following hypotheses: (i) It is possible to detect the affective states of interest by using multiple indices derived from physiological signals in real-time; (ii) Such recognized affective cues can be integrated within a machine's control architecture to make it capable of responding to them appropriately; (iii) The proposed affect-sensitive systems are expected to improve the overall human-machine interaction experience.

The potential applications of computer/robotic systems that can detect a person's affective (emotional) states and interact with him/her based on such perception are varied and numerous. Whether it is the domain of personal home robotic assistants that assist in cleaning and transportation, computerized tutors that engage students and permit a more enjoyable and productive discourse, service robots that act as aids in offices, hospitals, and museums – this novel aspect of human-machine interaction will impact them all.

In this work, the impacts of affective-sensitivity were investigated in both human-robot interaction (HRI) and human-computer interaction (HCI) contexts. We also experimentally demonstrated that, by endowing the robot/computer with the capability of affect recognition and adaptation, it is feasible to augment human-machine interaction systems to be used in the autism intervention by accommodating the individual needs and affective characteristics.

This chapter is organized as follows: in Section 2, a detailed description of past and ongoing research in the field of affective computing is presented. Here, the key modalities currently used for affect-recognition - facial expressions, vocal intonation, gestures/posture and physiology have been surveyed. Section 3 elaborates on the field of affective human-machine interaction that includes the review of efforts in both HRI and HCI studies and the key observations. Special emphasis was laid on the physiology-based affective communication. Section 4 presents the current state-of-the-art in applying interactive technology in autism intervention. The needs of computer/robotic systems that are capable of understanding and responding to the affective characteristics of the children with Autism Spectrum Disorders (ASD) are discussed. Finally Section 5 summarizes the backgrounds and contributions of the research based on 6 manuscripts.

2. Affective Computing

Affective computing can be defined as "computing that relates to, arises from, or deliberately influences emotions" (Picard, 1997). "Affective computing," according to Picard, includes various research themes, such as giving a computer/robot the ability to recognize and express emotions and developing its ability to respond intelligently to human emotion.

An affective or emotionally intelligent system is expected to possess a two-fold capability - express its own emotions in a manner understandable to humans and perceive emotions in humans. Research on machine-emotion synthesis has resulted in expressive robots and computerized agents that can articulate their emotions using human-like facial expressions and affective speech. This is a significant step towards making social machines that would permit natural HMI. However, in most such systems, there is no real understanding of human emotions and the synthesized emotions are mostly triggered by a limited rule-base. While these intelligent agents can smile, appear angry, confused or utter affective words, they are not capable of detecting human psychological states such as anxiety, frustration, engagement or boredom and reacting to them. The focus of our work is complementary to this body of research by addressing the later capability, i.e., how to endow a computer/robotic system with the ability to recognize human affective states and adapt its behavior appropriately based on such perception.

The need for a computer/robot to understand human emotion was discussed in (Picard, 1997; Simon, 1979; Sloman and Croucher, 1981). It can be argued that if an intelligent system can recognize a person's affective states implicitly, and can infer the cause of these states as related to the task, the human-machine interaction could achieve a different dimension. Such a capability, alone or in conjunction with other capabilities that allow explicit instructions from a human, is expected to provide a new paradigm for human-machine interaction.

According to Mehrabian (1969), 55 percent of the emotional meaning of a message is carried through facial expression, posture, and gestures, another 38 percent of meaning comes from tone of voice, and only about 7 percent of the emotional meaning of a message is communicated through explicit verbal channels. In psychophysiology research, there is good evidence that the physiological activity associated with emotions can be differentiated and systematically organized (Bradley, 2000). The correlations between physiological responses and underlying affective states have been investigated in various studies (Bradley, 2000; McCraty, et al., 1991; Sinha, et al., 1992; Picard, 1997).

The focus of this section is on the four above-mentioned modalities for detecting affect automatically – facial expression, speech, gesture/posture, and physiology. Each modality is reviewed with regard to the existing methodologies, classification accuracies, recent research developments, and their advantages and disadvantages.

2.1 Facial Expression

It is well known that people use facial expression as a powerful and instantaneous means to convey their emotions, intentions, and opinions while interacting socially. In recent years, there has been an upsurge of interest in the research problems of machine analysis of facial expressions (Edwards et al., 1998; Pantic and Rothkrantz, 2000; Sebe, et al., 2002). The problem of automatic recognition of human affective states from images of faces includes three sub-problem areas: finding faces, detecting facial features, and classifying these data into some affect categories.

Current studies assume, in general, that the presence of a face in the scene is ensured. Posed portraits of faces (uniform background and good illumination) constitute input data processed by the majority of the current systems. While head-mounted camera system was proposed in (Pantic and Rothkrantz, 2000) and much progress has been recently made in the development of vision systems for robust face detection (Yang, et al., 2002), most presently existing systems for facial affect recognition do not perform automatic face detection in an arbitrary scene. The features used are typically based on local spatial position or displacement of specific points and regions of the face, unlike the approaches based on audio (as described in Section 2.2), which use global statistics of the acoustic features. According to Pantic and Rothkrantz (2002), most of the proposed approaches to affect recognition via facial expressions tend to be static, analytic, or/and based on 2-D facial feature extraction. Most existing systems classify facial expression into one of the six known "basic" emotion categories (except for the automated system proposed by Pantic (2001), which categorizes facial expressions based on user-defined classes). The classification techniques used by the existing systems include: template-based methods

(Edwards et al., 1998), Fuzzy Logic (Pantic and Rothkrantz, 2000), artificial neural networks (ANN) (Kobayashi and Hara, 1992), and Bayesian Learning (Sebe, et al., 2002). The automated systems could achieve an accuracy of 64% to 98% when detecting three to seven emotions deliberately displayed by 5 to 40 subjects. This is a significant achievement since the accuracy of humans in detecting six basic emotional facial expressions is 70%-98% (Bassili, 1979).

Some of the limitations of the existing methods include: context of the situation and user personality are not taken into account when recognizing emotion from facial expression, temporal analysis of facial expressions to differentiate between emotions are not incorporated in existing facial recognition techniques, and simplifying assumptions are made to tame the problem of facial feature detection (e.g., images contain portraits of faces with no facial hair or glasses, the illumination is constant, the subjects are young and of the same ethnicity). Apart from the above mentioned limitations, the other shortcomings of existing systems are: inability to deal with head motions and inability to handle situations when part of the face is obstructed or the view is sideways instead of frontal.

2.2 Vocal Intonation

Vocal intonation can effectively measure the affective (emotional) content of speech. Speech consists of words spoken in a particular way. If we consider the verbal part (strings of words) only, without regarding the manner in which it was spoken, we might miss important aspects of the pertinent utterance and even misunderstand the spoken message by not attending to the nonverbal aspect of the speech. In simpler terms, "how we say" something is as important as "what we say" and the former may reinforce, replace, or contradict the latter.

In contrast to spoken language processing, which has witnessed significant advances, the processing of affective content in speech has only been explored recently. An important aspect of vocal affect analysis is extracting discriminatory auditory features from speech that would be useful in determining the affective content of the audio signal. The research in psycholinguistics showed a high correlation between some statistical measures of speech and the affective states of the speaker (Heuft, et al., 1996; Cowie and Douglas-Cowie, 1998; Amir and Ron, 1998; Iida et al., 1998; Murray and Arnott, 1993). The auditory features usually estimated from the input audio signal are: (i) pitch-related measures (the fundamental frequency of the acoustic signal determined by the rate at which vocal cords vibrate); (ii) intensity (the vocal energy); (iii) speech rate (the number of words spoken in a time interval); and (iv) voice quality (influenced by speech rate as well as harmonic structure). Table 1 shows a summary of human vocal qualitative characteristics associated with the emotions as reported by Murray and Arnott (1993). These are listed in relation to the neutral voice.

Table 1. Summary of human vocal affects described relative to neutral speech

	Anger	Happiness	Sadness	Fear	Disgust
Speech Rate	slightly faster	faster or slower	slightly slower	much faster	very much slower
Pitch Average	very much higher	much higher	slightly lower	very much higher	very much lower
Pitch Range	much wider	much wider	slightly narrower	much wider	slightly wider
Intensity	higher	higher	lower	normal	lower
Vocal Quality	breathy	blaring	resonant	irregular	grumbled

The audio signals are usually classified into different affective states based on the above features. Virtually all the existing work on automatic vocal affect analysis performs singular classification of input audio signals into a few emotion categories (Cowie et al., 2001). Utilized pattern recognition techniques include: ANNs (Nakatsu, et al., 2000), HMMs (Kang, et al., 2000), Gaussian mixture density models (Li and Zhao, 1998), Fuzzy membership indexing (Amir and Ron, 1998), and maximum-likelihood Bayes classifiers (Kang, et al., 2000). Scherer (1996) performed a large-scale study using 14 professional actors and reported that human ability to recognize emotions from purely vocal stimuli is about 60%. Automated vocal affect analyzers match this accuracy when recognizing two to eight emotions deliberately expressed by subjects recorded while pronouncing sentences having a length of 1 to 12 words. However, it should be noted that in many instances strong assumptions are made to make the problem of automating vocal affect analysis more tractable. For example, the recordings are noise free, the recorded sentences are short, delimited by pauses, and carefully pronounced by nonsmoking actors to express the required affective state. Overall, the test data sets are small (one or more words or one or more short sentences spoken by few subjects) containing exaggerated vocal expressions of affective states. In conclusion, it can be said that research in the area of affective vocal analysis has resulted in techniques that work well with structured, noise-free, small data sets but do not scale well to real-life situations where long continuous sentences would need to be analyzed.

2.3 Posture/Gesture

In addition to facial expressions and vocal intonation, people also rely on body gestures and postures to detect and express affective state. Ambady and Rosenthal (1992) found out that humans rely on the combined visual channels of face and body more than any other channel when they make judgments about human communicative behavior. Coulson (1992) presented experimental results on attribution of six emotions (anger, disgust, fear, happiness, sadness and surprise) to static body postures by using computer-generated figures. From his experiments he concluded that human recognition of emotion from posture is comparable to recognition from the voice, and some postures are recognized as well as facial expressions. Burgoon et al. (2005) discussed the issue of emotion recognition from bodily cues and provided useful references in the context of national security. Table 2 presents a list of body gesture cues as described by Coulson (1992) and Burgoon et al. (2005) together with the associated emotion categories.

Several methodologies were used for automatically detecting postures. Many researchers have used cameras as input devices. However, variations in lighting, background conditions, camera, subject positions, and subject appearance can complicate posture recognition via vision. To circumvent these problems, some investigators have used sensors such as switches, accelerometers, or pressure sensors mounted on a chair (Evreinov, 1999). As shown in Table 2, hand tracking is important for recognizing the gesture. One of the methods of tracking the movements of hand is via a glove which is equipped with a number of sensors that provide information about hand position, orientation, and flex of the fingers (Piekarski and Thomas, 2002). It was reported that using sensor gloves can deliver more accuracy in pose determination than using camera (Pavlovic et al., 1995).

Table 2. Summary of the body gesture and associated affective states

Anxiety	Hands close to the table surface; fingers moving; fingers tapping on the table
Anger	Body extended; hands on the waist; hands made into fists and kept low
Disgust	Body backing; left/right hand touching the neck or face
Fear	Body contracted; body backing; hands high up, trying to cover bodily parts
Happiness	Body extended; hands kept high; hands made into fists and kept high
Uncertainty	Shoulder shrug; palms up

Despite the importance of spontaneous gesture in normal human interactions, automatically deciphering the affective content of gestures has just begun attracting research attention. Mota and Picard (2003) have shown that the postures can be used to recognize the affective state of interest. In the above study, nine postures were found to be reliably recognized by humans, and an automated pattern recognition system was built to detect these postures. The system achieved an average accuracy of 87.6% when tested with data not included in the training set. In a recent work (Silva et al., 2006) a gesture description mechanism was proposed to capture the most expressive instant of gestures in terms of limb-to-torso distances and overall expansion of the boy in the frontal, lateral, and vertical dimensions. A Hidden Markov Models (HMM) based affective model was developed that can recognize the participants' emotions with a prediction rate of over 79%.

2.4 Physiology

While facial expressions, vocal intonation, and postures are discernible by people, physiological changes due to emotions are mostly unobservable. For instance, changes in blood pressure and heart rate cannot be directly observed. However, there is good evidence in Psychophysiology that the physiological activity associated with affective state is differentiated and systematically organized (Bradley, 2000). Indicators of peripheral physiological activity (e.g., indicators of autonomic activity such as cardiovascular activity, electrodermal activity, and electromyogram activity) have long been viewed as having significant potential to be used as convergent, or even primary, indicators of emotion (Lang, 1979; Lazarus, 1968). Autonomic measures have often been used with considerable success as indicators of emotional arousal (Smith, 1989).

Facial expressions and vocal intonation are the most popularly investigated modalities for affective computing. However, most existing systems based on the above two modalities are limited by the following issues: (i) they work under highly constrained conditions. For instance, facial recognition assumes availability of high resolution, frontal face images with no occlusion or motion, and vocal intonation requires few carefully spoken, noise free words with exaggerated affect expression; (ii) they show high accuracy for within-individual classification but poor performance across individuals; (iii) since both these modalities are voluntary, their application is constrained to situations when there is an overt expression of emotion (affective states such as fatigue, boredom and anxiety are not easily detectable); (iv) most systems are computationally expensive and not designed for real-time affect recognition. Following the above two modalities closely are physiology-based techniques and gesture/posture recognition. Gesture/posture-based affect recognition is still in primitive stages and currently there are very few systems that use this methodology. On the other hand, psychophysiological techniques are fast gaining ground as a potent means of affect recognition. It has been generally accepted that physiological modality circumvents several limitations of the vision and speech based methods (Picard, 1997). One of the chief advantages of using physiology is that physiological signals are continuously available and are not dependent on overt emotion expression. Hence, physiology-based affect detection could be very useful in situations where it is not possible to continuously monitor facial expressions of a person or in scenarios where a person's speech is not available to interpret his/her underlying emotion. In addition, physiology is usually not under voluntary control and hence provides an undiluted assessment of the underlying affective state. It is also reasonably independent of cultural, gender and age related biases (Bradley 2000). Furthermore, physiological data may offer an avenue for recognizing aspects of affect that may be less obvious for humans but more suitable for computers. Specifically, a computer system may be able to quickly implement signal processing and pattern recognition tools to infer underlying affective states that a human could not. Unobtrusive, small, wireless physiological sensors may be an ideal solution in these cases for real-time affect monitoring (Picard, 1997; Jafari et al., 2005; Wijesiriwardana et al., 2004).

This section discusses detecting affective states from the physiological signals. It first describes the popular emotion theories and their implications. Then, well-researched physiological measures that are powerful indicators of emotions are discussed. After that, a review of state-of- the-art in physiology-based affect recognition is presented. At last, the need for active learning for physiology-based affective computing is discussed.

2.4.1 Theories on Emotions

There are several theories on the psychophysiology of emotion, resulting from varied schools of thought. Two prominent approaches are discussed here – discrete emotions approach and dimensional emotion approach.

Discrete emotion theories claim the existence of historically evolved basic emotions, which are universal and can therefore be found in all cultures. Several psychologists have suggested a different number of these basic emotions, ranging from 2 to 18 categories, but conventionally "basic" emotions usually refer to the following six: anger, disgust, fear, happiness, sadness and surprise. Several arguments for the existence of these categories have been provided, like distinct universal facial signals, distinct universals in antecedent events, presence in other primates etc (Ekman, 1992).

Dimensional emotion theories use dimensions rather than discrete categories to describe the structure of emotions. According to a dimensional view, all emotions are characterized by their valence and arousal. Valence measures the degree to which the emotion is positive or negative, and arousal measures the strength of the emotion (Bradley, 2000). Some models have suggested an even greater number of dimensions (e.g. control), but arousal and valence have proved to be the two main dimensions in (Russell, 1980). When Russell started conducting self-report studies on the structure of emotion with the two-dimensional approach, he discovered a specific ordering of the words describing the felt emotions. The ratings did not fall in every area of the coordinate system, but instead clustered around the periphery of a circle. He called the resulting configuration the Circumplex of affect (Russell, 1980).

There is no general agreement on which of these theories is correct. One of the main empirical evidences of discrete emotion theory is based on the investigation of facial expression of emotions (Ekman, 1992). However, there is no agreement on a set of basic emotions. Even the criteria for choosing one set rather than another are not agreed upon, and in fact, focusing on different aspects of emotion notoriously tends to produce different lists (Cowie, et al., 2001). Furthermore, there is evidence that the affective state could be an aggregate of various affective categories at different arousal levels (Vansteelandt et al., 2005). The arousal-valence approach was questioned for its capability of being sufficient to differentiate equally between all emotions and classifying the emotions in a 2-D space defined by arousal and valence space (AV space). Cowie et al., (1999) found that some emotions that share the same degrees of arousal and valence but are perfectly distinguishable in everyday life (e.g. fear and anger). These universal views (discrete emotions approach and dimensional emotion approach) have been challenged by Russell (1994), who found that there are differences in the definition of emotions and recognition ability for subjects of different origins.

Considering the effort that has been devoted to the description of emotions, the lack of convergence suggests that there may well be no natural units to be discovered. It has been increasingly accepted in human-machine interaction society that, to develop an automatic analyzer of human affective feedback, pragmatic choices (e.g., application- and user-profiled choices) must be made regarding the selection of affective states to be recognized (Cowie, et al., 2001; Pantic and Rothkrantz, 2003).

2.4.2 Emotions and Physiology

While there is no common agreement on specific emotions being related to specific physiological patterns, it is generally agreed on that there is a significant correlation between them (Bradley, 2000). The Autonomic Nervous Systems (ANS), also known as the involuntary nervous system, controls actions that one does not have conscious control over. The ANS controls smooth muscle, gland activity, and cardiac muscle. It is this system and its control over physiological responses that is of interest in the study of emotions. The ANS is further divided into two branches. The sympathetic nervous system (SNS) has the dominant function in emergency situations or "fight or flight" situations. The parasympathetic branch (PNS) is the relaxed activity controller. The PNS promotes body maintenance such as food digestion. Mostly organ systems are dually innervated by the SNS and PNS. In the sections that follow, some of the widely used physiological measures are described here - cardiovascular activity, electrodermal activity (EDA), and electromyogram (EMG) activity.

Four important indicators of cardiovascular activity are commonly used by psychophysiologists: electrocardiogram (ECG), impedance cardiogram (ICG), photoplethysmogram (PPG), and phonocardiogram (PCG)/heart sound. ECG measures the heart activity through the electrical signal of the heart muscle. The PPG signal measures changes in the volume of blood in the finger tip associated with the blood volume pulse (BVP) cycle, and it provides an index of the relative constriction versus dilation of the blood vessels in the periphery. ICG analysis measures the impedance or opposition to the flow of an electric current through the body fluids. The heart sound signal measured sounds generated during each heartbeat. Several studies have investigated the relationship of the above-mentioned indicators of cardiovascular activity with underlying emotional states. Research has shown that positive emotions lead to alterations in heart rate (HR) variability, such as anger results in a sympathetically dominated power spectrum, while appreciation causes a power spectral shift toward MF (mid-frequency) and HF (high-frequency) activity (McCraty, et al., 1991). Heart rate has been found to be closely correlated with arousal and heart rate acceleration varies most consistently with stimulus arousal – increasing as arousal (pleasant or unpleasant) increases (Cook III, et al., 1991). Sinha et al. have showed that Sadness produced a characteristic pattern with moderate increase in blood pressure and vascular resistance compared with changes during neutral emotional state (Sinha, et al., 1992). Previous research has validated BVP measured at fingers as a measure of anxiety in response to a threat of physical harm. Results indicate that BVP is sensitive to the stress manipulation during both low and high stress periods and is correlated with self-reported anxiety (Bloom and Trautt, 1978). The cardiovascular activity has been used together with EMG to examine the positive and negative affective states of people (Cacioppo et al., 2000).

Electrodermal activity (EDA) consists of two main components - tonic response and phasic response. Tonic skin conductance refers to the ongoing or the baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event related changes that occur, caused by a momentary increase in skin conductance (resembling a peak superimposed on tonic skin conductance). EDA is closely linked with psychological concepts of emotion, arousal, and attention. Individual differences in phasic responses can be reliably associated with psychopathological states when being interpreted in context of the stimulus conditions that elicit EDA responses (Christie, and Friedman, 2002). EDA has been shown to be associated with task engagement and its arousal is produced by social stimulation that invokes stress tension, anxiety, or cognitive reactions. (Pecchinenda and Smith, 1996). It is also known that smaller values of EDA were associated with neutral states than with sadness, anger, fear, disgust, and amusement (Christie and Friedman, 2002).

Electromyogram (EMG) activity measures the electrical activity in the muscle during contraction. The EMG signal from corrugator supercilii muscle (eyebrow) captures a person's frown and detects the tension in that region, and the EMG signal from the zygomaticus major muscle captures the muscle movements while smiling. Upper trapezius muscle EMG activity measures the tension in the shoulders, one of the most common sites in the body for developing stress. Facial displays (frowns, grimaces, smiles etc.) of affective reactions are obvious overt behaviors associated with expression of emotions (Dimberg, 1990). Facial EMG activity has been used to examine the positive and negative affective states of people (Cacioppo et al., 2000). The Corrugator Supercilii muscles have been considered as a measure of distress (Ekman and Friesen, 1986). When viewing affective slides it was observed that Corrugator muscle activity increases for unpleasant stimuli and decreases for highly pleasant stimuli and there is a significant linear relationship between pleasantness and Zygomaticus EMG activity (Lang et al. 1993).

2.4.3 Physiology-based affect recognition

Although there have been published efforts in the field of Psychology and Human Factors that aimed at finding physiological correlates with small sets of emotions, most have focused on the analysis of variance comparisons and combining data over many subjects, where each was measured for a relatively small amount of time (Picard, et al., 2001). Recent studies have shown that emotions play an important role in a variety of cognitive processes and it would be useful to develop affect-sensitive machines that can understand human emotions (Picard, 1997). These findings have caused a splurge of researches in the area of human-machine interaction to apply more sophisticated pattern recognition techniques to physiology-based affective computing in order to develop affect recognizers that are capable of automatically inferring the affective states of humans working with machines.

Table 3. Summary of Affect-Recognition Systems Based on Physiological Signals

Author	Emotion Elicitation Method	Emotions Elicited	N	Measures	Affect Modeling Technique	Results
Picard, '01	computer controlled prompting system	Neutral, anger, hate, grief, platonic love, romantic love, joy, and reverence	1	Masseter EMG, BVP, (HR), EDA, and respiration.	Floating Forward Search (SFFS), Fisher Projection (FP),	49.4% and 50.6%correct classification for SFFS and FP, respectively
Kim '04	Visual and auditory stimuli	Sad, stressed, angry, and surprised	50	ECG, EDA, and SKT	Support Vector Machines (SVM)	78.4% (3 emotions) and 61.8% (4 emotions)
Regan, '07	Video games	Arousal-valence	12	EDA, EMG, and cardiovascular activity	Fuzzy logic	3% and 6% different from manual approach for both arousal and valence.
Kulic, '03	picture-based system	Arousal-valence	36	BVP, EDA, and corrugators EMG	Fuzzy logic	94% for arousal and 80% for valance
Scheirer, '02	A slow computer game interface	Frustration	5	EDA and BVP	HMM	64.7% correct classification.
Ark, '99	Instructed to show emotions via facial expressions	Anger, fear, sad, disgust, happy and surprise	8	SKT, EDA, and HR	Discriminant function analysis (DFA)	66.0% correct classification.
Nasoz, '03	Segments of movie	Sad, anger, fear, surprise, frustration, and amusement	3	SKT, EDA, and HR	KNN, DFA, and Marquardt Backpropagation (MB)	71.6% (KNN), 64.8 (DFA), and 83.7% (MB)

Table 3 summarizes state-of- the-art in developing affect recognizers from the physiological signals by using machine learning techniques. Some of the theoretical and practical challenges involved in physiology-based affect recognition are:

[1] What was the method of eliciting emotions?

[2] How many and which affective states were detected?

[3] How many and which physiological signals were used?

[4] How many participants were involved?

[5] Which machine learning techniques were used?

[6] What results were achieved?

It can be seen from Table 3 that physiology-based approach holds promise to detect emotions of a person interacting with a computer/robot. These works further substantiated the foundlings in psychophysiology literature that the physiological responses are closely correlated with the underlying affective states. Most systems have attempted affect-recognition using cardiovascular activities, (e.g., heart rate and blood volume pulse), EMG signals, and skin conductance. These signals are easy to record and analyze and have a well-understood cognitive basis. The approach of discrete emotions has been adopted by most of the works, where the goal was to determine which of the *n* target emotions was present. On an average 60%-80% accuracy has been achieved in distinguishing between 4-5 target affective states. It should be noted that none of these systems attempt to distinguish between varying levels of a single affective state. It is more challenging to predict the levels of arousal of a target emotion (for instance low/medium/high frustration) instead of determining discrete emotions (for instance anger, joy, sadness etc.). While Arousal-valence approach has been used in a few recent research works, no explicit modeling results were reported that can reliably transform the arousal and valence into the affective states of interest. On the other hand, as suggested by (Cowie, et al., 2001; Pantic and Rothkrantz, 2003), it is critical to choose target affective states based on application- and user-profiled considerations and thus develop an automatic analyzer of human affective feedback accordingly for a specific human machine interaction task. Another challenge for affective moedling is phenomenon of person-stereotypy (e.g., different individuals expressing the same emotion differently under same contexts), which makes it difficult to obtain universal patterns of emotions across individuals (Lacey and Lacey, 1958). This suggests that individual-specific approach should be applied in order to accommodate the differences encountered in emotion expression and an intensive study on each individual is demanded.

2.4.4 Active learning for physiology-based affective computing

To develop physiology-based affective models, it is a prerequisite to obtain a training set consisted of the physiological features and the corresponding subjective reports. However, in practice the experimental data measurement and labeling can be both time consuming and expensive. For example, in order to build physiology-based affective-models, a human-machine interaction task usually has to be performed for 2-5 minutes to get the physiological signals and subjective reports have to be collected afterward (which may interrupt the interaction) in order to get one labeled training data point (Liu, et al., 2008; Picard, Vyzas, and Healey, 2001; Rani, et al., 2006). As stated in (Picard, et al., 2004), more efficient machine learning paradigms are demanded for affective computing that depends relatively less on the availability of a large set of training examples.

A strategy to tackle the problem is to use active learning. Most established method for developing physiology-based affective models is passive learning, where the samples are randomly drawn from the dataset during the training (Liu, et al., 2008; Mandryk and Atkins, 2007; Nasoz, et al., 2003; Picard, Vyzas, and Healey, 2001; Rani, et al., 2006). There is no relation between the expected error rate and the training samples.

On the other hand, the search for effective training data sampling algorithms has been studied in the machine learning research. Active learning methods have been developed to reduce the dependence on the large and costly dataset by identifying informative samples for training (Cohn, Atlas, and Ladner, 1994). It has been shown that only a small portion of a large unlabeled data set may need to be labeled to train an active learner that achieves strong classification performance (Lewis and Catlett, 1994; Cohn, Atlas, and Ladner, 1994; Dagan and Engelson, 1995). Thus, active learning is an appealing tool for affective modeling in human-machine interaction. However, till date no published study has been found that specifically investigated this approach for physiology-based affective computing.

3. Human-Machine Interaction

Human-computer interaction (HCI) and human-robot interaction (HRI) are two main research subfields belonging to human-machine interaction study. Kiesler and Hinds (2004) argued that these two areas share many similarities and HRI should be based on HCI on which much research has been conducted. On the contrary, Fong, et al., (2001) addressed the difference between HCI and HRI as shown in Table 4.

Table 4. Comparison between HCI and HRI

HCI	HRI
• Controlled by men • 2 dimensions • Simple • Static User Model • Fixed or Portable • Mostly, vision & audio	• Autonomy • 3 dimensions • Complex • Dynamic User Model • Movable • Vision, Audio, and Tangibleness • Face to face • Learning and Decision

In recent years, information technology and robotics have made commendable progress towards making sophisticated, efficient, and multi-functional machines and ushered in many new areas of application (e.g., education, entertainment, personal assistance, rehabilitation, and search and rescue etc.). However, current intelligent systems are not fully reliable and smart enough to do such complex jobs without any human help. There is a real need in foreseeable future to synergistically combine various capabilities of computer/robotic systems with the human's intelligence and cognitive task understanding so that together they can address many of the current goals of various applications. To achieve this objective, it is important to make the machines more easily accessible to and intuitively operated by humans. However, one of the major stumbling blocks in deploying human-machine teams in these complex and unstructured task domains is that many of the existing computer/robotic systems are impervious to the psychological states of the human they work with due to the lack of any implicit channel of communication between the two.

The focus of this section is to present the review of the research investigation of the affective human-robot interaction and affective human-computer interaction, as well as the key observations. Special emphasis was laid on the physiology-based affective communication.

3.1 Affective Human-Computer Interaction

Affect-sensitive human-computer interaction has been the focus of research in recent years. There have been increasing numbers of studies that involve assessing the affective states of humans while interacting with computer applications. In the context of intelligent tutoring system, there have been efforts that aim at endowing a computerized tutor with the ability to adapt affectively in the teaching-learning process, which would permit a more natural, enjoyable and productive discourse. Conati (2002) proposed a probabilistic model to monitor a user's emotion and engagement during automated tutoring. The affective states of students (i.e., reproach, shame, and joy) were detected by the use of eye brow EMG, GSR and ECG through a dynamic decision network. The tradeoff between engagement and learning was achieved by a utility function that assigned appropriate weights to students' performance and engagement. Kapoor et al. (2001) present preliminary work done in the area of developing a Learning Companion, a computer-based system that can detect the affective aspects of a learner. Predinger et al. (2005) conducted an experimental study that examined GSR and EMG to investigate the effect of a life-like virtual teacher on the affective state of users under "affective persona" and "non-affective persona" conditions. Subtle expressivity was achieved by different body gestures and varying linguistic styles, which enabled the educational agent to affectively respond to the user's performance. In the field of computer games, the concept of Affective Gaming has been recently proposed by Gilleade et al. (2005), which aims at enhancing gaming experience by adapting the game course to the player's affective state. Three categories of high-level design heuristics were defined for developing such a gaming system: "Assist me," where games detect player's negative feeling and respond by providing assistance; "challenge me," where games detect the user's arousal and adapt the difficulty accordingly; and "emote me," where games provoke the intended emotions by tracking and utilizing the relationship between user's emotional response and specific game content.

It can be safe to conclude that physiology-based affect recognition is being increasingly explored for developing affect-sensitive interaction between humans and computers. However, most existing physiology-based affect recognition systems are limited by the following issues: (i) While affect recognition with high prediction performance were achieved in laboratory conditions (e.g., the participants watch pictures or videos), few works have developed affective models that can reliably predict the affective states of humans when they are involved in the natural real-world applications (e.g., less quantitative modeling results were reported in the affective auto-tutoring and gaming studies); (ii) Affective modeling works were done off-line and none of them have been validated in a real-time application; (iii) While the importance of including human in the loop of interaction has been increasingly recognized, current studies have the limitation of lacking systematic experimental investigation of the impacts of the affect-sensitive closed-loop human-machine interaction; (iv) Most works used across-individuals approach and did not did not consider person stereotypy. No existing systems have been found that are capable of determining the intensity of an affective state from the physiological signals.

3.2 Affective Human-Robot Interaction

There has been a steady progress in the field of intelligent and interactive robotics over the last two decades. Research on robot-emotion synthesis has resulted in expressive robots that can articulate their emotions using human-like facial expressions and affective speech. Some prominent examples of such robots are - Pong robot developed by the IBM group (Haritaoglu et al., 2001), Kismet and Leonardo developed in MIT (Breazeal, 2000; Hoffman and Breazeal, 2004), and ATR's Robovie-IIS (Kanda et al., 2002). Emotional intelligence in robots has found a powerful application in the toy industry. Robot pet such as the enormously popular Sony's Aibo robot dog is one such example. Many more socially interactive robots are being developed as personal robots, entertainment toys, and therapy assistants and to serve as test beds to validate social development theories. It is apparent that providing robots with synthetic emotions has received more attention than making them understand human emotions.

While there is significant progress towards making social robot to express its own emotions in a manner understandable to humans, till date there are very few human-robot interaction systems available in which real-time physiology-based feedback is utilized by a robot to interpret the underlying psychological state of the human and modify/adapt its (robot's) behavior as a result. Only two recent studies have investigated this aspect of human-robot interaction. Itoh et al. (2006) developed their own bioinstrumentation system to measure human stress when interacting with a fixed humanoid robot that had only an upper body. Their wearable system measured ECG, respiration, EDA (changes in skin resistance), pulse wave transit time, blood pressure, and upper body movements. If participants' stress level increased past a certain threshold then the robot would modify its actions to decrease participants' stress levels by shaking the participants' hand. Another study that utilized psychophysiological measures to evaluate how participants respond to robotic behaviors was performed by Kulic and Croft (2007). The level of arousal was employed as an indicator of user comfort with the robot, which was estimated by using multiple physiological indices, such as heart rate, skin conductance, and corrugator supercilii EMG. The results indicated that participants had lower arousal responses with the safe planned motions of the robotic manipulator arm and felt calmer when the robot motions were slower. Participants tended to show strong physiological responses to fast robotic arm movements. The eliminated arousal was utilized to modulate the robot velocity and behavior during interaction. The results suggested that physiological signals could provide useful information and add a level of perceived safety for humans interacting with robots. The experiments in (Itoh et al. 2006) relied heavily on inter beat interval (IBI) derived from ECG to measure the activity of the sympathetic (LF-HRV) and parasympathetic (HF-HRV or RSA) divisions of the ANS and thus derive the index of anxiety level. The affective modeling in (Kulic and Croft 2007) was achieved by using fuzzy logic. Enabling robots to perceive emotions in humans is a nascent research field and needs to be further explored. The preliminary results showed that physiology-based affect recognition and adaptation hold the promise to permit more natural human-robot interaction. However, current works (Itoh et al. 2006; Kulic and Croft 2007) did not validate of the developed affect inference mechanisms (e.g., examining the prediction accuracy) when they were used in the real-time applications. Furthermore, no systematical investigations were performed to evaluate how such affect recognition and adaptation impact on the users' perceived experiences when they are interacting with an affect-sensitive robot.

4. The Use of Interactive Technology in Autism Intervention

Autism Spectrum Disorders (ASD) are characterized by core deficits in social reciprocity and communication (DSM-IV-TR, American Psychiatric Association, 2000). Interventions often focus on social-problem solving and social skills training, so that participants can gain experience and exposure to various situations representative of everyday living. The ultimate goal of such interventions is for some generalization of these skills to carry over into real-life situations. Despite the urgent need and societal import of intensive treatment, appropriate intervention resources for children with ASD and their families are often extremely costly when accessible (Tarkan, 2002). Therefore, an important new direction for research on ASD is the identification and development of assistive therapeutic tools that can make application of intensive treatment more readily accessible. In response to this need, a growing number of studies have been exploring the application of advanced interactive technologies for future use in interventions to address the social deficits of children with ASD, namely computer technology (Bernard-Opitz, 2001), virtual reality (VR) environments (Parsons and Mitchell, 2002; Standen and Brown, 2005), and robotic systems (Dautenhahn and Werry, 2004; Michaud and Theberge-Turmel, 2002; Pioggia et al., 2005).

Various software packages have been developed and applied to address specific deficits associated with autism, e.g., understanding of false belief (Swettenham, 1996), attention (Trepagnier et al., 2006), expression recognition (Silver and Oakes, 2001), and social communication (Bernard-Opitz et al., 2001; Parsons et al., 2005).

The application of VR to the intervention of ASD is a relatively new area of research, which has gained ground only in the last 10 years. Initial results indicate that with the controllable complexity of a virtual world with minimized distractions, VR systems may allow for simplified but embodied social interaction that is less intimidating or confusing for children with ASD (Moore et al., 2000; Standen and Brown, 2005). A VR system that allows role-play in a setting designed to model the real world holds potential for children with ASD learning a range of appropriate social and behavioral skills (Parsons and Mitchell, 2002). Such a system could potentially allow for the manipulation and exacerbation of salient characteristics of interactions in a highly flexible environment that could potentially scaffold skills while minimizing potentially negative consequences. While changing and controlling surroundings may be beyond the scope of real-world interventions, such modifications are convenient and easily reproduced within VR environments.

Different from using computer software or VR environments, the interaction between children with ASD and physical robots during the intervention contributes important real-time and embodied characteristics of face-to-face social interaction among humans. Dautenhahn have explored how a robot can become a playmate that might serve a therapeutic role for children with autism in the Aurora project. Research suggested that children with ASD like to interact with reactive robots than the inanimate toys (Dautenhahn and Werry, 2004). Michaud and Theberge-Turmel (2002) investigated the impact of robot design on the interactions with children and emphasized that systems need to be versatile enough to adapt to the varying needs of different children. Pioggia et al. (2005) developed an interactive life-like facial display system for enhancing emotion recognition in individuals with ASD. Robots have also been used to teach basic social interaction skills using turn-taking and imitation games, and the use of robots as social mediators and as objects of shared attention can encourage interaction with peers and adults (Dautenhahn and Werry, 2004; Kozima, et al., 2005). Robotic technology poses the advantage of furnishing robust systems that can support multimodal interaction and provide a repeatable, standardized stimulus while quantitatively recording and monitoring the performance progress of the children with ASD to assess the intervention approaches (Scassellati, 2005).

By employing human-machine interaction technologies, computer/robot-based therapeutic tools can partially automate the time-consuming routine behavioral therapy sessions and may allow intensive intervention to be conducted at home (Dautenhahn and Werry, 2004).

Even though there is increasing research in applying interactive technologies in autism intervention, no published studies were found that specifically addressed how to automatically detect and respond to affective cues of children with ASD. However, such ability could be critical given the importance of human affective information in human machine interaction (Fong et al., 2003; Picard, 1997) and the significant impacts of the affective factors of children with ASD on the intervention practice (Seip, 1996). For example, an experienced therapist generally continuously monitors the affective cues of the children with ASD and adjusts the course of the intervention accordingly: ‘likes and dislikes chart’ is recommended to record the children’s preferred activities and/or sensory stimuli during interventions that could be used as reinforcers and/or ‘alternative behaviors’ (Seip, 1996); children with autism are particularly vulnerable to anxiety and intolerant of feelings of frustration, which requires a therapist to plan tasks at an appropriate level of difficulty (Ernsperger, 2003); the engagement of children with ASD is the ground basis for the ‘floor-time therapy’ to help them develop relationships and improve their social and communication skills (Wieder and Greenspan, 2005). Similarly, a computer/robotic systems for autism intervention must also be able to understand the affective needs of these children - an ability that the current ASD intervention assistive systems lack - to achieve effective interaction that addresses the role of affective states in human-machine interaction and intervention practice.

5. Scope and Summary of the Research

We initially performed a systematic investigation of the strengths and weaknesses of these four machine learning methods when being employed for the physiology-based affect recognition. Individual-specific modeling approach was used to account for the phenomenon of person stereotypy, which was capable of delivering competitive prediction on the intensity of affective states. Then, we experimentally investigated the impact of an affect-based dynamic difficulty adjustment on player's interaction with a computer game, which is capable of physiology-based affect recognition and real-time difficulty adjustment in a closed-loop manner. Later, we proposed and implemented a closed-loop affect-sensitive human-robot interaction framework, which is capable of (i) performing accurate real-time affect recognition based on the affective models generated from the data collected before; and (ii) modifying robot's behaviors accordingly. A robot-based basketball game is designed where a robotic "coach" monitors the human participant's anxiety level using the developed affect model and dynamically changes its behavior parameters to allow users' skill improvement while maintaining desired anxiety levels. After that, we investigated the feasibility of modeling the affective states of children with ASD via psychophysiological analysis. Two cognitive tasks were designed and implemented to elicit varying levels of arousal in the target affective states (i.e., liking, anxiety, and engagement) in the children with ASD studied. Multiple subjective reports from an autism therapist, a parent, and the participant were analyzed to account for the suspected unreliability of the subjective self-reports from children with ASD. Then, we proposed and implemented an affect-sensitive robot-assisted autism intervention framework. The affective states of children with ASD were detected via a physiology-based affect recognition technique in real time. A reinforcement learning based behavior adaptation mechanism is employed to enable the robot to adapt its behaviors autonomously as a function of the predicted child's affective state. This is the first time that the impact of affect-sensitive closed-loop interaction between a robot and a child with ASD has been demonstrated experimentally. Finally, we investigated the use of Support Vector Machine active learning to alleviate the costly and time consuming video review and labeling efforts in physiology-based affective modeling for children with ASD while still maintaining sufficient model performance. The research is presented in 6 manuscripts which are given as:

Manuscript 1: An Empirical Study of Machine Learning Techniques for Affect Recognition in Human-robot Interaction

Background

Given the fact that ensuring sensitivity to the other party's emotions plays a significant role in human interactions (Cowie et al., 2001), it should permit more meaningful and natural human-robot interaction when a robot can detect the affective cues of the person it is working with. It has been generally agreed in Psychophysiology that emotions and physiology are closely intertwined and one influences the other (Bradley, 2000). Physiology-based affect recognition has been actively used by several research groups in recent year to relate physiological phenomena/patterns to the underlying emotions (Picard et al., 2001; Conati, 2002; Nasoz et al., 2003; Kim et al., 2004). Researchers have established that it is possible to distinguish between various affective states by using multiple features derived from diverse physiological signals and marked a significant step towards applying machine learning and signal processing technology to the existing practices in psychophysiology. However, some shortcomings of these systems are: (i) The manner of affect elicitation usually involves audio/visual stimuli, or in some cases deliberate emotion expression that do not correspond well with real-life situations; (ii) Most of them distinguish between discrete affective states such as joy, anger, disgust, fear etc. and not between varying levels of arousal within a emotion of interest(for instance distinction between low, moderate and high anxiety); (iii) Relatively less work on the relationship of certain physiological features (such as inter beat interval, spontaneous blink, pulse transit time etc.) with affective states is available. Furthermore, even though several machine learning methods have been successfully employed to build affect recognizers from physiological indices, a systematic comparison of various methods- their strengths and weaknesses has not been possible largely because of the following points of diversity in each study: (i) Definition of emotion – discrete or continuous; (ii) Nature of physiological features used; (iii) Manner of self-reporting to get subjective affective states of participants; and (iv) Baselining techniques used. Given these diversities, it is hard to find a common ground for comparing methods and analyzing their merits to provide more insights in choosing different affect recognition algorithms.

Summary of Contribution

There are three contributions in this work. First is to design and implement two cognitive tasks that successfully elicited varying levels of arousal in the target affective states (i.e., Anxiety, Boredom, Engagement, Frustration, and Anger). The second is to use individual-specific modeling approach to account for the phenomenon of person stereotypy (i.e, within a given context, different individuals express the same emotion with different characteristic response patterns (Lacey and Lacey, 1958)) that was capable of delivering competitive prediction on the intensity of affective states. The third is to perform a systematic comparison of the strengths and weaknesses of these four machine learning methods (K-Nearest Neighbor, Regression Tree, Bayesian Network, and Support Vector Machines) when being employed for the physiology-based affect recognition. The comparison parameters included: classification accuracy, time/space efficiency, model's interpretability, and impact of feature selection. The correlation of the physiological features to the target affective states was also investigated. The results were validated by an experimental study on 15 participants. Manuscript 1 is based on the following papers:

- Rani, P., Liu, C., Sarkar, N., and Vanman, E. *"An Empirical Study of Machine Learning Techniques for Affect Recognition in Human-Robot Interaction"*, Pattern Analysis and Applications Journal, vol. 9, no. 1, pp. 58-69, 2006.
- Liu, C., Rani, P., and Sarkar, N., *"Comparison of Machine Learning Techniques for Affect Detection in Human Robot Interaction"*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Alberta, Canada, 2005, pp. 2662 – 2667.

Manuscript 2: Dynamic Difficulty Adjustment in Computer Games through Real-Time Affective Feedback

Background

With steady progress in recent years, computer game has become one of the most popular and economically successful forms of human-computer interaction (HCI). Static game difficulty levels (i.e. selected by players manually) are not sufficient to avoid getting the player overwhelmed/bored (Koster, 2004) and could be annoying as well as cause interruption of the game play (Chen, 2007). In order to address this issue, dynamic difficulty adjustment (DDA) mechanisms have been proposed to enable the game-playing experiences automatically tailored to the individual characteristics (Hunicke and Chapman, 2004). In most current DDA research works, the performance of the player has been used as a main measure of the characteristics of the players. However, as noted by Pagulayan et al. (2002), unlike productivity software, computer game's paramount evaluation factor should be the affective experience provided by the play environment instead of the user's performance. While high-level design heuristics were proposed in the affective gaming field (Gilleade et al., 2005) and there have been research efforts in developing affect-sensitive computerized tutoring systems (Conati, 2002; Prendinger et al. 2005), the impact on human users when computer game respond to recognized affective states (i.e., interact in a closed-loop manner) is still largely unexplored.

Summary of Contribution

The main contribution of this work is to experimentally investigate the impact of an affect-based DDA on player's interaction with a computer game that is capable of physiology-based affect recognition and real-time difficulty adjustment in a closed-loop manner. The physiological features were extracted on-line as inputs and the reliable real-time prediction performance was achieved to detect the target affective state (i.e., players' anxiety). The state-flow models were utilized to dynamically adjust the game difficulty level based on the performance or detected anxiety level, respectively. This is the first time, to our knowledge, that how the gaming experience can be augmented by using the affect-based DDA is demonstrated though a systematic user study. It was observed that: (i) The perceived anxiety-level was reduced for the majority of the participants during the affect-based DDA session; (ii) The performance of the majority of the participants improved during the affect-based DDA session; (iii) Most participants perceived the game with the affect-based DDA to be more challenging than the one with the performance-based DDA; and (iv) Most participant perceived that the game with the affect-based DDA to be more satisfying than the one with the performance-based DDA. Manuscript 2 is based on the following papers:

- Liu, C., Rani, P., and Sarkar, N., *"Dynamic Difficulty Adjustment in Computer Games through Real-Time Affective Feedback"*, International Journal of Human-Computer Interaction, to be print, 2008.
- Rani, P., Sarkar, N. Liu, C., *"Maintaining Optimal Challenge in Computer Games Through Real-Time Physiological Feedback"*, Chapter 4, Task Specific Information Processing in Operational and Virtual Environments, Foundations of Augmented Cognition, edited by Dylan D. Schmorrow, Lawrence Erlbaum Associates Publishers, pp. 184-192, 2006..

Manuscript 3: Interaction between Human and Robot – An Affect-Inspired Approach

Background

An affective or emotionally intelligent robot is expected to possess a two-fold capability - perceive emotions in humans and express its own emotions in a manner understandable to humans. Research on robot-emotion synthesis has resulted in expressive robots that can articulate their emotions using human-like facial expressions and affective speech (Fong et al., 2003; Breazeal and Aryananda, 2002). However, in most such robots, there is no real understanding of human emotions and the robot emotions are triggered by a limited rule-base. It appears that for intelligent and intuitive human-robot interaction, it is imperative that apart from being capable of synthesizing affect, the robot should be capable of perceiving human affective states and responding to them appropriately to address such perception and to achieve a close-loop interaction. While physiology-based affect recognition has been explored by several research groups (Conati, 2002; Kim et al., 2004; Nasoz et al., 2003; Picard et al., 2001), till date there is no human-robot interaction system available in which real-time physiology-based feedback is utilized by a robot to interpret the underlying affective states of the human and modify/adapt its (robot's) behavior as a result.

Summary of Contribution

The main contribution of this work is to propose and implement a closed-loop human-robot interaction framework, which is capable of (i) performing accurate real-time affect recognition based on the affective models generated from past data; and (ii) modifying robot's behaviors accordingly. A robot-based basketball game is designed where a robotic "coach" monitors the human participant's anxiety level using the affect model generated from past data and dynamically changes its behavior parameters to allow users' skill improvement while maintaining desired anxiety levels. Such an affective feedback based human-robot interaction system was evaluated against a performance feedback based system to compare their effects on the user's performance, perceived challenge, anxiety, and overall experience. Manuscript 3 is based on the following papers:

- Rani, P., Liu, C., and Sarkar, N., *"Interaction between Human and Robot - an Affect-inspired Approach"*, Interaction Studies, vol. 9, no. 2, pp. 230-257, 2008.

- Liu, C., Rani, P., and Sarkar, N., *"Human-Robot Interaction Using Affective Cues"*, 15th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN), Hatfield, United Kingdom, 2006, pp. 285 - 290.
- Liu, C., Rani, P., and Sarkar, N., *"Affective State Recognition and Adaptation in Human-Robot Interaction: A Design Approach"*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Beijing, China, 2006, pp. 3099 - 3106.

Manuscript 4: Physiology-based Affect Recognition for Computer Assisted Intervention of Children with Autism Spectrum Disorder

Background

Autism is a neurodevelopmental disorder characterized by core deficits in social interaction, social communication, and imagination. Despite the urgent need and societal import of intensive treatment (Rutter, 2006), appropriate intervention resources for children with ASD and their families are often extremely costly when accessible (Tarkan, 2002). To provide alternative intervention approaches, a number of recent studies have been exploring advanced computer[1] assisted interactive technologies in intervention of children with ASD, such as computer technology, virtual reality (VR) environments, and robotic systems. Even though there is increasing research in computer assisted intervention, we found no published studies that specifically addressed automatic detection of affective cues of children with ASD. This could be important since (i) researches suggest that endowing a computer with an ability to understand implicit affective cues should permit more meaningful and natural human-computer interaction

[1] we use the term computer to imply both computer and robot assisted ASD interventions in this work

(Picard, 1997), and (ii) it is common in autism therapy that therapists who work with children with ASD continuously monitor various affective information or cues of the children in order to adapt their intervention strategies (Seip, 1996). Physiology-based affective modeling has advantages over other observational modalities (e.g., facial expression or vocal intonation) in evaluating the emotional responses of the children with ASD, since it permits continuous gathering of rich data in the face of potential communicative limitations of children with ASD (both nonverbal and verbal), particularly regarding expression of expression of affective states (DSM-IV-TR, American Psychiatric Association, 2000; Green, et al., 2002).

Summary of Contribution

There are three main contributions in this work. First is to design and implement computer-based cognitive tasks that successfully elicited target affective states (i.e., liking, anxiety, and engagement) in the children with ASD studied. Special efforts have been made in the task design (e.g., participants' age range and cognitive ability, task user interface, task configuration, task instruction, order of task presentation, and subjective report questionnaire, etc.) to produce changes in physiological signals with presumed underlying changes in affective states and have the participants engage in the cognitive tasks while the generated varying physiological signals are being collected. The second is to investigate and analyze multiple subjective reports from an autism therapist, a parent, and the participant to account for the suspected unreliability of the subjective self-reports from children with ASD. The designed experimental setup and protocol allowed an autism therapist and a parent of the participant to be effectively involved in the study and thus to provide the subjective reports. The third is to generate and collect multiple channels of physiological data and derive a large set of physiological features, which was used by Support Vector Machines (SVM) to develop affective models for children with ASD. An individual-specific approach was employed to account for the phenomenon of person-stereotypy and spectrum nature of autism (DSM-IV-TR, American Psychiatric Association, 2000). A therapist-like affective model was achieved that yields reliable prediction performance. This is the first time that the affective states of children with ASD have been experimentally detected via physiology-based affect recognition technique. In addition, this work also investigated (i) the effects of reducing the number of physiological signals to achieve more economical modeling, and (ii) the correlation between the affective model's prediction performance and the agreement between the therapist and parent on the subjective reports about how they thought the participant was feeling during the tasks. Manuscript 4 is based on the following papers:

- Liu, C., Conn, K., Sarkar, N., and Stone, W., *"Physiology-based Affect Recognition for Computer Assisted Intervention of Children with Autism Spectrum Disorder"*, International Journal of Human-Computer Studies, vol.66, no. 9, pp. 662–677, 2008.
- Liu, C., Conn, K., Sarkar, N., and Stone, W., "*Affect Recognition in Robot Assisted Rehabilitation of Children with Autism Spectrum Disorder*," IEEE International Conference on Robotics and Automation (ICRA), Roma, Italy, 2007, pp. 1755-1760.

Manuscript 5: Online Affect Detection and Robot Behavior Adaptation for Intervention of Children with Autism

Background

Recently the use of robotic technologies has been explored as potential adjuncts to autism intervention. Research suggests that robots can allow simplified but embodied social interaction that is less intimidating or confusing for children with ASD (Dautenhahn and Werry, 2004). Robots have been used to interact with children with ASD in imitation tasks and as social mediators and can help facilitate them in interacting with other individuals (Dautenhahn and Werry, 2004; Kozima, 2005). From the technological viewpoint, robots have the advantage of being robust but flexible systems that can reliably repeat as well as adaptively modify experiments while quantitatively recording data (Scassellati, 2005). Their use has shown initial promise to become an effective tool in autism intervention and could offset the cost in the long run. However, current robot-assisted autism intervention systems are not capable of interpreting the affective cues of the children with ASD, nor are they able to respond to such perception accordingly. Such abilities could be critical given the importance of human affective information in HRI (Fong et al., 2003; Picard, 1997) and the significant impacts of the affective factors of children with ASD on the intervention practice (Seip, 1996). Consequently, these robotic systems are limited in addressing the core aspect of ASD, namely creating social situations based on emotional understanding that can be suitably adjusted depending on the need of the individuals on the autism spectrum.

Summary of Contribution

This work presented a physiology-based affect-inference mechanism for robot-assisted intervention where the robot can detect the affective states of a child with ASD and adapt its behaviors accordingly. Psychophysiological analysis is performed that uses a large set of physiological indices and the subjective reports of the affective states from a therapist, a parent, and the child himself/herself. A robot uses a Support Vector Machines based affective model to implicitly detect the affective cues in real-time. A reinforcement learning based behavior adaptation mechanism is employed to enable the robot to adapt its behaviors autonomously as a function of the predicted child's affective state. The robot learned the individual liking level of each child with regard to the game configuration and selected appropriate behaviors to present the task at his/her preferred liking level. Results show the robot automatically predicted individual liking level in real time with 81.1% accuracy. This work is the first time that the affective states of children with ASD have been detected via a physiology-based affect recognition technique in real time. This is also the first time that the impact of affect-sensitive closed-loop interaction between a robot and a child with ASD has been demonstrated experimentally. While the results are achieved in a non-social interaction task, it is expected that the real-time affect recognition and response system described in this work will provide a basis for future research into developing robot-assisted intervention tools to help children with ASD explore social interaction dynamics in an affect-sensitive and adaptive manner. Manuscript 5 is based on the following papers:

- Liu, C., Conn, K., Sarkar, N., and Stone, W., *"Online Affect Detection and Robot Behavior Adaptation for Intervention of Children with Autism"*, IEEE Transactions on Robotics, vol.24, no.4, pp. 883-896, 2008.
- Liu, C., Conn, K., Sarkar, N., Stone, W., "*Online Affect Detection and*

Adaptation in Robot Assisted Rehabilitation for Children with Autism", 16th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN), 2007.

Manuscript 6: Active Learning Using Support Vector Machine for Physiology-based Affective Modeling for Children with Autism

Background

There have been two types of constraints that impede the efficient development of human machine interaction applications when the participation of human subjects and the classification are demanded (e.g., in affective computing): i) sample collection/processing and ii) sample labeling. For example, to build physiology-based affective-models, a human-machine interaction task usually has to be performed for 2-5 minutes to get the physiological signals and subjective reports have to be collected in order to get one labeled training data point (Liu, et al., 2008; Picard, Vyzas, and Healey, 2001; Rani, et al., 2006). The experimental data measurement and labeling can be both time consuming and expensive.

In this work, we addressed the second issue in the context of developing physiology-based affective model for children with ASD. As discussed in (Liu, et al., 2008; Chapter V), due to the unreliability of self-reports from the children with ASD, an autism therapist and a parent of the participant were also involved in the study to provide the subjective reports. This requirement posed additional challenges in participant recruitment and experiment coordination/schedule. An alternative approach could be video-recording the experiments and allowing therapist/parent to label the video segments at his/her own time. However, it can be expected costly and time consuming to label such video records. For example, in a study like (Liu, et al., 2008), for each child with ASD, there would be 6 hours of video needed to be reviewed and around 86 segments needed to be labeled. In such circumstances a method that allows the construction of reliable affect recognizers while only needs the labeling of a small fraction of samples can be of advantage, speeding up the procedure and possibly reducing costs due to extra analysis.

Summary of Contribution

We investigated this challenge by using Support Vector Machine active learning (SVM-AL, Tong and Chang, 2001; Tong and Koller 2001) to alleviate the efforts of sample labeling. We ran the simulation on the dataset obtained in our previous work (Liu, et al., 2008), where the physiological features of the children with ASD and the labels for the target affective state from the therapist are given. However, the labels in the dataset will not be fed to the system until they are requested, which emulates the process of video review and labeling of the therapist.

By using the margin-based query (Tong and Chang, 2001) to select the informative samples for the label requests, SVM-AL is capable of improving the relative prediction performance (*RPP*) of affective models efficiently with the use of relatively less labeled samples. Specifically, we observed that: i) SVM-AL has larger *RPP* after the first several rounds of training (10-15 labeled training data on average) than SVM-PL; ii) acceptable model performance (e.g., with 80% or 90% *RPP*) can be achieved by asking the therapist to review and annotate only about 50%-60% of the dataset; iii) It could be possible to obtain a better performance by using a fraction of samples, which are informatively selected, than using the whole training dataset. While active learning is an appealing tool for sample selecting/labelling (Cohn, D., Atlas, L., & Ladner 1994; Dagan and Engelson, 1995; Lewis and Catlett, 1994), till date no published study has been found that specifically explored its application in affective computing.

References

Ambady, N. & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A metaanalysis. Psychological Bulletin; 111(2), 256–274.

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders: DSM-IV-TR (4th ed.). Washington, DC: American Psychiatric Association.

Amir, N. & Ron, S. (1998). Towards an automatic classification of emotion in speech. Proc. of ICSLP, Sydney, 555–558.

Ark, W., Dryer, D. C., & Lu, D. J. (1999). The Emotion Mouse. Proceedings of HCI International, 818 – 823.

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. Journal of Personality Social Psychology, 37, 2049–2058.

Bernard-Opitz, V., Sriram, N., & Nakhoda-Sapuan, S. (2001). Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. J Autism Dev Disord, 31(4), 377-384.

Bloom, J. & Trautt, G. M. (1978). Finger pulse volume as a measure of anxiety: Further evaluation. Psychophysiology, 14(6), 541-544.

Bradley, M. M. (2000). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary & G. Berntson (Eds.), Handbook of Psychophysiology (pp. 602-642). New York: Cambridge University Press.

Breazeal, C. (2000). Sociable machines: expressive social exchange between humans and robots. Sc.D. dissertation, Department of Electrical Engineering and Computer Science, MIT.

Breazeal, C., & Aryananda, L. (2002). Recognition of affective communicative intent in robot-directed speech. Autonomous Robots, 12(1), 83-104.

Burgoon, J.K., Jensen, M.L., Meservy, T.O., Kruse, J., & Nunamaker J.F. (2005). Augmenting human identification of emotional states in video. Proceedings of the international conference on intelligent data analysis.

Cacioppo, J.T., Berntson, G.G., Larsen, J.T., Poehlmann, K.M., & Ito, T.A., (2000). The psychophysiology of emotion. In: Lewis, M., & Haviland-Jones, J.M. (Eds.), Handbook of Emotions. The Guilford Press, New York, 173–191.

Chen, J. (2007). Flow in games (and everything else). Communications of the ACM, 50(4), 31-34.

Cohn, D., Atlas, L., & Ladner, R. (1994). Improving generalization with active learning. Mach. Learn., 15, 201–221.

Cook III, E. W., Hawk, L. W., Davis, T. L., & Stevenson, V. E. (1991). Affective individual differences and startle reflex modulation. Journal of Abnormal Psychology, 100, 5-13.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. Applied Artificial Intelligence, 16(7-8), 555-575.

Coulson M. (1992). Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. Journal of Nonverbal Behavior, 28(2), 117–139.

Cowie R. & Douglas-Cowie, E. (1998). Automatic statistical analysis of the signal and prosodic signs of emotion in speech. Proc. of ICSLP, Philadelphia, 1989–1992.

Cowie, R., Douglas-Cowie, E., Apolloni, B., Taylor, J., Romano, A., & Fellenz, W. (1999). What a neural net needs to know about emotion words. In: CSCC'99 Proceedings, 5311–5316.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine, 18(1), 32-80.

Christie, I. C. & Friedman, B. H. (2002). Multivariate discrimination of emotion-specific autonomic nervous system activity. Psychophysiology, 39(1), 2002.

Dagan, I. & Engelson, S. P. (1995). Committee-based sampling for training probabilistic classifiers. Proc. 12th Int. Conf. Machine Learning, 150–157.

Darwin, C. (1965). The Expression of Emotions in Man and Animals, John Murray, Ed., 1872. Reprinted by Univ. Chicago Press.

Dautenhahn, K., & Werry, I. (2004). Towards interactive robots in autism therapy: background, motivation and challenges. Pragmatics & Cognition, 12, 1-35.

Dimberg, U. (1990). Facial electromyography and emotional reactions. Psychophysiology, 27, 481–494.

Edwards, G. J., Cootes, T. F., & Taylor, C. J. (1998). Face recognition using active appearance models. Proc. ECCV, 2, 581–695.

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

Ekman, P., & Friesen, W. V. (1986). A new pan cultural facial expression of emotion. Motivation and Emotion, 10(2), 1986.

Ernsperger, L. (2003). Keys to success for teaching students with autism. Future Horizons.

Evreinov, G., Agranovski, A., Yashkin, A., & Evreinova, T. (1999). PadGraph, Proceedings of HCI International, 2, 985-989.

Fong, T., Thorpe, C. & Bauer, C. (2001). Collaboration, dialogue, and human-robot interaction. Proceeding of 10th International Symposium of Robotics Research, Lorne, Victoria, Australia.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. Robotics and Autonomous Systems, 42(3-4), 143-166.

Gilleade, K., Dix, A., & Allanson, J. (2005). Affective Videogames and Modes of Affective Gaming: Assist Me, Challenge Me, Emote Me, Proceedings of DIGRA 2005.

Green, D., Baird, G., Barnett, A. L., Henderson, L., Huber, J., & Henderson, S. E. (2002). The severity and nature of motor impairment in Asperger's syndrome: a comparison with Specific Developmental Disorder of Motor Function. Journal of Child Psychology and Psychiatry and Allied Disciplines, 43(5), 655-668.

Haritaoglu, I., Cozzi, A., Koons, D., Flickner, M., Yacoob, Y., Zotkin, D., & Duriswami, R. (2001). Attentive toys. International Conference on Multimedia and Expo.

Heuft, B., Portele, T. & Rauth, M. (1996). Emotions in time domain synthesis. in Proc. of ICSLP, Philadelphia, 1974–1977.

Hoffman, G. & Breazeal, C. (2004). Robots that work in collaboration with people. CHI 2004 Workshop on Shaping Human Robot Interaction.

Hunicke, R., & Chapman, V. (2004). AI for Dynamic Difficult Adjustment in Games, Challenges in Game Artificial Intelligence AAAI Workshop (pp. 91-96). San Jose.

Iida, A., Campbell, N., Iga, S., Higuchi, F., & Yasumura, M. (1998). Acoustic nature and perceptual testing of corpora of emotional speech. Proc. of ICSLP, Sydney, 225–228.

Itoh, K., Miwa, H., Nukariya, Y., Zecca, M., Takanobu, H., Roccella, S., Carrozza, M.C., Dario, P., Takanishi, A. (2006) Development of a Bioinstrumentation System in the Interaction between a Human and a Robot. 2006 IEEE/RSJ International Conference, 2620-2625.

Jafari, R., Dabiri, F., Brisk, P., & Sarrafzadeh, M. (2005). Adaptive and fault tolerant medical vest for life critical medical monitoring. Paper presented at the 20th ACM Symposium on Applied Computing Santa Fe, NM.

Kang, B. S., Han, C. H., Lee, S. T., Youn, D. H., & Lee, C. (2000). Speaker dependent emotion recognition using speech signals. in Proceedings of International Conference on Spoken Language Processing, 383–386.

Kanda, T., Ishiguro, H., Ono, T., Imai M., & Nakatsu, R. (2002). Development and Evaluation of an Interactive Humanoid Robot Robovie. IEEE International Conference on Robotics and Automation, 1848-1855.

Kiesler S. & Hinds P. (2004).Introduction to This Special Issue on Human-Robot Interaction. Human-Computer Interaction, 19, 1~8.

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. Medical & Biological Engineering & Computing, 42, 419-427.

Koster, R. (2004). Theory of Fun for Game Design. Phoenix: Paraglyph Press.

Kobayashi, H. & Hara, F. (1992). Recognition of mixed facial expressions by a neural network. Proc. ROMAN, 387–391.

Kozima, H., Nakagawa, C., & Yasuda, Y. (2005). Interactive robots for communication-care: A case-study in autism therapy. in IEEE International Workshop on Robot and Human Interactive Communication, Nashville, TN., 341-346.

Lacey, J. I., & Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy. Am J Psychol, 71(1), 50-73.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. International Journal of Human-Computer Studies, 65(4), 329-347

Mehrabian A. & Friar, J. T. (1969). Encoding of attitude by a seated communicator vis postures and position cues. Journal of Consulting and Clinical Psychology, 5.

Michaud, F. and Theberge-Turmel, C. (2002). Mobile robotic toys and autism. in Socially Intelligent Agents: Creating Relationships with Computers and Robots, K. Dautenhahn, A. H. Bond, L. Canamero, and B. Edmonds, Eds.: Kluwer Academic Publishers, 125-132.

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

Lang, P. J. (1979). A bio-informational theory of emotional imagery. Psychophysiology, 16, 495-512.

Lazarus, R. S. (1968). Emotions and adaptation: conceptual and empirical relations. in W. J. Arnold, Ed., Nebraska Symposium on Motivation, 16, 175-266, Lincoln: University of Nebraska Press.

Lewis D. D. & Catlett, J. (1994). Heterogeneous uncertainty sampling for supervised learning. Proc. 11th Int. Conf. Machine Learning, 148–156.

Li Y. & Zhao, Y. (1998). Recognizing emotions in speech using short-term and long-term features. in Proceedings of International Conference on Spoken Language Processing, 2255–2258.

Liu, C., Conn, K., Sarkar, N., & Stone, W. (2008). Physiology-based affect recognition for computer assisted intervention of children with autism spectrum disorder. International Journal of Human-Computer Studies, 66(9), 662–677.

McCraty, R., Atkinson, M,, Tiller, W. A., Rein, G., & Watkins, A. D. (1991). The effects of emotions on short-Term power spectral analysis of heart rate variability. American Journal of Cardiology, 76 (14), 1089-1093.

Moore, D. J., McGrath, P., & Thorpe, J. (2000). Computer aided learning for people with autism - A framework for research and development. Innovations in Education and Training International, 37(3), 218-228.

Mota, S. & Picard, R. W. (2003). Automated posture analysis for detecting learner's interest level. Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction.

Murray, I. R. & Arnott, L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. Journal Acoustical Society of America, 93(2), 1097–1108.

Nasoz, F., Alvarez, K., Lisetti, C., & Finkelstein, N. (2003). Emotion recognition from physiological signals for presence technologies. Int J Cogn Technol Work Spec Issue Presence, 6(1).

Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R. L., & Fuller, T. (2002). User-centered design in games. In J. A. Jacko & A. Sears (Eds.), The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications (pp. 883-906). Mahwah, NJ: Lawrence Erlbaum Associates.

Pantic, M. & Rothkrantz, L. J.M. (2003). Towards an affect-sensitive multimodal human-computer interaction," in Proceedings of the IEEE, 91(9), 1370-1390.

Pantic, M., Rothkrantz, L.J.M. (2000).Automatic analysis of facial expression: The state of the art. IEEE Trans. Pattern Anal. Machine Intell. 22, 1424–1445.

Pantic, M. & Rothkrantz, L. J. M. (2000). Expert system for automatic analysis of facial expression. Image Vis. Comput. J., 18 (11), 881–905.

Pantic, M. (2001). Facial expression analysis by computational intelligence techniques. Ph.D. dissertation, Delft Univ. Technol., Delft, The Netherlands.

Parsons, S., & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. J Intellect Disabil Res, 46(Pt 5), 430-443.

Parsons, S., Mitchell, P., & Leonard, A. (2005). Do adolescents with autistic spectrum disorders adhere to social conventions in virtual environments? Autism, 9(1), 95-117.

Pavlovic, V. I., Sharma, R., & Huang, T. S. (1995). Visual interpretation of hand gestures for human-computer interaction: A review. Technical Report UIUC-BI-AI-RCV-95-10, University of Illinois at Urbana-Champaign.

Pecchinenda, A., & Smith, C. A. (1996). The affective significance of skin conductance activity during a difficult problem-solving task. Cognition and Emotion, 10(5), 481-504.

Pioggia, G., Igliozzi, R., Ferro, M., Ahluwalia, A., Muratori, F., & De Rossi, D. (2005). An android for enhancing social skills and emotion recognition in people with autism. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13(4), 507-515.

Prendinger, H., Mori, J., & 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.

Picard, R. W. (1997). Affective Computing. Cambridge: The MIT Press.

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

Picard, R.W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D., & Strohecker, C. (2004). Affective learning - a manifesto. BT Technical Journal, 4, 253-269.

Piekarski, W. & Thomas, B. H. (2002). The tinmith system: demonstrating new techniques for mobile augmented reality modeling. in Third Australasian conference on User interfaces, 61–70.

Rani, P., Liu, C., Sarkar, N., & Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human-robot interaction”, Pattern Analysis and Applications Journal, 9(1), 58-69.

Reeves, B., & Nass, C. I. (1996). The media equation : how people treat computers, televisions, and new media as real people and places. New York: Cambridge University Press.

Russell, J.A., (1980). A circumplex model of affect. Journal of Personality and Social Psychology 39, 1161–1178.

Russell, J.A. (1994). Is there universal recognition of emotion from facial expression? A review of the crosscultural studies. Psychological Bulletin 115, 102–141.

Rutter, M. (2006). Autism: its recognition, early diagnosis, and service implications. J Dev Behav Pediatr, 27(2 Suppl), S54-58.

Scassellati, B. (2005). Quantitative metrics of social response for autism diagnosis. in Proc. IEEE International Workshop on Robot and Human Interactive Communication, Nashville, TN, 585- 590.

Scherer, K.R. (1996). Adding the affective dimension: a new look in speech analysis and synthesis. Proc. International Conf. on Spoken Language Processing, 1808– 1811.

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

Sebe, N., Lew, M. S., Cohen, I., Garg, A., & Huang, T. S. (2002). Emotion recognition using a cauchy naive bayes classifier. Proc. ICPR, 1, 17–20.

Seip, J. (1996). Teaching the autistic and developmentally delayed: A guide for staff training and development. Delta, British, Columbia.

Silva, P. R. D., Osano, M., Marasinghe, A., & Madurapperuma, A. P. (2006). A computational model for recognizing emotion with intensity for machine vision applications," IEICE Transactions, vol. 89-D(7), 2171-2179.

Silver, M., & Oakes, P. (2001). Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others. Autism, 5(3), 299-316.

Sinha, R., Lovallo, W. R., & Parsons, O. A. (1992). Cardiovascular differentiation of emotions. Psychosomatic Medicine. 54(4), 422-435.

Simon, H. (1979). Motivational and Emotional Controls of Cognition. in Models of Thought, Yale University Press, 29-38.

Sloman A. & Croucher, M. (1981). Why robots will have emotions. in the Proceedings of the Seventh International Conference on AI, 197-202.

Smith, C. A. (1989). Dimensions of appraisal and physiological response in emotion. Journal of Personality and Social Psychology, 56, 339-353.

Standen, P. J., & Brown, D. J. (2005). Virtual reality in the rehabilitation of people with intellectual disabilities: review. Cyberpsychol Behav, 8(3), 272-282; discussion 283-278.

Swettenham, J. (1996). Can children with autism be taught to understand false belief using computers? J Child Psychol Psychiatry, 37(2), 157-165.

Tarkan, L. (October 21, 2002). Autism therapy is called effective, but rare. New York Times.

Tong, S. & Chang, E. (2001). Support vector machine active learning for image retrieval. Proc. IEEE Int. Conf. Computer Vision Patter Recognition, 150–157.

Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. Journal of Machine Learning Research, 45-66.

Trepagnier, C. Y., Sebrechts, M. M., Finkelmeyer, A., Stewart, W., Woodford, J., & Coleman, M. (2006). Simulating social interaction to address deficits of autistic spectrum disorder in children. Cyberpsychol Behav, 9(2), 213-217.

Yang, M. H., Kriegman, D. J., & Ahuja, N. (2002). Detecting faces in images: A survey. IEEE Trans. Pattern Anal. Machine Intell., 24, 34–58.

York, M., Bedard S., & Colindres, C. (1984). Categories of Implicit Interpersonal Communication. Perceptual and Motor Skills, 59(3), 855-862.

Vansteelandt, K., Van Mechelen, I., & Nezlek, J. B. (2005). The co-occurrence of emotions in daily life: A multilevel approach. Journal of Research in Personality, 39(3), 325-335.

Wieder, S. and Greenspan, S. (2005). Can children with autism master the core deficits and become empathetic, creative, and reflective? The Journal of Developmental and Learning Disorders, 9.

Wijesiriwardana, R., Mitcham, K., & Dias, T. (2004). Fibre-meshed transducers based real time wearable physiological information monitoring system. Paper presented at the International Symposium on Wearable Computers, Washington, DC.

CHAPTER II: MANUSCRIPT 1

AN EMPIRICAL STUDY OF MACHINE LEARNING TECHNIQUES FOR AFFECT RECOGNITION IN HUMAN-ROBOT INTERACTION

Pramila Rani, Changchun Liu, Nilanjan Sarkar & Eric Vanman

(Published by Pattern Analysis and Applications)

Abstract

Given the importance of implicit communication in human interactions, it would be valuable to have this capability in robotic systems wherein a robot can detect the motivations and emotions of the person it is working with. Recognizing affective states from physiological cues is an effective way of implementing implicit human-robot interaction. Several machine learning techniques have been successfully employed in affect-recognition to predict the affective state of an individual given a set of physiological features. However, a systematic comparison of the strengths and weaknesses of these methods has not yet been done. In this paper we present a comparative study of four machine learning methods- K-Nearest Neighbor, Regression Tree, Bayesian Network and Support Vector Machine as applied to the domain of affect recognition using physiological signals. The results showed that Support Vector Machine gave the best classification accuracy even though all the methods performed competitively. Regression Tree gave the next best classification accuracy and was the most space and time efficient.

Key words: Affect Recognition, Machine Learning, Psychophysiology, and Emotional Robotics

1. Introduction

There has been a steady progress in the field of intelligent and interactive robotics over the last two decades ushering in a new era of utilitarian autonomous systems. Personal and service robots will soon be seen at homes, offices, classrooms, hospitals, and factories. They will guard the premises, vacuum the floors, keep a watch over children, make photocopies, serve as "personal assistants," and in general help make our lives more comfortable. The recent "World Robotics 2004" survey [1] states "In the long run, service robots will be everyday tools for mankind". It also reports that over 600,000 household robots were in use by the end of 2003, a number that is anticipated to exceed 4 million units by 2007. As robots and people begin to co-exist and cooperatively share a variety of tasks, "natural" human-robot interaction that resembles human-human interaction is becoming an increasingly important aspect of robots. Reeves and Nass in their counterintuitive yet outstanding work [2] have already shown that people's interactions with computers, TV and similar machines/media are fundamentally social and natural, just like interactions in real life.

Human interactions are characterized by explicit as well as implicit channels of communication. While the explicit channel transmits overt messages, the other one transmits implicit messages about the communicator. Ensuring sensitivity to the other party's emotions is one of the key tasks associated with the second, implicit channel 0. In making robots respond naturally and sociably to humans implies that robots should have a degree of sensibility to human emotions and temperaments. It has been shown previously that emotions aid in perception, understanding and intelligent behaviour [4]. Therefore, endowing robots with a degree of emotional intelligence should permit more meaningful and natural human-robot interaction. The potential applications of robots that can detect a person's affective states and interact with him/her based on such perception are varied and numerous. Whether it is the domain of personal home aids that assist in cleaning and transportation, toy robots that engage and entertain kids, professional service robots that act as assistants in offices, hospitals, and museums, or the search, rescue and surveillance robots that accompany soldiers and fire-fighters – this novel aspect of human-robot interaction will impact all of them.

While the earliest social robots were developed in 1940s [5], work in the area of emotion recognition by robots has gained grounds only in the last decade. Various modalities such as facial expression, vocal intonation, gestures, and postures can be utilized to determine the underlying emotion of a person interacting with the robot [6][7]. An exhaustive survey of affect-detection based on vision and speech is provided in a work by Pantic et al. [8]. Physiology is yet another effective way of estimating the emotional state of a person and is being actively used by several research groups. In psychophysiology (the branch of psychology that is concerned with the physiological bases of psychological processes), it has been generally agreed on the fact that emotions and physiology (biological signals such as heart activity, muscle tension, blood pressure, skin conductance etc.) are closely intertwined and one influences the other. Affective computing methods are now being enthusiastically applied to human-computer interaction and other domains such as driving, flying, and machine operation [9]-[12]. However, the application of this technique in the robotics domain is relatively less [13].

In the previous research works in emotion recognition, changes in emotions have been considered either along a continuous dimension (e.g., valence and arousal) or among discrete states. Various machine learning and pattern recognition methods have been applied for determining the underlying affective state from cues such as facial expressions, vocal intonations, and physiology. Fuzzy logic [14], K-Nearest Neighbor algorithm [15], linear and nonlinear regression analysis [16], discriminant analysis [17], and combination of Sequential Floating Forward Search and Fisher Projection methods [18] have been used in the past to infer affective states. Apart from the above-mentioned methods neural networks [19] , Bayesian classification methods [20], Hidden Markov Model [21], and Dynamic Decision Network 0 have also been used.

Even though several methods have been successfully employed to build affect recognizers from physiological indices, a systematic comparison of various methods- their strengths and weaknesses has not been possible largely because of the following points of diversity in each study:

(i) Definition of emotion – discrete or continuous

(ii) Nature of physiological features used

(iii) Manner of self-reporting to get subjective affective states of participants

(iv) Baselining techniques used

Given these diversities, it is hard to find a common ground for comparing methods and analysing their merits. This paper makes an attempt to investigate the performance of four popular machine learning methods – K-Nearest Neighbour (KNN), Regression Tree (RT), Bayesian Network (BNT) and Support Vector Machine (SVM), when applied to the domain of affect recognition using physiological indices. All the methods were tested using the same physiological data from the same tasks, hence attaining uniformity in the various aspects of emotion elicitation, data processing, feature extraction, baselining, and self-reporting procedures.

Figure 1 shows an overview of the method. The input feature set was derived from physiological signals after the application of a series of pre-processing and signal analysis techniques. The output set was derived from the self-report of the participant. Each vector of input features had a corresponding output vector consisting of self-report for all the affective states. This data set was utilized for all four machine learning techniques.

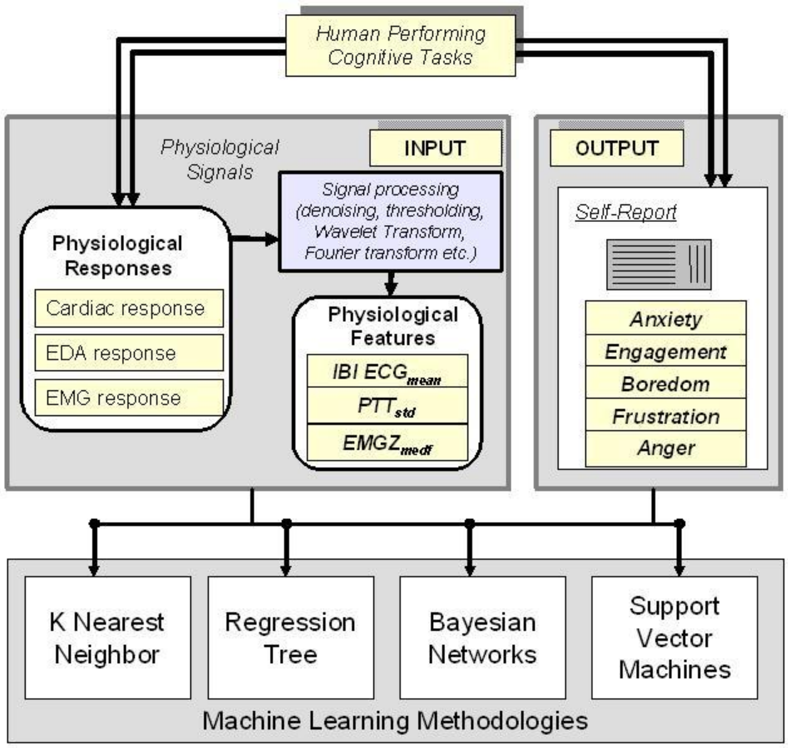


Figure 1. Method Overview

The next step in this research would be to imbed the affect-recognizer in a robot's existing functionality so that it can be responsive to the affective states of the people it works with. This leads to a new set of research issues regarding robot control architecture for behaviour switching and performance metrics for systematic evaluation. This paper does not deal with those aspects but regards it as the next phase in the development of an implicit human-robot interaction framework that will be covered in our future works.

The paper is organized as follows: Section II describes the existing methods that are used for classifying affective states based on physiology and their respective accuracies. A brief description of the physiological signals and the features derived from these signals employed for affect recognition are presented in Section III. The particulars of cognitive tasks designed for affect elicitation are described in Section IV. The details of experimental setup are presented in Section V. In Section VI, we describe the four methods that have been employed in this empirical study - KNN, RT, BNT and SVM. This is followed by the results and discussion in Section VII. Finally, Section VIII summarizes the contributions of the paper and provides important conclusions.

2. Existing Methods and Classification Accuracies

Several researchers in human machine interaction have focused on physiology-based affect-recognition. Picard and colleagues have employed a combination of Sequential Floating Forward Search and Fisher Projection methods to classify eight emotions with 81% accuracy [17]. K-Nearest Neighbor, Discriminant Function Analysis and Marquardt Backpropagation algorithms were applied to differentiate among six emotions by Lisetti and Nasoz and the correct classification accuracies - 71%, 74% and 83% were achieved respectively [23]. Artificial Neural Network has also been used to assess the mental workload and the mean classification accuracies were 85%, 82%, and 86% for the baseline, low task difficulty and high task difficulty conditions, respectively [24]. Support Vector Machines based emotion-recognizers have also been investigated in [25]0, where correct classification accuracies of 78.4%, 61.8%, and 41.7% were reported for the reorganization of three, four and five emotions, respectively.

3. Physiological Signals and Features for Affect Recognition

There is good evidence that the physiological activity associated with affective state can be differentiated and systematically organized. The transition from one emotional state to another, for instance, from state of boredom to state of anxiety is accompanied by dynamic shifts in indicators of Autonomic Nervous System (ANS) activity. The physiological signals we examined are: various features of cardiovascular activity, including interbeat interval, relative pulse volume, pulse transit time, heart sound, and pre-ejection period; electrodermal activity (tonic and phasic response from skin conductance) and electromyogram (EMG) activity (from corrugator supercilii, zygomaticus, and upper trapezius muscles) [27]. These signals were selected because they i) are shown to capture important information about the underlying targeted affective states, (ii) can be measured non-invasively; and iii) are relatively resistant to movement artifacts.

Table1. Physiological Indices

Physiological Response	Features Derived	Label Used	Unit of Measurement
Cardiac activity	Sympathetic power (from ECG)	Sym	Unit/Square Second
	Parasympathetic power (from ECG)	Para	Unit/Square Second
	Very Low Frequency Power (from ECG)	VLF	Unit/Square Second
	Ratio of powers	Sym Para Para VLF Sym VLF	No unit
	Mean IBI	IBI ECG_{mean}	Milliseconds
	Std. of IBI	IBI ECG_{std}	Standard Deviation (no unit)
	Mean amplitude of the peak values of the PPG signal (Photoplethysmogram)	PPG $Peak_{mean}$	Micro Volts
	Standard deviation (Std.) of the peak values of the PPG signal	PPG $Peak_{std}$	Standard Deviation (no unit)
	Mean Pulse Transit Time	PTT_{mean}	Milliseconds
Heart Sound	Mean of the 3rd,4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Mean d3 Mean d4 Mean d5	No unit
	Standard deviation of the 3rd,4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Std d3 Std d4 Std d5	No unit
Bioimpedance	Mean Pre-Ejection Period	PEP_{mean}	Milliseconds
	Mean IBI	IBI ICG_{mean}	Milliseconds
Electrodermal activity	Mean tonic activity level	$Tonic_{mean}$	Micro-Siemens
	Slope of tonic activity	$Tonic_{slope}$	Micro-Siemens/Second
	Mean amplitude of skin conductance response (phasic activity)	$Phasic_{mean}$	Micro-Siemens
	Maximum amplitude of skin conductance response (phasic activity)	$Phasic_{max}$	Micro-Siemens
	Rate of phasic activity	$Phasic_{rate}$	Response peaks/Second
Electromyographic activity	Mean of Corrugator Supercilii activity	Cor_{mean}	Micro Volts
	Std. of Corrugator Supercilii activity	Cor_{std}	Standard Deviation (no unit)
	Slope. of Corrugator Supercilii activity	Cor_{slope}	Micro Volts/Second
	Mean Interbeat Interval of blink activity	IBI $Blink_{mean}$	Milliseconds
	Std. of Interbeat Interval of blink activity	IBI $Blink_{std}$	Standard Deviation (no unit)
	Mean amplitude of blink activity	Amp $Blink_{mean}$	Micro Volts
	Standard deviation of blink activity	$Blink_{std}$	Standard Deviation (no unit)
	Mean of Zygomaticus Major activity	Zyg_{mean}	Micro Volts
	Std. of Zygomaticus Major activity	Zyg_{std}	Standard Deviation (no unit)
	Slope. of Zygomaticus Major activity	Zyg_{slope}	Micro Volts/Second
	Mean of Upper Trapezius activity	$Trap_{mean}$	Micro Volts
	Std. of Upper Trapezius activity	$Trap_{std}$	Standard Deviation (no unit)
	Slope. of Upper Trapezius activity	$Trap_{slope}$	Micro Volts/Second
	Mean and Median frequency of Corrugator, Zygomaticus and Trapezius	$Zfreq_{mean}$ $Cfreq_{median}$ $Tfreq_{mean}$ etc.	Hertz
Temperature	Mean temperature	$Temp_{mean}$	Degree Centigrade
	Slope of temperature	$Temp_{slope}$	Degree Centigrade/Second
	Std. of temperature	$Temp_{std}$	Standard Deviation (no unit)

Multiple features (as shown in Table 1) were derived for each physiological measure. Some of these features are described in our previous work [28]. "Sym" is the power associated with the sympathetic nervous system activity of the heart (in the frequency band 0.04-0.15 Hz.). "Para" is the power associated with the heart's parasympathetic nervous system activity (in the frequency band 0.15-0.4 Hz.). "VLF" is the power associated with the Very Low Frequency band (less than 0.04 Hz.). InterBeat Interval (IBI) is the time interval in milliseconds between two "R" waves in the ECG waveform in millisecond. IBI ECGmean and IBI ECGstd are the mean and standard deviation of the IBI. Photoplethysmograpm signal (PPG) measures changes in the volume of blood in the finger tip associated with the pulse cycle, and it provides an index of the relative constriction versus dilation of the blood vessels in the periphery. Pulse transit time (PTT) is the time it takes for the pulse pressure wave to travel from the heart to the periphery, and it is estimated by computing the time between systole at the heart (as indicated by the R-wave of the ECG) and the peak of the pulse wave reaching the peripheral site where PPG is being measured. Heart Sound signal measures sounds generated during each heartbeat. These sounds are produced by blood turbulence primarily due to the closing of the valves within the heart. The features extracted from the heart sound signal consisted of the mean and standard deviation of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform. Bioelectrical impedance analysis (BIA) measures the impedance or opposition to the flow of an electric current through the body fluids contained mainly in the lean and fat tissue. A common variable in recent psychophysiology research, pre-ejection period (PEP) derived from impedance cardiogram (ICG) and ECG measures the latency between the onset of electromechanical systole, and the onset of left-ventricular ejection and is most heavily influenced by sympathetic innervation of the heart. Electrodermal activity consists of two main components – Tonic response and Phasic response. Tonic skin conductance refers to the ongoing or the baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event related changes that occur, caused by a momentary increase in skin conductance (resembling a peak). The EMG signal from Corrugator Supercilii muscle (eyebrow) captures a person's frown and detects the tension in that region. It is also a valuable source of blink information and helps us determine the blink rate. The EMG signal from the Zygomaticus Major muscle captures the muscle movements while smiling. Upper Trapezius muscle activity measures the tension in the shoulders, one of the most common sites in the body for developing stress.

Various signal processing techniques such as Fourier transform, wavelet transform, thresholding, and peak detection, were used to derive the relevant features from the physiological signals. All these features are powerful indicators of the underlying affective state of the person showing this response. We have exploited this dependence of a person's physiological response on affect to detect and identify affective states of anxiety, engagement, boredom, frustration and anger in real-time using advanced signal processing techniques.

One of the prime challenges with affective computing is the phenomena of person stereotypy, i.e, within a given context, different individuals express the same emotion with different characteristic response patterns. For each of the fifteen participants there were distinct physiological indices that showed high correlation with each affective state.

Any feature with an absolute correlation greater than equal to 0.3 with a given affective state was considered significant and was selected as inputs of the classifiers. Feature selection was performed in a person-specific manner. For each participant, a set of features that were highly correlated with his/her different affective states were chosen. So there were five feature sets corresponding to the five affective states per participant. For example when performing anxiety classification of participant 6 using a given classification technique, the feature set consisting of a subset of the entire feature set that was highly correlated with Participant 6's reported anxiety was chosen. Affect-recognition was performed utilizing two types of feature sets – (i) the entire feature set and, (ii) the correlated feature set respectively.

4. Cognitive Task for Affect Elicitation

Two PC based cognitive tasks were designed to elicit the above mentioned affective states in the participants. Physiological data from participants were collected during the experiment.

The aim of the tasks was to invoke in the participants varying intensities of the following five affective states: engagement, anxiety, boredom, frustration and anger. The tasks chosen were solving anagrams and playing Pong. The anagram solving task has been previously employed to explore relationships between both electrodermal and cardiovascular activity with mental anxiety [29]. Emotional responses were manipulated in this task by presenting the participant with anagrams of varying difficulty levels. A long series of trivially easy anagrams caused boredom, an optimal mix of solvable and difficult anagrams caused engagement, unsolvable or extremely difficult anagrams elicited frustration, and giving time deadlines generated anxiety. All these conditions were well tested during the task design and piloting stage.

The Pong task consisted of a series of trials each lasting up to four minutes, in which the participant played a variant of the early, classic video game “Pong”. This game has been used previously by researchers to study anxiety, performance, and gender differences [30]. Various parameters of the game were manipulated to elicit the required affective responses. These included: ball speed and size, paddle speed and size, sluggish or over-responsive keyboard and random keyboard response. Low speeds and large sizes of ball and paddle made games boring after a while, whereas high speed ball and paddle along with smaller sizes of the two made the game engaging. Very high speeds caused anxiety at times. Sluggish or over-responsive keyboard induced frustration and anger. The relative difficulties of various trial configurations were established through pilot work.

Each participant took part in six sessions (on six different days) of the above two tasks – three one hour sessions of solving anagrams and three one hour sessions of playing Pong. These task tasks spanned a period of one month. In each session, before starting the actual tasks, a ten minute baseline recording was done which was used later to offset day-variability.

5. Experimental Setup

The objective of the experiment was to elicit varying intensities of emotional states in participants as they performed computer-based cognitive tasks. Fifteen participants (eight women and seven men) took part in the experiment. Their age range was from 21 to 57 years. After initial briefing regarding the computer tasks, sensors were attached to the participant's body. Each session consisted of 3 minute epochs (for anagram tasks) and upto 4 minute epochs (for Pong tasks), each epoch followed by a questionnaire for self-reporting. The participants reported assessment of their own affective states via self-reports. During the tasks, the participant's physiology was monitored with the help of wearable biofeedback sensors and Biopac data acquisition system (www.biopac.com). The digitally sampled sensor information was sent to the computer using an Ethernet cable. The signals monitored consisted of electrocardiogram, bio-impedance, electromyogram (from the corrugator, zygomaticus and upper trapezius muscles), electrodermal activity, peripheral temperature, blood volume pulse, and heart sound.

During the experiment, a total of 15 datasets were collected (one for each participant). Each data set consisted of 46 input features and 5 output affective states - engagement, anxiety, boredom, frustration and anger. Each output state had three classes - low, medium and high. These three levels for each affective state were obtained by discretizing the output. The self-reports were normalized to [0,1] and then discretized such that 0-0.33 was labelled low, 0.34-0.67 was medium and 0.68-1.0 was labelled high. All the five affective states were discretized separately so that there were three levels in each affective state. Classifications were performed on each affective state individually using each of the four methods. Each dataset contained approximately 100 epochs.

6. Machine Learning Methods Applied

Determining the intensity (high, medium, low) of a particular affective state from the physiological response resembles a classification problem. In this classification problem, the attributes are the physiological features and the target function is the degree of arousal. Developing such an affect recognition system was challenging because of the physiological data sets being complex. The complexity was primarily due to the (i) inherent noise in the physiological signals (e.g. movement artifacts, day variability) (ii) high dimensionality (There are currently 46 features and this will increase as the number of affect detection modalities increases.), (iii) mixture of data types, and (iv) non-standard data structures. In this paper we have employed the following four methods to determine the underlying affective state of an individual given a set of physiological indices under the same operating conditions - KNN, RT, BNT and SVM. All of these methods have distinctive characteristics that make them good candidates for this empirical comparison. KNN is a widely used instance-based learning method that scores high on simplicity. RT is a popular inductive inference learning method that has a built-in feature selection capability. BNT has the distinct advantage of uncovering causal relationships among attributes, hence providing added knowledge regarding the problem domain. SVM is supported by statistical learning theory and usually shows good generalisation performance.

6.1 K-Nearest Neighbor Classifier

K-Nearest Neighbor computes the similarity between the test instance and the training instance. It finds out the category that the test instance is most similar by considering the k top–ranking nearest instances. In this work, similarity score summing method was used to assign the test instance X to the class with the maximal sum of similarity score:

$$C(X) = \arg\max_{m} \sum_{X_j \in KNN} Sim(X, X_j)\, y(X_j, c_m) \tag{1}$$

where X_j is one of the k neighbors in the training set, $y(X_j, c_m) \in \{0,1\}$ indicates whether X_j belongs to class c_m, and $Sim(X, X_j)$ measures the similarity between X

and X_j . The similarity value between two instances is based on a distance metric. The Euclidean distance metric was used here. In order to select appropriate value of k , Monte Carlo simulations were performed by minimizing the leave-one-out cross-validation error. KNN is sensitive to the noisy and irrelevant features. To cope with this problem we employed a correlation based feature selection approach to avoid the irrelevant features.

6.2 Regression Tree

A regression tree takes as input a situation or an object characterized by a set of properties and outputs a decision [31]. Each node corresponds to a test of one input feature and the branches that emerge from that node are possible test result values (positive and negative). The terminal or leaf nodes represent the value of the decision that will be returned if that node is reached. Classification And Regression Trees (CARTs) have been extensively applied in the medical field. Important applications include: diagnosing heart attacks, cancer diagnosis, speech recognition and classification of age by gait measurement [32][33].

The regression tree creation begins by choosing the best feature to split the examples. The best feature is the one that changes the classification the most. Two primary issues exist, (i) Choosing the best feature to split the examples at each stage, and (ii) Avoiding data overfitting. Many different criteria could be defined for selecting the best split at each node. In this work, the Gini Index function was used to evaluate the goodness of all the possible split points along all the features [31]. Trees were pruned based on an optimal pruning scheme that first pruned branches that gave the least improvement in error cost. Pruning was performed to remove the redundant nodes since bigger, overfitted trees have higher misclassification rates.

6.3 Bayesian Networks

Bayesian networks apply probability and belief theory to build graphical models that encode probabilistic relationships among events of significance. It is a directed acyclic graph that consists of two main components: (i) a network structure that encodes a set of conditional independence relations amongst a set of variables, and (ii) a set of tables of local probability distributions associated with each variable. Together, these components define the joint probability distribution of the variables. Such a graphical model along with statistical techniques can be a powerful tool for data analysis [34]. Some of the main advantages are: (i) Since the model encodes dependencies among all the events, it can handle cases when some data entries are missing, (ii) knowledge of causal relationships between events can be gained, that enables one to better understand a problem domain and predict the results of unexpected events, (iii) the Bayesian model is an ideal one for combining prior knowledge (which often comes in causal form) and data since it captures well both the causal and probabilistic characteristics of a problem domain.

Given all the above mentioned advantages of Bayesian classification, it can be seen that this method is a potent technique to learn and classify affective patterns of human beings. Several researchers have already used Bayesian techniques for learning blueprints of affective states through audio, visual and physiological cues [20].

For creating Bayesian network, in this paper, we used the Max-Min Hill Climbing Algorithm (MMHC) [34]. It takes as input an array of data and returns a high scoring BN. The continuous features were discretized using D2- a supervised splitting technique based on Entropy [36]. For inference we used the junction tree method provided by the BNT toolbox (http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html). The technique first compiles the Bayesian network into a secondary structure called a junction tree representing joint distributions over non-disjoint sets of variables. The new evidence is inserted, and then a message passing technique updates the joint distributions and makes them consistent. Finally, using marginalization, the distributions for each variable can be calculated.

6.4 Support Vector Machines

Support Vector Machine, pioneered by Vapnik [37], is an excellent tool for classification problems [38]. Its appeal lies in its strong association with statistical learning theory as it approximates structural risk minimization principle. Good generalization performance can be achieved by maximizing the margin, where margin is defined as the sum of the distances of the hyperplane from the nearest data points of each of the two classes.

SVM is a linear machine working in a high dimensional feature space formed by an implicit embedding of lower dimensional input data into the feature space through the use of a nonlinear mapping. This allows using linear algebra and geometry to separate the data normally only separable with nonlinear rules in the input space. To allow efficient computation of inner products directly in the feature space and circumvent the difficulty of specifying the non-linear mapping explicitly, all operations in learning and testing modes are done in SVM using so-called kernel functions satisfying Mercer conditions [37]. The explicit form of the nonlinear mapping need not be known.

The most distinctive fact about SVM is that the learning task is reduced to a dual quadratic programming problem by introducing the so-called Lagrange multipliers [37][39]. The solution with respect to the Lagrange multipliers gives the optimal hyperplane defined by the nearest training points, support vectors, for which the corresponding Lagrange multipliers are non-zero. This induces sparseness in the solution and gives rise to efficient approaches to optimization.

The SVM approach is able to deal with noisy data and overfitting by allowing for some misclassifications on the training set. This makes it particularly suitable for affect recognition because the physiology data is noisy and the training set size is often small. In this work, in order to deal with the nonlinearly separable data, soft margin classifiers with slack variables were used to find a hyperplane with less restriction [39]. RBF was selected as the kennel function because it often delivers better performance [37]. Although SVM separates the data only into two classes, the recognition of more classes can be done by applying some voting scheme, e.g., "one against one" and "one against all" approaches. We chose "one against one" in our task since it usually produces better results [40]. Grid search based ten-fold cross-validation is used to determine the parameters of the classifier. With the kernel representation, SVM provides an efficient technique that can tackle the difficult, high dimensional affect recognition problem.

7. Results and Discussion

7.1 Predictive Accuracy

The performance of KNN, RT, BNT and SVM in classifying unknown instances is shown in Fig. 2. The method of cross-validation used for all the four methods was leave-one-out. All of the methods gave competitive recognition accuracies, which verified our emotion elicitation protocol and the principle of emotion recognition from physiological signals. This was promising considering that this task was challenging in two respects (i) the emotions were elicited from participants engaged in real-life cognitive computer tasks as opposed to having actors/participants deliberately express a given emotion, and (ii) varying levels of arousal of any given emotion (for instance low frustration, high frustration) were identified instead of determining discrete emotions (for instance anger, joy, sadness etc.). This is a difficult task as the distinction between target classes is more subtle in latter than in the former case.

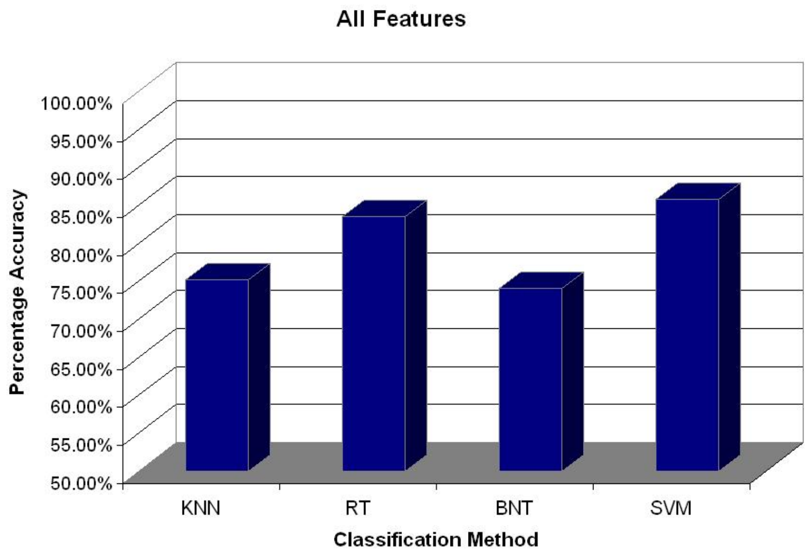


Figure 2. Classification accuracy of the methods using all the features

Figure 2 shows that the mean correct classification accuracies across all participants (averaged across all affective states) were - 75.16% for KNN, 83.50% for RT, 74.03% for BNT and 85.81% for SVM. The statistical significance of the difference in classification accuracies was tested using the sign-test between any two methods. It was found that all these differences were statistically significant with a greater than 95% significance level. One of the reasons for the low performance of KNN is that it does not perform any generalizations regarding the data set. The inductive bias of KNN method is that neighboring instances probably have similar categories. However, when the distance between two instances is calculated all features are given equal weight. This becomes problematic when the discriminative or useful features constitute only a small subset of the entire feature set. Therefore, the accuracy of KNN method is sensitive to the number of noisy features.

On the contrary, SVM delivers the best performance because it achieves a trade off between the complexity of the network and the training error so as to prevent overfitting. The performance of BNT was probably affected by the following limitations: (i) absence of an initial structure or set of constraints that could guide the generation and evaluation of high scoring networks (this is largely due to the incomplete understanding in psychophysiology regarding the interactions between physiology and affective states), (ii) limited size of data sets (due to restricted number of human-in-the-loop experiments that could be performed) that prevented the estimation of accurate conditional probabilities.

Table 2. Classification accuracy of the methods for the affective states (%)

	Anxiety	Boredom	Engagement	Frustration	Anger
KNN	80.38	73.92	70.63	70.89	79.98
RT	88.54	77.17	78.82	79.51	93.47
BNT	80.64	71.48	65.26	70.86	81.93
SVM	88.86	84.23	84.41	82.81	88.74

Table 2 shows the performance of the various learning methods for the five affective states under investigation. The classifiers have different overall performances for different kinds of emotions. For example, classification accuracy is consistently better for anxiety than for frustration. One possible reason could be that the task design resulted in elicitation of particular emotions more successfully than the others.

7.2 Efficiency

Real-time embedded applications in robotics require time and space efficiency of the learning algorithms employed. Hence, we investigated the training speed and memory requirements of the above four learning methods in the affect recognition task. The training and testing times were normalized with respect to the RT method as it was the fastest. It was found that both BNT and SVM were two times slower in training. In testing, BNT was 30 times slower whereas SVM and KNN were only 3 times slower. It is expected that with large data sets and higher number of features, SVM will be faster than KNN because of the sparse solution that SVM gives. There were other differences in the approaches of the methods in general. While BNT and RT did not require any parameter tuning, in case of KNN and SVM choosing appropriate parameters was imperative for good results.

Regarding the space efficiency, KNN stores all training instances and hence, does not extend well to very large datasets. RT on the other hand, stores only the index of the relevant features and the corresponding thresholds. This makes RT easily scalable. BNT stores the directed acyclic graph along with the conditional probability table of each node and SVM stores the support vectors which determine the discriminant hyperplane. Both BNT and SVM are more space efficient than KNN but less than RT.

7.3 Interpretability

Although RT and BNT do not work as accurately as SVM, they are still valuable candidates for the affect recognition. As an inductive learning method, Regression trees allow us to identify important physiological indicators and transfer the learned results into a set of simple rules. With the capability of capturing the causality among the physiological features and the affective states in a probabilistic manner, BNT can provide an insight into the underlying relationships among physiological features and emotions, many of which are still unknown.

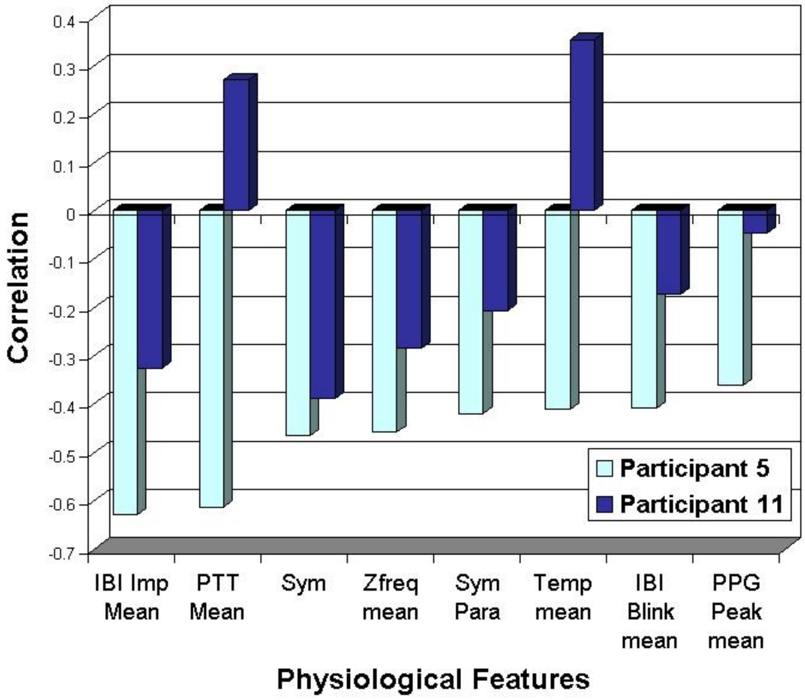


Figure 3. Person stereotypy with respect to the affective state of anxiety for Participants 5 and 11

7.4 Feature Selection

Figure 3 shows the physiological features that were highly correlated with the state of anxiety for participant 5 and the corresponding correlation of the same features with the state of anxiety for participant 11. It can be seen from Figure 3 that two features – mean of pulse transit time (PTTmean) and mean of temperature (Tempmean) are correlated differently for the two participants. While both are correlated positively with anxiety for participant 11, they are negatively correlated for participant 5. However, features like mean interbeat interval of impedance (IBI Impmean), sympathetic activity power (Sym) and mean frequency of EMG activity from zygomaticus major (Zfreqmean) are similarly related for both participants.

Also, for any individual, the set of useful indices were different for the different affective states. For instance, as seen in Figure 4, the bars in dark indicate the correlation of features that were found useful in detecting the state of engagement in Participant 5. The bars in light indicate the correlation of the same features with the state of anxiety for the same participant and it can be readily seen that most of the features that are useful in detecting engagement are not useful in detecting anxiety. This led us to believe that the classification accuracies of the above methods might increase if we used only the highly correlated features instead of the entire set.

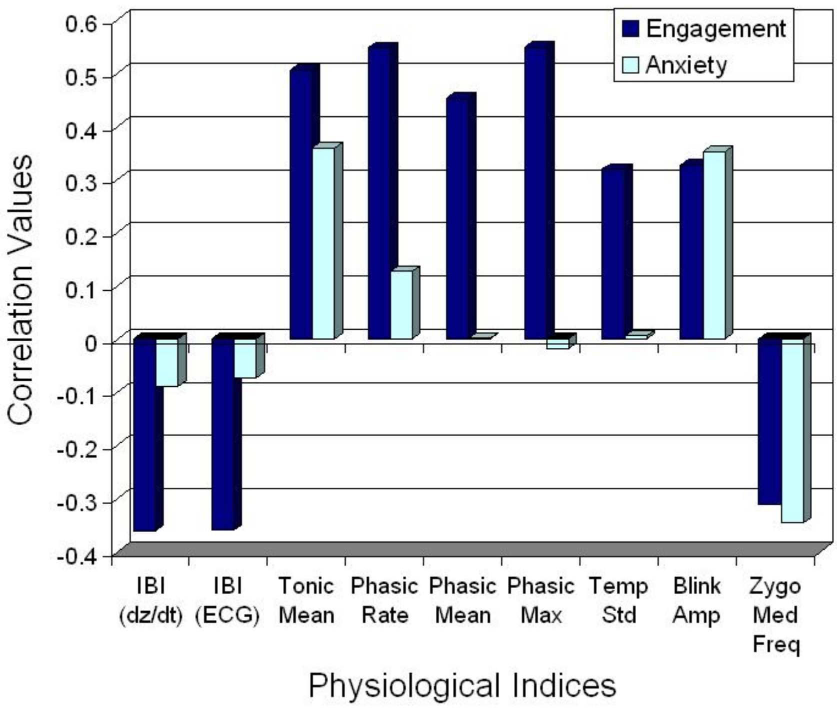


Figure 4. Comparison between Anxiety and Engagement for Participant 5

Figure 5 shows the new classification accuracy when only the highly correlated features instead of the entire feature set were used for affect learning. We performed sign test to determine statistical significance of the results with or without feature selection. It was observed that while the accuracy improved by 3.62% for KNN and 3.65% for BNT with a significance level of 95%, the performance of RT and SVM was not impacted significantly.

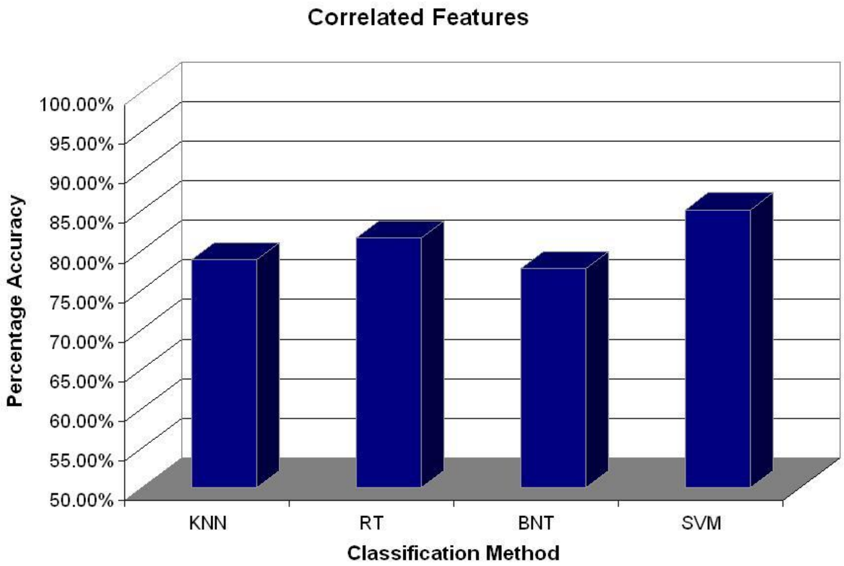


Figure 5. Classification accuracy of the methods using only the highly correlated features

As previously mentioned, KNN is sensitive to noise. By selecting only the features that were highly correlated with the target affective state we excluded the less important features. The distance metric now gave more concise measure of the similarity between instances leading to better performance. In case of BNT also there was an improved performance with selected features. One conjecture would be that conditioning on many variables with the limited data set previously may have caused poor approximation of the probability values and also introduced spurious edges. Hence, filtering out the less significant features improves the BNT performance. RT's performance maintained nearly the same probably because this method already has an inbuilt feature selection capability wherein it selects useful features by the method of information gain. Correlation based feature selection in the input space also did not improve the performance of SVM in this task. SVM employs nonlinear mappings that result in a high dimensional feature space. Hence, enforcing linear relationships by using correlated features may not have worked well in this situation.

8. Conclusion and Future Work

A comparative study of the merits of four popular learning methods as applied to affect detection was presented. In this work we focussed on determining affective states from physiological signals. We discussed the nature of the physiological signals and the derived features from them, that were used for affect recognition. Two cognitive tasks – solving anagrams and playing Pong, were designed to elicit affective states of anxiety, engagement, boredom, frustration and anger in participants. Fifteen participants took part in this study where each was involved in the tasks for 6 hours. Their physiology was continuously monitored during the tasks using biofeedback sensors.

The problem under investigation was as follows: given a set physiological features, each labelled as an indicator of a particular level of arousal of a given affective state, determine the performance of the following learning methods in predicting the class of unseen instances – KNN, RT, BNT, and SVM. There is no such comparison reported in the literature that uses the same physiological data from the same tasks, hence attaining uniformity in the various aspects of emotion elicitation, data processing, feature extraction, baselining, and self-reporting procedures. However, such a comparative study is important for the development of affect recognition in human-robot interactions. The contribution of the present work lies in providing a basis for choice of different affect recognition algorithms.

It was found that SVM with a classification accuracy of 85.81% performed the best, closely followed by RT (83.50%), KNN (75.16%) and BNT (74.03%). Using informative features (the ones that were highly correlated with the affective states) improved the performance for KNN and BNT by almost 4%. In terms of space and time efficiency, RT ranked higher than the other methods.

Future work will involve performing closed-loop experiments involving implicit human-robot interaction based on affective states. We will investigate a human-robot interaction task where the robot will implicitly sense human affective states using affect recognition algorithms described here, and alter its behaviour in order to address human need.

References

[1] World Robotics 2004 -- Statistics, Market Analysis, Forecasts, Case Studies and Profitability of Robot Investment" (Sales No. GV.E.04.0.20 or ISBN No. 92-1-101084-5)

[2] B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Televisions and New Media Like Real People and Places*, New York: Cambridge Univ. Press, 1996.

[3] R. Cowie, E, Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J. G. Taylor, "Emotion recognition in human-computer interaction," *IEEE Signal processing magazine*, vol.18, no.2, pp. 32-80, Jan 2001.

[4] R. Picard, *Affective Computing*, Cambridge, The MIT Press, 1997.

[5] W. G. Walter, *The Living Brain*, New York, W. W. Norton, 1963.

[6] C. Breazeal and L. Aryananda, "Recognizing affective intent in robot directed speech", *Autonomous Robots*, vol. 12, no. 1, pp. 83-104, 2002.

[7] G.C Littlewort, M.S. Bartlett, J. Chenu, I. Fasel, T. Kanda, H. Ishiguro, and J.R Movellan, "Towards social robots: Automatic evaluation of human-robot interaction by face detection and expression classification" , *Advances in Neural Information Processing Systems*, vol. 16. pp. 1563-1570, 2004

[8] M. Pantic and L.J.M. Rothkrantz, "Towards an affect-sensitive multimodal human-computer interaction ," *in Proceedings of the IEEE*, vol.91, no.9, pp: 1370-1390, September 2003.

[9] C. Conati and X. Zhou, "Modeling Students' Emotions from Cognitive Appraisal in Educational Games", *Proceedings of 6th International Conference on Intelligent Tutoring Systems*, France, 2002

[10] R. W. Backs, J. K. Lenneman, J. M. Wetzel, P. Green, "Cardiac measures of driver workload during simulated driving with and without visual occlusion, " *Human Factors*, vol.45, no.4, pp. 525 -539, 2003.

[11] E. Hudlicka and M.D. McNeese, "Assessment of user affective and belief states for inference adaptation: application to an air force pilot task", *User Modeling and User Adapted Interaction*, vol. 12, pp 1-47.

[12] Y. Hayakawa and S. Sugano, "Real time simple measurement of mental strain in machine operation", *ISCIE 1998 Japan-U.S.A. symposium on Flexible Automation*, pp: 35-42, Otsu, Japan, 1998.

[13] Dana Kulic and E. Croft, "Estimating Intent for Human-Robot Interaction*," in Proceedings of IEEE International Conference. on Advanced Robotics*, pp. 810-815, 2003

[14] N. Tsapatsoulis, K. Karpouzis, G. Stamou, F. Piat and S. Kollias, "A Fuzzy System for Emotion Classification based on the MPEG-4 Facial Definition Parameter Set," *in Proceedings of EUSIPCO-2000*, Finland, 2000.

[15] V.A. Petrushin. "Emotion recognition agents in real world." *AAAI Fall Symposium on Socially Intelligent Agents: Human in the Loop,* 2000.

[16] T. Moriyama, H. Saito, and S. Ozawa. "Evaluation of the relation between emotional concepts and emotional parameters on speech." *IEICE Journal*, J82-DII(10), pp. 1710-1720, 1999.

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

[18] R.W. Picard, E. Vyzas, and J. Healy. "Toward machine emotional intelligence: analysis of affective psychological states." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175-1191, 2001.

[19] J. Zhao and G. Kearney. “Classifying facial movement by backpropagation neural networks with fuzzy inputs.” *in Proceedings of International Conference on Neural Information Processing*, pp. 454-457, 1996.

[20] Y. Qi and R. W. Picard. “Context-sensitive Bayesian Classifiers and Application to Mouse Pressure Pattern Classification,” *in Proceedings of International Conference on Pattern Recognition,* Canada, 2002.

[21] I. Cohen, A. Garg, and T.S. Huang. “Emotion recognition using multilevel HMM.” *in Proceedings of NIPS Workshop on Affective Computing,* Colorado, 2000.

[22] C. Conati, "Probabilistic Assessment of User's Emotions in Educational Games," *Journal of Applied Artificial Intelligence, special issue on “Merging Cognition and Affect in HCI”,* vol. 16, pp. 555-575, 2002.

[23] F. Nasoz, K. Alvarez, C. Lisetti, and N. Finkelstein, "Emotion Recognition from Physiological Signals for Presence Technologies," *International Journal of Cognition, Technology, and Work - Special Issue on Presence*, Vol. 6, no. 1, 2003.

[24] G. F. Wilson, C. A. Russell, "Real-time assessment of mental workload using psychophysiological measures and artificial neural networks," *Human Factors*, vol. 45, no. 4, pp. 635 -643, 2003.

[25] K. Takahashi, "Remarks on emotion recognition from bio-potential signals", *in Proceedings of 2nd International Conference on Autonomous Robots and Agents*, New Zealand, 2004.

[26] K. H. Kim, S. W. Bang, S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical & Biological Engineering & Computing*, vol. 42, pp. 419-427, 2004.

[27] M. M. Bradley, "Emotion and motivation," in Handbook of Psychophysiology, J.T. Cacioppo, L. G. Tassinary, and G. Berntson, Eds., pp. 602-642, New York: Cambridge University Press, 2000.

[28] P. Rani, N. Sarkar, C. Smith, and L. Kirby, “Anxiety Detecting Robotic Systems – Towards Implicit Human-Robot Collaboration,” *Robotica*, vol. 22, no. 1, pp. 85-95, 2004.

[29] A. Pecchinenda and C. A. Smith, "The Affective Significance of Skin Conductance Activity During a Difficult Problem-solving Task ," *Cognition and Emotion*, vol. 10, no. 5, pp. 481-504, 1996.

[30] R. M. Brown, L. R. Hall, R. Holtzer, S. L. Brown, N. L. Brown, "Gender and Video Game Performance," *Sex Roles*, vol. 36, no. 11-12, pp. 793 – 812, June 1997.

[31] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA., 1984.

[32] P. Kokol, M. Mernik, J. Završnik, K. Kancler and I. Malèiæ, "Decision Trees and Automatic Learning and Their Use in Cardiology," *Journal of Medical Systems,* vol. 9, no. 4, pp. 201-206, 1994.

[33] S. Downey and M. J, Russell, "A Decision Tree Approach to Task Independent Speech Recognition," *in Proceedings of Inst Acoustics Autumn Conf on Speech and Hearing*, vol. 14, no. 6, pp. 181-188, 1992.

[34] D. Heckerman. *A Tutorial on Learning with Bayesian Networks*, in Learning in Graphical Models, M. Jordan (Ed), Cambridge, MIT Press, 1999.

[35] L.E. Brown, I. Tsamardinos, C.F. Aliferis. "A Novel Algorithm for Scalable and Accurate Bayesian Network Learning," *in Proceedings of the 11th World Congress on Medical Informatics (MEDINFO),* California, September 2004.

[36] J. Catlett, "On changing continuous attributes into ordered discrete attributes," *in Proceedings of Fifth European Working Session on Learning*, Berlin: Springer-Verlag, pp. 164–177, 1991

[37] V. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.

[38] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," *in Proceedings of ECML-98, 10th European Conference on Machine Learning*, pp. 137–142, Heidelberg, DE, 1998.

[39] C. Burges, "A tutorial on support vector machines for pattern recognition," *Knowledge Discovery and Data Mining*, U. Fayyad, Ed. Norwell, MA: Kluwer, pp.1-43, 2000.

[40] C. W. Hsu and C. J. Lin, "A comparison of methods for multi-class support vector machines," *IEEE Transactions on Neural Networks*, 13, pp. 415-425. 2002.

CHAPTER III: MANUSCRIPT 2

DYNAMIC DIFFICULTY ADJUSTMENT IN COMPUTER GAMES THROUGH REAL-TIME ANXIETY-BASED AFFECTIVE FEEDBACK

Changchun Liu, Pramila Agrawal, Nilanjan Sarkar, and Shuo Chen

(Accepted by International Journal of Human-Computer Interaction)

Abstract

A number of studies in recent years have investigated the dynamic difficulty adjustment (DDA) mechanism in computer games in order to automatically tailor gaming experience to individual player's characteristics. While most of these existing works focus on game adaptation based on player's performance, affective state experienced by the players could play a key role in gaming experience and may provide a useful indicator for a DDA mechanism. In this paper, an affect-based DDA was designed and implemented for computer games. In this DDA mechanism, a player's physiological signals were analyzed to infer his/her probable anxiety level, which was chosen as the target affective state, and the game difficulty level was automatically adjusted in real time as a function of the player's affective state. Peripheral physiological signals were measured through wearable biofeedback sensors and several physiological indices were explored to determine their correlations with anxiety. An experimental study was conducted to evaluate the effects of the affect-based DDA on game play by comparing it with a performance-based DDA. This is the first time, to our knowledge, that the impact of a real-time affect-based DDA has been demonstrated experimentally.

1. Introduction and Motivation

There has been a steady progress in the field of computer games in recent years that has become one of the most popular and economically successful forms of human-computer interaction (HCI) systems (Zaphiris, 2007). The worldwide market for computer game hardware, software and accessories is expected to grow from £11.7 billion in 2002 to £17 billion in 2007 (RocResearch, 2004) as more novel play environments are developed for entertainment and education (Stokes, 2005). While gaming technology has continued to evolve, there has been general dissatisfaction of players with the current computer games due to their inadequacy of providing optimal challenge levels to accommodate individual player's characteristics such as skills, capacities to learn and adapt, and emotional traits (Gilleade, Dix, & Allanson, 2005; Sweetser & Wyeth, 2005). Static difficulty levels that are manually selected by the players are not sufficient to avoid getting the player overwhelmed or bored since players are likely to be unable to assess which challenge level matches their skills (Koster, 2004). Additionally, asking the players to frequently choose the difficulty levels could be annoying as well as cause interruption of the game play (Chen, 2007).

In order to address this issue, a growing number of studies have been investigating the dynamic difficulty adjustment (DDA) mechanisms to enable the game-playing experiences automatically tailored to the individual characteristics. Demasi and Cruz have developed a "challenge function" by using heuristics (e.g., time to complete a task and rate of successful shots, etc.) to map a given game state to a value that specifies the difficulty felt by a user (Demasi & Cruz, 2002). Reinforcement learning has been employed to allow computer-controlled agent to learn optimal strategies in a fighting game while choosing suboptimal actions to fit the players' performance when necessary (Andrade et al., 2005). Spronck et al. proposed a rule-based approach, called dynamic scripting, that includes the model of the opponent player. It assigns each behavior rule a probability of being picked and then modifies the probability dynamically based on the success or failure rate of each rule (Spronck et al., 2006). The DDA has been increasingly recognized by the game development community as a key characteristic for a successful game. For instance, in Resident Evil 4 (www3.capcom.co.jp/bio4/english.html), a 3rd person shooter game with 5 levels of difficulty, the difficulty adjustment can be automatically accomplished based on the player's performance.

In most current DDA research works, the performance of the player has been used as a main measure of the characteristics of the players. However, as noted by Pagulayan et al. (2002), unlike productivity software, computer game's paramount evaluation factor should be the affective experience provided by the play environment instead of the user's performance. A case study on several popular computer games (e.g., Combat Flight Simulator (PC), Combat Evolved (Xbox), etc.) suggested that standard performance-based usability methods may not be sufficient to evaluate gaming experience and issues related to affective aspects of the game (e.g., fun) should be considered (Pagulayan, et al., 2005). Mandryk and Atkins (2007) also regarded the emotional experience is the key measurement of a game playing and used fuzzy physiological approach to determine the underlying affective states related to game play in an off-line manner. Echoing similar opinion, the concept of "Affective Gaming" has been proposed in recent years (Gilleade & Allanson, 2003; Gilleade, Dix, & Allanson, 2005; Magerkurth et al., 2005; Sykes & Brown, 2003), which focuses on exploring the impacts of affective factors in computer game design and adaptation. Furthermore, players can have different motivations to play a game (Koster, 2004). For a player who derives satisfaction from completing difficult tasks, one has to be cautious in decreasing the difficulty level even when he/she has been defeated for several times; whereas for another player, it may not be appropriate to increase the difficulty level even when his/her performance is excellent. We believe that the affective state of a player is likely to be a critical factor in many gaming experience, and that the next generation of DDA mechanism should consider both player's performance and affective state information.

The primary objective of this research is to explore the feasibility of recognizing a player's affective states via a physiology-based affect recognition technique during gaming and investigate how the gaming experience can be augmented by using the recognized affective state to automatically adjust game difficulty level in real-time. Note that we recognize the fact that a DDA mechanism that considers only affective state information may not be optimal. A versatile DDA mechanism should also consider player's performance, his/her personality, and the context and complexity of the game among other issues to generate a rewarding gaming experience. However, we first want to establish that real-time affective feedback is possible during the gaming process and that such a feedback can impact the experience of game play. The goal here is to advance the state-of-the-art in affective gaming, which has gained significant importance in the Human-Computer Interaction (HCI) community in recent years (Gilleade & Allanson, 2003; Gilleade, Dix, & Allanson, 2005; Magerkurth et al., 2005; Sykes & Brown, 2003). In order to achieve this objective, we divide our research into two major phases: (i) to obtain the affective model in Phase I, and (ii) to investigate the impact of affectsensitiveness on the gaming experience in Phase II. The primary contribution of this paper lies in Phase II work. However, since the Phase II work is dependent on affective models developed in Phase I, we believe it is necessary to briefly discuss Phase I work. The detailed results of the Phase I work were published in (Rani et al., 2006) and are omitted here. This is the first time, to our knowledge, that the impact of an affect-based DDA on player's interaction with a computer game that is capable of physiology-based affect recognition and real-time difficulty adjustment in a closed-loop manner has been investigated experimentally.

There are several modalities such as facial expression (Bartlett et al., 2003), vocal intonation (Lee & Narayanan, 2003), gestures (Asha et al., 2005), and physiology (Kulic & Croft, 2007; Leon, et al., 2004; Mandryk & Atkins, 2007; Rani et al., 2004) that can be utilized to recognize the affective states of individuals interacting with a computer. In this work we choose to create affective model based on physiological data for several reasons. One of the chief advantages of using physiology is that physiological signals are continuously available and are not dependent on overt emotional expression. Our aim is to recognize affective states of people engaged in real-life activities, such as playing computer games. Even if a person does not overtly express his/her emotion through speech, gestures or facial expression, a change in the physiological signal pattern associated with the changes of underlying affective states is likely to occur, which could be detectable. Furthermore, physiology is usually not under one's voluntary control and hence may provide an undiluted assessment of the underlying affective state. It is also reasonably independent of cultural, gender, and age related biases (Brown et al., 1997). Besides, there is evidence that the transition from one affective state to another is accompanied by dynamic shifts in indicators of Autonomic Nervous System (ANS) activity (Bradley, 2000; Picard, 1997). The physiological signals that have been used in this research consist of various cardiovascular, electrodermal, electromyographic, and body temperature signals, all of which have been extensively investigated in psychophysiology literature (Bradley, 2000).

An important question when estimating human affective state is how to represent the affective state. Although much existing research categorizes human affective states into what is called a set of "basic emotions," there is no consensus on a set of basic emotions among the researchers (Cowie et al., 2001). This fact implies that it requires pragmatic choices to select a target affective state for a given application (Cowie et al., 2001). In this paper, we chose anxiety to be the target affective state for the affect-based DDA design. The DDA mechanism will allow the computer game to recognize anxiety and respond to it in an appropriate manner. Anxiety was chosen for two primary reasons. First, anxiety plays an important role in various human-computer interaction tasks that can be related to performance, challenge, and ability (Brown et al., 1997; Chen, 2007). Second, the correlation of anxiety with physiology is well established in the psychophysiology literature (Rohrmann, Hennig, & Netter, 1999) and thus provides us with a scientific basis to infer it. In this study, we develop an affective model of a player that is capable of determining the intensity of anxiety (i.e., low/medium/high) instead of discrete emotions. Another important fact that should be noted for affective modeling is the phenomenon of person stereotypy. There is evidence that within a given context, different individuals express the same emotion with different physiological response patterns (Lacey & Lacey, 1958). The novelty of the presented affective modeling is that it is individual-specific in order to accommodate the differences encountered in emotion expression.

Note that a player's performance and affective state could be fused together in a DDA mechanism. However, in this paper we focus on the impact of an affect-based DDA on the gaming experience. Hence, we separated a performance-based DDA from an affect-based DDA and compared their effects on a computer game. Additionally, we implemented the DDA mechanism using state-flow diagrams, where the states were represented by a set of predefined difficulty levels. Although it is possible to use a player's affective state information to manipulate game environment settings and agents' behaviors in a moment-by-moment manner, such control often depends on heuristic knowledge and specific genre of a game (Hunicke & Chapman, 2004) and is beyond the scope of this paper. However, since most existing computer games have embedded predefined difficulty levels, the presented approach could be integrated with a large class of games.

The rest of the paper is organized as follows: the next section reports on related works in physiology-based affect recognition, intelligent tutoring system that used affective cues, and affective gaming. A description of the physiological signals and the features that were derived from these signals for affective modeling are presented in Section 3. In Section 4, we describe the machine learning algorithm used for detecting affective cues. Section 5 presents experimental designs for affective model building (Phase I) and evaluation of the effects of the affect-based DDA (Phase II). This is followed by a detailed results and discussion section (Section 6). Finally, Section 7 summarizes the contributions of the paper and provides future directions of this research.

2. Related Work

The use of physiology as a method to evaluate the affective state has attracted increasing attention in recent years. Multiple physiological measures such as electromyography (EMG), electroencephalography (EEG), and heart rate variability (HRV), have been used jointly to assess stress (Rani et al., 2002), workload (Kramer, Sirevaag, & Braune, 1987), and mental effort (Vicente, Thornton, & Moray, 1987). Galvanic Skin Response (GSR), EMG and Electrocardiogram (ECG) have been examined in (Mandryk & Atkins, 2007) to determine the underlying affective states related to game play. Various machine learning techniques including fuzzy logic (Mandryk & Atkins, 2007; Rani et al., 2002), discriminant function analysis (Nasoz et al., 2003), auto-associative neural networks (Leon et al., 2007), and support vector machines (Kim, Bang, & Kim, 2004) have been applied to differentiate discrete emotions (e.g., anger, joy, sadness etc.). In our previous work (Rani et al., 2004), we have shown the relationship between anxiety and several physiological parameters like HRV, facial EMG, GSR, blood pulse volume, and peripheral temperature. Although the existing studies provide valuable supports for the validity of physiology-based affect recognition, the impact on human users when computers respond to recognized affective states (i.e., interact in a closed-loop manner) is still largely unexplored.

In the context of intelligent tutoring system, there have been research efforts that aim at endowing a computerized tutor with the ability to adapt affectively in the teaching-learning process, which would permit a more natural, enjoyable and productive discourse. Conati (2002) proposed a probabilistic model to monitor a user's emotion and engagement during automated tutoring. The affective states of students (i.e., reproach, shame, and joy) were detected by the use of eye brow EMG, GSR and ECG through a dynamic decision network. The tradeoff between engagement and learning was achieved by a utility function that assigned appropriate weights to students' performance and engagement. Prendinger et al. (2005) conducted an experimental study that examined GSR and EMG to investigate the effect of a life-like virtual teacher on the affective state of users under "affective persona" and "non-affective persona" conditions. Our work differs from those studies in several aspects. First, our work focuses on investigating a DDA mechanism in the context of computer games. Specifically we are interested in evaluating the effects of an affect-based DDA on gaming experience by comparing it with a performance-based DDA though a systematic user study. Second, we identify the varying levels of anxiety instead of determining the occurrence of specific discrete emotions. Determining the intensity of an affective state could be a more challenging problem than differentiating discrete emotional states (Rani et al., 2006). Third, we adopt an individual-specific approach to overcome person-stereotypy and explore a more comprehensive set of physiological indices. We develop affective model for each individual player with reliable real-time predictions (as described in Section 6), whereas works in (Conati, 2002; Prendinger, Mori, & Ishizuka, 2005) presented across-individuals approach that did not consider person stereotypy.

Finally, our work falls into a nascent research field of HCI, called Affective Gaming (Gilleade & Allanson, 2003; Gilleade, Dix, & Allanson, 2005; Magerkurth et al., 2005; Sykes & Brown, 2003), that aims at enhancing gaming experience by adapting the game course to the player's affective state. Although concepts of affective gaming have been discussed for game design in (Gilleade & Allanson, 2003; Gilleade & Dix, 2004), the existing studies have the limitation of lacking systematic experimental investigation, which has been addressed in our work.

3. Physiological Indices for Recognizing Anxiety

There is good evidence that the physiological activity associated with the affective state can be differentiated and systematically organized (Bradley, 2000). The relationships between both electrodermal and cardiovascular activities with anxiety were investigated in (Dawson et al., 1990; Pagani, Lombardi, & Guzzetti, 1986; Rohrmann, Hennig, & Netter, 1999; Watts, 1975). When a human being is anxious, it is commonly observed that the parasympathetic activity of his/her heart decreases and the sympathetic activity increases (Pagani & Guzzetti, 1986). It was also reported that anxiety may cause an increase in skin conductance level (Watts, 1975). Previous research has validated blood pulse volume (FPV) measured at fingers is sensitive to the stress manipulation and is correlated with self-reported anxiety during the anticipation period (Bloom & Trautt, 1978). Measures of EMG activity of the chosen muscles (e.g., Corrugator Supercilii muscles) were also shown to be strong indicators of anxiety (Ekman & Friesen, 1986). In our work, we used this relationship between physiological response and the underlying affective states to develop an affect-recognition system.

The physiological signals we examined were: features of cardiovascular activity, including interbeat interval, relative pulse volume, pulse transit time, heart sound, and pre-ejection period; electrodermal activity (tonic and phasic response from skin conductance) and electromyogram (EMG) activity (from Corrugator Supercilii, Zygomaticus, and upper Trapezius muscles). These signals were selected because they are likely to demonstrate variability as a function of the targeted affective states, as well as they can be measured noninvasively, and are relatively resistant to movement artefacts (Lacey & Lacey, 1958).

Multiple features (as shown in Table 1) were derived for each physiological measure by using various signal processing techniques such as Fourier transform, wavelet transform, adaptive thresholding, and peak detection. Some of these features were described in our previous work (Rani et al., 2004). "Sym" is the power associated with the sympathetic nervous system activity of the heart (in the frequency band 0.04-0.15 Hz). "Para" is the power associated with the parasympathetic nervous system activity of the heart (in the frequency band 0.15-0.4 Hz). "VLF" is the power associated with the Very Low Frequency band (less than 0.04 Hz). Interbeat Interval (IBI) is the time interval in milliseconds between two "R" waves in the ECG waveform. "IBI ECG_{mean}" and "IBI ECG_{std}" are the mean and standard deviation of the IBI, respectively. The R-peak detection algorithm first performed band-pass filtering on the raw ECG signal. The resulting signal was then smoothed by a 10ms moving average window. Peaks were then detected in the resulting signal, and heuristic detection rules were applied to avoid missing R peaks or detecting multiple peaks for a single heart beat. These rules included obtaining the amplitude threshold (the difference between a peak and the corresponding inflection point) at which a peak should be considered a beat, enforcing a minimum interval of 300ms and maximum interval of 1500ms between peaks, checking for both positive and negative slopes in a peak to ensure that baseline drift is not misclassified as a peak, and allowing backtracking with reexamination/interpolation when peak missing was detected.

Table 1. Physiological Indices

Physiological Response	Features Derived	Label Used	Unit of Measurement
Cardiac activity	Sympathetic power (from ECG)	Sym	Unit/Square Second
	Parasympathetic power (from ECG)	Para	Unit/Square Second
	Very Low Frequency Power (from ECG)	VLF	Unit/Square Second
	Ratio of powers	Sym Para Para VLF Sym VLF	No unit
	Mean IBI	IBI ECG_{mean}	Milliseconds
	Std. of IBI	IBI ECG_{std}	Standard Deviation (no unit)
	Mean amplitude of the peak values of the PPG signal (Photoplethysmogram)	PPG $Peak_{mean}$	Micro Volts
	Standard deviation (Std.) of the peak values of the PPG signal	PPG $Peak_{std}$	Standard Deviation (no unit)
	Mean Pulse Transit Time	PTT_{mean}	Milliseconds
Heart Sound	Mean of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Mean d3 Mean d4 Mean d5	No unit
	Standard deviation of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Std d3 Std d4 Std d5	No unit
Bioimpedance	Mean Pre-Ejection Period	PEP_{mean}	Milliseconds
	Mean IBI	IBI ICG_{mean}	Milliseconds
Electrodermal activity	Mean tonic activity level	$Tonic_{mean}$	Micro-Siemens
	Slope of tonic activity	$Tonic_{slope}$	Micro-Siemens/Second
	Mean amplitude of skin conductance response (phasic activity)	$Phasic_{mean}$	Micro-Siemens
	Maximum amplitude of skin conductance response (phasic activity)	$Phasic_{max}$	Micro-Siemens
	Rate of phasic activity	Phasicrate	Response peaks/Second
Electromyographic activity	Mean of Corrugator Supercilii activity	Cor_{mean}	Micro Volts
	Std. of Corrugator Supercilii activity	Cor_{std}	Standard Deviation (no unit)
	Slope of Corrugator Supercilii activity	Cor_{slope}	Micro Volts/Second
	Mean Interbeat Interval of blink activity	IBI $Blink_{mean}$	Milliseconds
	Std. of Interbeat Interval of blink activity	IBI $Blink_{std}$	Standard Deviation (no unit)
	Mean amplitude of blink activity	Amp $Blink_{mean}$	Micro Volts
	Standard deviation of blink activity	$Blink_{std}$	Standard Deviation (no unit)
	Mean of Zygomaticus Major activity	Zyg_{mean}	Micro Volts
	Std. of Zygomaticus Major activity	Zyg_{std}	Standard Deviation (no unit)
	Slope of Zygomaticus Major activity	Zyg_{slope}	Micro Volts/Second
	Mean of Upper Trapezius activity	$Trap_{mean}$	Micro Volts
	Std. of Upper Trapezius activity	$Trap_{std}$	Standard Deviation (no unit)
	Slope of Upper Trapezius activity	$Trap_{slope}$	Micro Volts/Second
	Mean and Median frequency of Corrugator, Zygomaticus, and Trapezius	$Zfreq_{mean}$ $Cfreq_{median}$ $Tfreq_{mean}$	Hertz
Temperature	Mean temperature	$Temp_{mean}$	Degree Centigrade
	Slope of temperature	$Temp_{slope}$	Degree Centigrade/Second
	Std. of temperature	$Temp_{std}$	Standard Deviation (no unit)

Photoplethysmograph signal (PPG) measures changes in the volume of blood in the fingertip associated with the pulse cycle, and it provides an index of the relative constriction versus dilation of the blood vessels in the periphery. Pulse transit time (PTT) is the time it takes for the pulse pressure wave to travel from the heart to the periphery. PPT is determined by computing the time between systole at the heart (as indicated by the R-wave of the ECG) and the peak of the pulse wave reaching the peripheral site where PPG is being measured. Bioelectrical impedance analysis (BIA) measures the impedance or opposition to the flow of an electric current through the body fluids. A common variable in recent psychophysiology research, pre-ejection period (PEP) measures the latency between the onset of electromechanical systole and the onset of left-ventricular ejection. PEP is derived from impedance cardiogram (ICG) time-derivative and ECG and is most heavily influenced by sympathetic innervations of the heart. The peak detection mechanisms to determine the peaks of BVP and ICG time-derivative were similar to the ECG R-peak detection algorithm, while additional heuristic rules were added to reduce the degradation of the signal quality due to motion artifacts and avoid spurious peak detection with backtracking. Unlike ECG signals, the peak amplitudes of PPG and ICG showed a larger deviation over a given period of time. An adaptive thresholding rule was integrated in the peak detection algorithm to address this deviation, which continuously changed/updated the threshold value to determine whether candidates for peaks qualified to be valid peaks.

Electrodermal activity (EDA) consists of two main components - tonic response and phasic response. Tonic skin conductance refers to the ongoing or the baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event related changes that occur, caused by a momentary increase in skin conductance (resembling a peak). The raw EDA signal was smoothed by a 25ms moving average window and then down-sampled by 10 to remove the high frequency measurement noise. The phasic skin conductance detection algorithm used the following heuristics for considering a particular peak as a valid skin conductance response: (i) the slope of the rise to the peak should be greater than 0.05 microsiemens/minute; (ii) the amplitude should be greater than 0.05 microsiemens; and (iii) the rise time should be greater than 0.25 sec. All the signal points that were not included in the response constituted the tonic part of the skin conductance signal.

The EMG signal from Corrugator Supercilii muscle (eyebrow) captures a person's frowns and detects the tension in that region. This EMG signal is also a valuable source of blink information and helps determine the blink rate. The EMG signal from the Zygomaticus Major muscle captures the muscle movements while smiling. Upper Trapezius muscle activity measures the tension in the shoulders, one of the most common sites in the body for developing stress. The analysis of the EMG activities in the frequency domain involved applying Fast Fourier transform (FFT) on a given EMG signal, integrating the EMG spectrum, and normalizing it to [0,1] to calculate the two features of interest - the median frequency and mean frequency for each EMG signal. The blink-related features were determined from the corrugator supercilii EMG signals after being preprocessed by a low-pass filter (10 Hz).

The heart Sound signal measures sound generated during each heartbeat. These sounds are produced by blood turbulence primarily due to the closing of the valves within the heart. The features extracted from the heart sound signal consist of the mean and standard deviation of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform. Variations in the peripheral temperature mainly come from localized changes in blood flow caused by vascular resistance or arterial blood pressure and reflect the autonomic nervous system activity. The signal was down-sampled by 10 and filtered to remove high-frequency noise, from which the time-domain features (e.g., mean, SD, and slope) were calculated.

Any feature (derived from physiological signals) with an absolute correlation greater than equal to 0.3 with the target affective state was considered significant and was selected as inputs of the recognizers. It should be noted that the phenomenon of person-stereotypy makes it difficult to obtain universal patterns of emotions across individuals. As mentioned above, to overcome person-stereotypy we adopted an individual-specific approach where an affective model for each individual was developed in the Phase I study (e.g., we determine the physiological pattern of anxiety for each participant).

4. Anxiety Recognition based on Regression Tree

Determining a person's probable anxiety level from his/her physiological response resembles a classification problem where the attributes are physiological features and the target function is the anxiety level. The main challenge for this classification system to work, however, was the complex nature of the input physiological data sets. This complexity was primarily due to: (i) high dimensionality of the input feature space (there were 46 features), (ii) mixture of data types, and (iii) non-standard data structures. Additionally, a few physiological data were noisy where the biofeedback sensors had picked up movement artifacts.

In our earlier work (Rani et al., 2006), we compared several machine learning algorithms namely K nearest neighbors (KNN), Bayesian Network technique (BNT), Support Vector Machines (SVM) and Regression Tree (RT) for affect recognition from physiological signals and found that regression tree technique was efficient for affective modeling in terms of predictive accuracy and time and space cost. Hence in this work we employed RT to determine the underlying target affective state of a player given a set of physiological features. We omit the details of the theory and learning method of RT in this paper, which can be found in (Breiman et al., 1984) and our previous work (Rani et al., 2006).

5. Experimental Investigation

5.1 Subjects

Fifteen individuals (eight females, and seven males) volunteered to participate in the Phase I experiments. Their age ranged from 18 to 34 years, except for one participant, who was 54 years old. They were from diverse professional and ethnic backgrounds. They all had college degrees and had experience of playing computer games. Due to the nature of the tasks, the following were considered when choosing the participants: (i) their fluency in English, (ii) their familiarity with computers and ease of operation of keyboard and mouse, and (iii) general health (the absence of any problem in hearing or sensing). Participants were solicited through phone, emails and flyers posted around the Vanderbilt University area. They were given monetary compensation for their voluntary participation. Out of the fifteen participants, nine also took part in Phase II experiments.

The Institutional Review Board (IRB) approval was sought and received for conducting these experiments. In the IRB application, all details of the experiment were reported and it was emphasized that the health and safety of the participants was by no means endangered by participating in these experiments. It was also mentioned that the range of anxiety that the participants could experience was no more intense than the levels of anxiety that are commonly experienced in daily life. A detailed consent form was also drafted that acquainted the participants with the experimental procedure and their role in it. Participants were allowed to participate in the experiment only after their consent had been obtained through a signed consent form.

5.2 Game Design

Two computer games were designed and implemented to evoke varying intensities of anxiety from the participants. Physiological data from participants were collected during the experiments. The two games consisted of solving anagram and playing Pong. The anagram game has been previously employed to explore relationships between both electrodermal and cardiovascular activity with anxiety (Pecchinenda & Smith, 1996). Emotional responses were manipulated in this game by presenting the participants with anagrams of varying difficulty levels, as established through pilot work. An optimal mix of solvable anagrams caused low level of anxiety at times. Unsolvable or extremely difficult anagrams and giving time deadlines generated anxiety. In Pong sessions the participants played a variant of the classic computer game “Pong”. This game has also been used in the past by researchers to study anxiety and performance (Brown et al., 1997). Various parameters of the game were manipulated to elicit anxiety. These included ball speed and size, paddle speed and size, sluggish or over-responsive keyboard, and random keyboard response. The anxiety levels ranged from a low level of anxiety caused by a low ball speed and large sizes of the ball and the paddle, to high level of anxiety caused by a very high ball speeds and sluggish or over-responsive keyboard.

Each game session was subdivided into a series of discrete epochs that were bounded by self-reported affective state assessments. During the assessment, participants reported their perceived level of anxiety on a pop-up dialog box. It occurred every 3 minutes for the anagram game and every 2-4 minutes for the Pong game. This information was collected using a battery of self-report questions rated on a nine-point Likert scale where 1 indicated the lowest level and 9 indicated the maximum level. The reported level of anxiety were labeled and used for affective modeling in Phase I and assessing the real-time prediction performance of affective model in Phase II.

Based on a previous pilot study, different configurations of game parameters were determined to vary the difficulty level. During piloting, participants played a number of epochs of each game with selected configurations. After each epoch, the difficulty of the configuration perceived by them (on a nine-point Likert scale) were reported and recorded, as well as their performances, such as the number of correct answers for anagram game and the number of balls that they successfully hit in Pong game. After the piloting was over, these results were compiled to determine the perceived difficulty level of each configuration. The configurations were sorted and grouped according to their difficulty ratings. It was found that there were three distinct clusters of configurations that were well separated along the difficulty scale. These clusters were named Levels I, II and III, in the increasing order of difficulty. The averaged performance of participants for a given configuration was used to determine the threshold for that configuration.

In Phase II study, Pong game was used to study the impacts of an affect-based mechanism and a performance-based DDA mechanism to the gaming experience. The target number of hits (*TNH*) was defined as 10% higher than the average across the thresholds of all configurations for a given difficulty level. After each epoch was over, the participant's performance was rated as excellent ($hits \geq \lfloor 1.2TNH \rfloor$), good ($\lfloor 0.8TNH \rfloor \leq hits < \lfloor 1.2TNH \rfloor$) or poor ($hits < \lfloor 0.8TNH \rfloor$).

5.3 Phase I: Affective Modeling

In order to obtain physiological data to build affective models, the experiment in Phase I were designed to elicit varying intensities of the target affective state in participants as they played the computer games. We only provide brief but necessary information regarding affective modeling here. A detailed description of Phase I work can be found in (Rani et al., 2006). The training data set consisted of labeled self-reports of anxiety and various features (as described in Section 3) that were extracted off-line from the collected physiological data. The affective model was developed using the Regression Tree method. Each participant took part in a total of six sessions – three anagram game sessions and three pong game sessions. Each session was approximately one hour long and consisted of 16 epochs on an average. An epoch was a 2 - 4 minutes followed by self-reporting that usually lasted for an interval of 30 seconds to 1 minute. After the self-reporting the next epoch would begin. At the beginning of each session, baseline physiological signals were recorded in order to offset day-variability. Phase I study spanned a period of about two months.

In order to develop affective model, we built mappings from the extracted physiological features to the intensity (i.e., low/medium/high) of anxiety. This mapping was cast as a classification problem. The training datasets were formed by merging physiological features and self-reports of the participants as shown in Figure 1. The physiological data and self-reports were recorded in two separate files at the time of experiment. Later, the physiology data file was partitioned into the data blocks pertaining to every epoch in a separate file. Then, each epoch file was processed to extract the relevant features from the physiology data. The reports on anxiety was normalized to [0, 1] and then discretized such that 0–0.33 was labeled low, 0.34–0.67 was medium and 0.68–1.0 was labeled high. During the experiment, a total of 15 datasets were collected (one for each participant). Each dataset has n ($n \approx 96$) data vectors. The prediction accuracy of the developed model was evaluated by the leave-one-out cross validation method. Results of affective modeling can be found in Section 6.1.

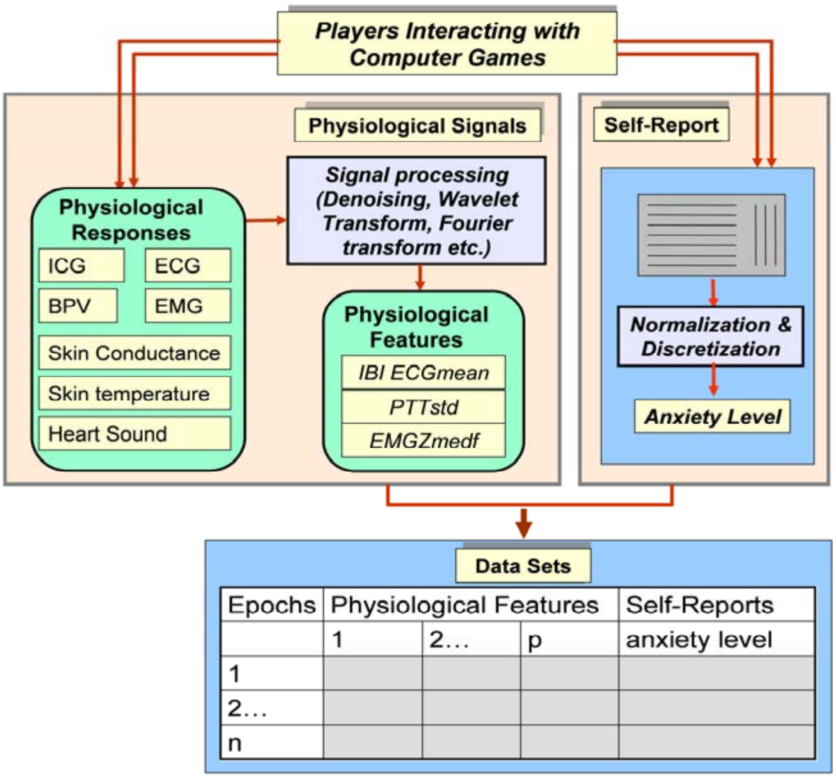


Figure1. Formation of Data Sets for Affective Modeling

5.4 Phase II: Affect-based Dynamic Difficulty Adjustment

In Phase II, two sessions of Pong game with two different DDA mechanisms were conducted for each participant: One in which the game difficulty was adapted based on player's performance; and another in which the real-time recognized player's anxiety level was employed to alter game difficulty. Phase II study spanned a period of about one month.

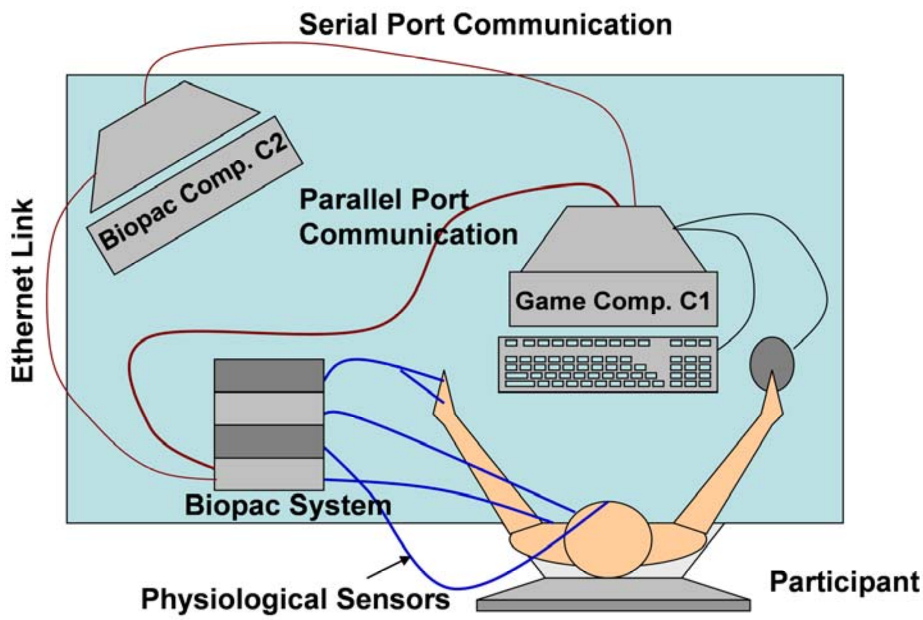


Figure 2. Phase II Experimental Set-up

5.4.1 Experiment setup

The set-up for the Pong game, which can adjust difficulty level dynamically based on recognized affective state/performance, is shown in Figure 2. The participant played the game on computer C1 while his/her physiological data was acquired via the Biopac system connected to C2. Physiological signals were transmitted from the Biopac transducers to C2 through an Ethernet link at 1000 Hz after being amplified and digitized by the Biopac system. C1 was also connected to the Biopac system via a parallel port, through which the game related event markers were recorded along with the physiological data in a time-synchronized manner. The serial communication between C1 and C2 enabled them to communicate with each other. C2 performed the following functions: (i) established serial communication with C1 to acquire the performance rating of the participant, (ii) acquired signals from Biopac system (that included the physiological signals and the event markers), (iii) ran signal processing routines to process the physiological data to extract features on-line, and (iv) used affective model developed in Phase I to recognize the anxiety level in real-time. Hence C2 could determine the affective state of the participant as well as his/her current performance. This information along with the knowledge of the current game difficulty was utilized to determine the next level of difficulty of the game. There was a serial communication protocol established between C1 and C2 that ensured that begin/end of Pong epochs on C1 was appropriately synchronized with the physiological data acquisition on C2.

5.4.2 Experimental design

Nine out of the fifteen participants who had taken part in Phase I study participated in Phase II experiments. Each of these nine participants played a total of two pong playing sessions ("Png1" and "Png2"). In Png1, the game difficulty was adjusted based on performance without any regard to the anxiety of the participant; while In Png2, the game difficulty was changed based on the real-time recognized anxiety level of the participant without regard to the performance.

Each Pong session was approximately 45 minutes long and consisted of 12 epochs of 2 minutes each. The remaining time was spent in set-up, attaching sensors, self-reports and taking breaks. After every epoch, the participant reported his/her assessment of one's own anxiety on a nine point Likert scale. In addition, at the end of each of the whole completed session, the participant answered questions pertaining to their overall experience during the entire session on a nine point Likert scale, which included their overall anxiety, enjoyment, challenge, and self-evaluation of their performance. These questions were asked to determine the aggregate gaming experience of each completed session (as described in Section 6.2).

During any Pong epoch, the game proceeded as follows:

1) A pop-up dialog box describing the rules of the game and other game-related instructions appeared on the game computer.
2) The participant was notified of the goal (number of minimum hits, maximum allowable misses and the time available) via a pop-up dialog box on the game computer.
3) Once the game started, the participant used the up and down arrow keys on the computer to control the paddle to hit the moving ball on-screen.
4) During any given epoch, the number of hits, misses and the number of seconds remaining were continuously updated on the bottom panel of the Pong screen.
5) After each epoch was over, the participant's performance was rated as excellent, good, or poor.
6) The end of a given epoch was followed by an interval of 30 seconds to 1 minute for self-reporting. After the self-reporting was completed, the next epoch would begin.

In Pong game in Phase II, three levels of difficulty - Level I (easy), Level II (moderately difficult), and Level III (very difficult) and three levels of performance -

poor, good, and excellent were identified. We also classified anxiety in three levels– low, medium and high. Figure 3 and Figure 4 show the state-flow models that were utilized to dynamically adjust difficulty level based on performance (P) and anxiety (A), respectively. It can be seen that in the performance-based DDA, excellent performance resulted in an increase in the level of difficulty (except when the player was already at the highest level), good performance caused the level to remain constant at the current level, and poor performance resulted in a decrease in difficulty level (except when the player was already at the lowest level). In the affect-based DDA, it can be seen that low anxiety resulted in increase in the level of difficulty (except when the player was already at the highest level), medium anxiety caused the level to remain constant at the current level, and high anxiety resulted in a decrease in difficulty level (except when the player was already at the lowest level).

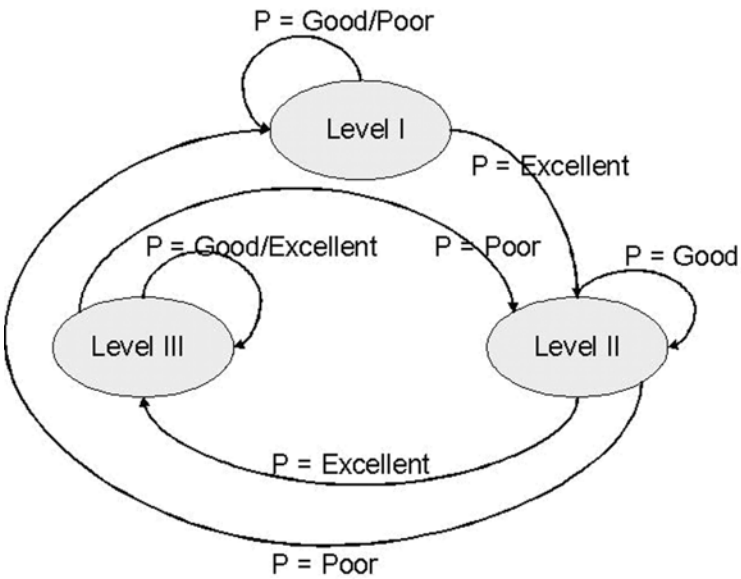


Figure 3. State-flow Diagram for Performance-Based DDA

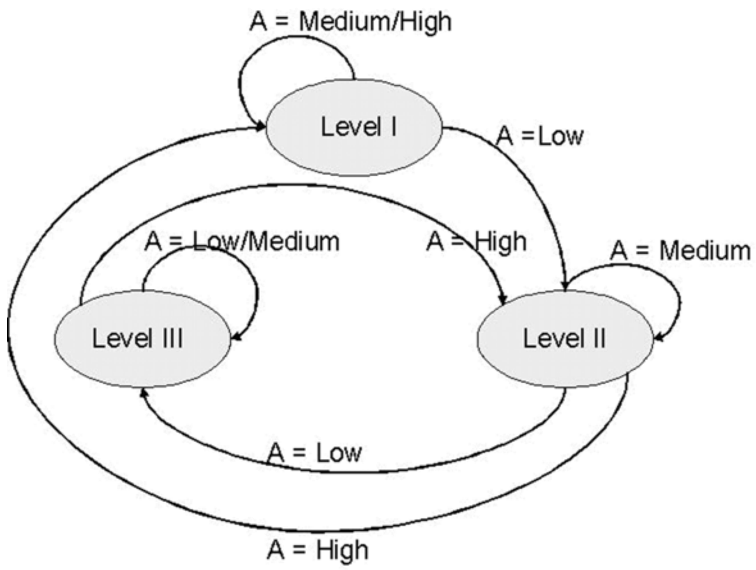


Figure 4. State-flow Diagram for Affect-Based DDA

The following conditions were imposed to avoid bias in data due to habituation, session-order, and to deal with day-variability: (i) in order to prevent habituation, at least 10 days time interval between any two Pong sessions was enforced; (ii) the sessions (performance-based and affect-based) were randomized to avoid any bias due to the order of sessions; (iii) all the other experimental conditions were kept constant over all sessions.

6. Results and Discussion

In this section we first briefly discuss the Phase I result that presents the off-line performance of the affective models. We then discuss the Phase II results from real-time closed-loop experiments in detail, which are the primary contribution of this work.

6.1 Phase I: Off-line Affective Modeling

In Phase I, we performed a comparative assessment of several machine learning methods for developing affective models. Figure 5 shows that the mean percentage accuracy (averaged across all the game epochs for all participants) to distinguish between different levels of anxiety were 88.5% for RT, 80.4% for KNN, 80.6% for BNT, and 88.9% for SVM. The results showed that all the above-mentioned methods performed well. This was in accordance with the claim of psychophysiologists that there is a distinct relationship between physiology and underlying affective states. Among the four machine learning techniques that were examined, both RT and SVM gave more reliable classification accuracy. However, it should be noted that RT does not require any parameter tuning, whereas in the case of SVM, choosing appropriate parameters (e.g., regularization parameter and kernel parameters) was imperative (Vapnik, 1998). A detailed analysis of the time and space efficiency of RT-based affective modeling can be found in our previous work (Rani et al., 2006). Since regression tree technique was efficient for anxiety recognition, it was used in Phase II for game difficulty adjustment.

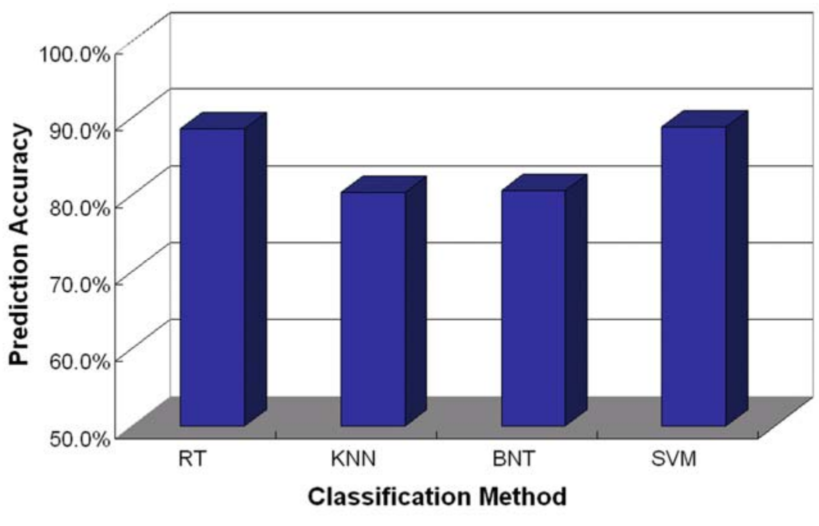


Figure 5. Prediction Accuracy forAffective State of Anxiety

6.2 Phase II: Comparison of Affect-based DDA with Performance-based DDA in Real-time Gaming

The results presented here are based on the validation Pong game sessions: Png1 (with performance-based DDA) and Png2 (with affect-based DDA). We observed several important results that are summarized below.

- *The real-time prediction accuracy of the affective models was high.* Once the affective modeling is accomplished in Phase I, the model can accept as input the physiological features, extracted on-line, and produce as output the probable level of anxiety of a participant when he/she is playing the computer game. The average real-time prediction accuracy, which represents how closely the on-line physiology-based quantitative measure of anxiety level matched with that of the subjective rating of anxiety, was 78% across all the 9 participants. Note that our affective model was evaluated based on physiological data obtained on-line from a real-time application. However, even then this real-time prediction accuracy is comparable to the results achieved through off-line analysis as reported in the literature (Kim, Bang, & Kim, 2004; Nasoz et al., 2003; Rani, 2005).
- *The performance of the majority of the participants improved during the affect-based DDA session.* The improvement in performance after the performance-based and anxiety-based sessions was shown in the "Performance" column of Table 2. In each of these sessions, the first and the last epoch were identical test epochs and the difference in the number of hits of the last and the first epoch gave the performance improvement. As can be seen that 7 out of 9 participants showed a greater improvement in performance after the affect-based session while 2 did not show any improvement (Participants 5 and 9).

Among the 7 participants who showed an improvement in affect-based game adaptation, 2 participants actually had a degradation of performance during the performance-based game adaptation. Using repeated measure ANOVA test, it was observed that the null hypothesis (asserting that there was no change in performance between performance-based and anxiety-based game sessions) could be rejected ($p < 0.05$).

Table 2. Perceived anxiety, Performance Improvement, Challenge, and Satisfaction Index (SI) across Performance-based (P) and Anxiety-based (A) Sessions

Participant ID	**Performance**		**Challenge**		**SI**		**Anxiety**	
	P	**A**	**P**	**A**	**P**	**A**	**P**	**A**
P1	2	5	5	5	13	13	5	1
P2	5	10	5	7	13	16	6	4
P3	1	5	5	7	10	17	6	3
P4	-3	10	4	5	14	11	2	6
P5	10	10	7	7	15	19	5	7
P6	3	6	8	9	22	23	8	5
P7	22	24	5	7	18	21	8	6
P8	-6	12	4	5	13	18	7	5
P9	0	0	4	7	14	16	5	5

- *Most participants perceived the game with the affect-based DDA to be more challenging than the one with the performance-based DDA.* At the end of each completed game session, the participants had reported the level of challenge that they had experienced and from this self-report, it was seen that most participants perceived the game with the affect-based DDA to be more challenging than the one with the performance-based DDA (“Challenge” column of Table 2). Except P1 and P5 who reported constant challenges across the two sessions, all the other participants reported an increase in challenge during the anxiety-based session. Using repeated measure ANOVA test, it was observed that the null hypothesis (asserting that there was no change in challenge between performance-based and anxiety-based sessions) could be rejected ($p < 0.01$).

• *Most participant perceived that the game with the affect-based DDA to be more satisfying than the one with the performance-based DDA.* An index called *Satisfaction Index* (SI) was defined by combining the values of challenge (C), enjoyment (E) and performance appraisal (P) reported by the participants at the end of each session.

$$SI = C + E + P \tag{1}$$

The *SI* could be a possible measure of the overall satisfaction of the participant during a given game session controlled by either the performance-based DDA or the affect-based DDA. There have been many efforts to develop metrics for measuring enjoyment in computer games, but no formal standards have yet been developed for evaluating fun, enjoyment or satisfaction. Echoing similar opinion, Wiberg (2005) states "research into the aspect of user satisfaction has so far been neglected in the research discipline of HCI …When discussing fun and entertainment in the context of usability, the most closely related notion is 'user satisfaction'". In the work by Sweetser and Wyeth (2005), the authors present a model of enjoyment based on eight elements -concentration, challenge, skills, control, clear goals, feedback, immersion, and social interaction. They claim that each of these elements contributes to achieving enjoyment in games. We used challenge, skill (as indicated by performance) along with a direct report on enjoyment to compute the *Satisfaction Index*.

"SI" column of Table 2 shows the values of SI during the two game sessions for each participant. 7 out of 9 participants reported an increase in the *SI* during the game session with the affect-based DDA. Of the 2 participants who did not report higher satisfaction during the anxiety-based session, P1 reported no change where as P4 reported a decrease in overall satisfaction. It should be noted that P4 also reported an increase in anxiety during the affect-based session, and P1 reported no change in perceived challenge during the two sessions. Using repeated measure ANOVA test, it was observed that the null hypothesis could be rejected ($p < 0.05$).

- *The perceived anxiety-level was reduced for the majority of the participants during the affect-based DDA session*. The results discussed so far suggest that anxiety-based DDA has positively influenced user satisfaction, feeling of challenge, and performance. In addition, we were also interested to know how the participant felt about their anxiety during gaming. The anxiety of the participants as reported by them (perceived anxiety) at the end of the completed affect-based session and the completed performance-based session was shown in the "Anxiety" column of Table 2. It can be seen that out of 9 participants, 6 reported a decrease in anxiety, 2 reported an increase and 1 reported no change in anxiety during the anxiety-based session as compared to the performance-based session. While the majority of the participants felt that they were less anxious when playing the game with affect-based DDA, it is interesting to note that no statistically significant difference in perceived anxiety was observed between the affect-based sessions and the performance-based sessions ($p =0.24$, repeated measure ANOVA) when using the reports of perceived anxiety collected at the end of the session. In order to explore the nature of perceived anxiety during the gaming process we analyzed the anxiety reports after each epoch, which we believe represent a more accurate record of perceived anxiety over the entire game duration. As described in Section 5.4.2, each session consisted of 12 epochs and besides the reports at the end of each completed session, a participant also reported his/her assessment of one's own anxiety after every epoch. This epoch-based anxiety reports may allow a finer grain analysis on the difference of perceived anxiety during the process of the game playing in the two conditions (anxiety-based vs. performance-based). A nested random-effect mixed model test was performed to evaluate the significance of the anxiety difference between the two sessions and it was observed that such difference was statistically significant (mean =3.41 for affect-based sessions and mean =4.42 for performance-based session, $p < 0.01$). Given the facts that majority of the participants felt less anxious after the affect-based sessions and that there was significant differences in the perceived anxiety during the game playing process in the two conditions, it suggests that by utilizing the information regarding the probable anxiety level of the participant to continuously adapt the game difficulty, the affect-based DDA has the potential to impact the gaming experience positively and keep the participants in a lower anxiety state.

7. Conclusions

In recent years several researchers have investigated dynamic difficulty adjustment (DDA) mechanisms to improve game-playing experiences such that the games can be automatically tailored to individual characteristics. However, most existing works on DDA mechanisms focus on player's performance as the determining factor. These DDAs do not possess the ability of deciphering affective cues of the players. While performance assessment is important and useful, affective states of the players can have major impacts on the gaming experience. This paper reported our efforts in developing an affect-based DDA mechanism to allow a computer game to infer and respond to the affective state while interacting with the players. The affective state (e.g., anxiety, in this case) was recognized using psychophysiological analysis. We explored a comprehensive set of physiological indices for affective modeling. The gaming experiences of the participants were evaluated and compared when a performance-based DDA mechanism and an affect-based DDA mechanism were applied to the same computer game. This is the first time, to our knowledge, that the impacts of an affect-base DDA to player's interaction with a computer game that is capable of physiology-based affect recognition and real-time difficulty adjustment in a closed-loop manner has been investigated experimentally.

Four machine learning methods were investigated to classify the anxiety level. A Regression Tree based affective model yielded reliable prediction with approximately 88% success while the other three approaches also performed competitively. When the developed RT-based model was applied in Phase II to recognize the anxiety level during the game play in real-time, it gave 78% correct predictions. While relatively less existing works investigated affect recognition in real-time applications and while further exploration in this direction is needed, this result suggested that physiology-based affective modeling provides a promising methodology to objectively quantify player's emotion when interacting with computer games. A systematic experimental study was conducted to evaluate the impacts of an affect-based DDA on the game play by comparing it with a performance-based DDA. It was observed that 6 out of 9 participants showed lower anxiety during the anxiety-based session than in the performance-based session, and 7 participants showed a greater improvement in performance during the anxiety-based session. 77% of the participants reported more challenging gaming experience and the overall satisfaction of gaming was enhanced by the affect-based DDA for majority of participants. These results suggest that gaming experience could be enhanced when a computer game is capable of recognizing player's affective states and adjusting game difficulty accordingly.

Note that the presented work requires physiological sensing that has its own limitations. For example, one needs to wear physiological sensors, and use of such sensors could be restrictive under certain circumstances. However, given the rapid progress in wearable computing, e.g., physiological sensing clothing and accessories (Jafari et al., 2005; Sung & Pentland, 2005; Wijesiriwardana et al., 2004), we believe that physiology-based affect recognition can be appropriate and useful to achieve affect-sensitive gaming.

One limitation of this work is that six one-hour gaming sessions were conducted for each participant in order to collect the training data for affective modeling. Further work is needed to reduce the length of time and the data for model building so that the affect-based DDA can be efficiently applied to game applications. The next research goal would be to explore the trade-off between prediction accuracy and training set size and investigate new machine learning techniques to optimize training data to compensate for its scarcity. Active learning (Vijayakumar & Ogawa, 1999) is one method that could hold promise for such a purpose. Active learning method can assume some control over what next game epoch to be introduced during the affective modeling process to get a more informative training point. It is also expected that the required training process would be reduced when the player's physiology is used together with other channels of affect-related information, such as eye gaze (Prendinger, Ma, & Ishizuka, 2007) and posture (Tan, Slivovsky, & Pentland, 2001). The presented work, however, demonstrated the feasibility that a player's affective state can be deciphered from his/her physiological response during gaming and a DDA mechanism can be designed that can adjust the game difficulty in real-time based on the affective state information. The experimental investigation showed the benefits of such a DDA mechanism. It is expected these results will encourage future research into affect-based DDA design for computer games. Additionally, besides anxiety, other affective states (e.g., excitement and frustration) are also considered to be important in game playing (Gilleade & Dix, 2004; Mandryk, et al., 2007). While the affective modeling methodology in this work could be used to detect the intensity of anxiety, excitement, and frustration simultaneously, more sophisticated difficulty adaptation mechanisms would be demanded to incorporate multiple inferred affective cues and account for other game playing information of interests, such as the player's performance, his/her personality, and the context and complexity of the game. We will investigate fast and robust DDA mechanisms that would permit a computer's adaptive response in the more complex gaming applications and allow the affect-sensitive DDA to be adopted in the future computer games.

References

Andrade, G., Ramalho, G., Santana, H., & Corruble, V. (2005). *Challenge-Sensitive Action Selection: an Application to Game Balancing.* Paper presented at the IEEE/WIC/ACM International Conference on Intelligent Agent Technology Compiègne, France.

Asha, K., Ajay, K., Naznin, V., George, T., & Peter, F. D. (2005). *Gesture-based affective computing on motion capmture data.* Paper presented at the Int. Conf. on Affective Computing and Intelligent Interaction (ACII), Beijing, China.

Bartlett, M. S., Littlewort, G., Fasel, I., & Movellan, J. R. (2003). *Real time face detection and facial expression recognition: development and applications to human computer interaction.* Paper presented at the Computer Vision and Pattern

Recognition Workshop, Madison, Wisconsin.

Bloom, J. & Trautt, G. M. (1978). Finger pulse volume as a measure of anxiety: Further evaluation. *Psychophysiology*, 14(6), 541-544.

Bradley, M. M. (2000). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary & G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 602-642). New York: Cambridge University Press.

Breiman L., Friedman J. H., Olshen R. A., & Stone C. J. (1984). *Classification and Regression Trees*. Chapman & Hall/CRC.

Brown, R. M., Hall, L. R., Holtzer, R., Brown, S. L., & Brown, N. L. (1997). Gender and Video Game Performance. *Sex Roles, 36*(11-12), 793-812.

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM, 50*(4), 31-34.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence, 16*(7-8), 555-575.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., et al. (2001). Emotion recognition in human-computer interaction. *Ieee Signal Processing Magazine, 18*(1), 32-80.

Dawson, M. E., Schell, A. M., & Filion, D. L. (1990). The electrodermal system. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of Psychophysiology: Physical, social, and inferential elements* (pp. 295-324). New York: Cambridge University Press.

Demasi, P., & Cruz, A. (2002). *Online Coevolution for Action Games.* Paper presented at the 3rd international conference on intelligent games and simulation, London.

Ekman, P., & Friesen, W. V. (1986). A new pan cultural facial expression of emotion. *Motivation and Emotion*, 10(2), 159-168.

Gilleade, K. M., & Allanson, J. (2003). *A Toolkit for Exploring Affective Interface Adaptation in Videogames.* Paper presented at the HCI International 2003, New Jersey.

Gilleade, K. M., & Dix, A. (2004). *Using frustration in the design of adaptive videogames.* Paper presented at the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology, Singapore.

Gilleade, K., Dix, A., & Allanson, J. (2005). Affective Videogames and Modes of Affective Gaming: Assist Me, Challenge Me, Emote Me, *Proceedings of DIGRA 2005*.

Hunicke, R., & Chapman, V. (2004). AI for Dynamic Difficult Adjustment in Games, *Challenges in Game Artificial Intelligence AAAI Workshop* (pp. 91-96). San Jose.

Jafari, R., Dabiri, F., Brisk, P., & Sarrafzadeh, M. (2005). *Adaptive and fault tolerant medical vest for life critical medical monitoring.* Paper presented at the the 20th ACM Symposium on Applied Computing Santa Fe, NM.

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing, 42*(3), 419-427.

Koster, R. (2004). *Theory of Fun for Game Design*. Phoenix: Paraglyph Press.

Kramer, A. F., Sirevaag, E. J., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated fight missions. *Human Factors, 29*(2), 145-160.

Kulic, D., & Croft, E. (2007). Physiological and subjective responses to articulated robot motion. *Robotica, 25*, 13-27.

Lacey, J. I., & Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy. *Am J Psychol, 71*(1), 50-73.

Lee, C. M., & Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing, 13*(2), 293-303.

Leon, E., Clarke, G., Callaghan, V., & Sepulveda, F. (2004). Real-time detection of emotional changes for inhabited environments. *Computers & Graphics-Uk, 28*(5), 635-642.

Leon, E., Clarke, G., Callaghan, V., & Sepulveda, F. (2007). A user-independent real-time emotion recognition system for software agents in domestic environments.

Engineering Applications of Artificial Intelligence, 20(3), 337-345.

Magerkurth, C., Cheok, A. D., Mandryk, R. L., & Nilsen, T. (2005). Pervasive games: bringing computer entertainment back to the real world. *ACM Computers in Entertainment, 3*(3), 11-29.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies, 65*(4), 329-347

Nasoz, F., Alvarez, K., Lisetti, C. L., & Finkelstein, N. (2003). Emotion recognition from physiological signals for presence technologies. *International Journal of Cognition, Technology, and Work – Special Issue on Presence, 6*(1).

Pagani, M., Lombardi, F., & Guzzetti, S. (1986). Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympathovagal interaction in man and conscious dog. *Circulation Research*, 59, 178-193.

Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R. L., & Fuller, T. (2002). User-centered design in games. In J. A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (pp. 883-906). Mahwah, NJ: Lawrence Erlbaum Associates.

Pagulayan, R. J., Steury, K. R., Fulton, B., Romero, R. L. (2005). Designing for fun: user testing case studies. In Blythe, M. A. et al.(Eds), *Funology; From Usability to Enjoyment* (pp. 137-150). Kluwer Academic Publishers: Dordrecht.

Pecchinenda, A., & Smith, C. A. (1996). The affective significance of skin conductance activity during a difficult problem-solving task. *Cognition & Emotion, 10*(5), 481-503.

Picard, R. W. (1997). *Affective Computing*. Cambridge: The MIT Press.

Prendinger, H., Ma, C. L., & Ishizuka, M. (2007). Eye movements as indices for the utility of life-like interface agents: A pilot study. *Interacting with Computers, 19*(2), 281-292.

Prendinger, H., Mori, J., & 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.

Rani, P. (December 2005). *Psychophysiology based Affective Communication for Implicit human-Robot Interaction.* Doctoral Dissertation,Vanderbilt University, Nashville.

Rani, P., Liu, C. C., Sarkar, N., & Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis and Applications, 9*(1), 58-69.

Rani, P., Sarkar, N., Smith, C. A., & Kirby, L. D. (2004). Anxiety detecting robotic system - towards implicit human-robot collaboration. *Robotica, 22*, 85-95.

Rani, P., Sims, J., Brackin, R., & Sarkar, N. (2002). Online stress detection using psychophysiological signals for implicit human-robot cooperation. *Robotica, 20*, 673-685.

RocResearch. (2004). *Video Game Industry*: RocResearch Ltd.

Rohrmann, S., Hennig, J., & Netter, P. (1999). Changing psychobiological stress reactions by manipulating cognitive processes. *International Journal of Psychophysiology, 33*(2), 149-161.

Sykes, J. and Brown, S. (2003). Affective gaming. Measuring emotion through the gamepad. In Proceedings of *CHI-03*, 732–733.

Spronck, P., Ponsen, M., Sprinkhuizen-Kuyper, I., & Postma, E. (2006). Adaptive Game AI with Dynamic Scripting. *Machine Learning, 63*(3), 217-248.

Stokes, B. (2005). Videogames have changed: time to consider Serious Games? *The Development Education Journal 11*(2).

Sung, M., & Pentland, A. S. (2005). Minimally-Invasive Physiological Sensing for Human-Aware Interfaces, *HCI International 2005*. Las Vegas, USA.

Sweetser, P., & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment, 3*(3).

Tan, H. Z., Slivovsky, L. A., & Pentland, A. (2001). A sensing chair using pressure distribution sensors. *Ieee-Asme Transactions on Mechatronics, 6*(3), 261-268.

Vapnik, V. (1998). *Statistical learning theory*. New York: Wiley.

Vicente, K. J., Thornton, D. C., & Moray, N. (1987). Spectral-Analysis of Sinus Arrhythmia - a Measure of Mental Effort. *Human Factors, 29*(2), 171-182.

Vijayakumar, S., & Ogawa, H. (1999). Improving generalization ability through active learning. *Ieice Transactions on Information and Systems, E82d*(2), 480-487.

Watts, J. M. J. (1975). Anxiety and the habituation of the skin conductance response. *Psychophysiology*, 12(5), 596-601.

Wiberg, C. (2005). Affective computing vs. usability? Insights of using traditional usability evaluation methods, *CHI 2005 Workshop on Innovative Approaches to Evaluating Affective Interfaces*. Portland, USA.

Wijesiriwardana, R., Mitcham, K., & Dias, T. (2004). *Fibre-meshed transducers based real time wearable physiological information monitoring system.* Paper presented at the International Symposium on Wearable Computers, Washington, DC.

Zaphiris, P. (2007). HCI issues in computer games. *Interacting with Computers, 19*(2), 135-139.

CHAPTER IV: MANUSCRIPT 3

INTERACTION BETWEEN HUMAN AND ROBOT – AN AFFECT-INSPIRED APPROACH

Pramila Rani, Changchun Liu, & Nilanjan Sarkar

(Published by Interaction Studies)

Abstract

This paper presents a human-robot interaction framework where a robot can infer implicit affective cues of a human and respond to them appropriately. Affective cues are inferred by the robot in real-time from physiological signals. A robot-based basketball game is designed where a robotic "coach" monitors the human participant's anxiety to dynamically reconfigure game parameters to allow skill improvement while maintaining desired anxiety levels. The results of the above-mentioned anxiety-based sessions are compared with performance-based sessions where in the latter sessions, the game is adapted only according to the player's performance. It was observed that 79% of the participants showed lower anxiety during anxiety-based session than in the performance-based session, 65% showed a greater improvement in performance after the anxiety-based session and 71% of the participants reported greater overall satisfaction during the anxiety-based sessions. This is the first time, to our knowledge, that the impact of real-time affective communication between a robot and a human has been demonstrated experimentally.

Key words: Human-robot interaction, implicit communication, physiological sensing, affective computing, anxiety, closed-loop interaction

1. Introduction

There has been a steady progress in the field of intelligent and interactive robotics over the last two decades ushering in a new era of personal and service robots. The World Robotics 2005 survey (http://www.unece.org/press/pr2005/05stat_p03e.pdf) reports that over 1,000,000 household robots were in use last year, a number that is anticipated to exceed several million in the next few years. As robots and humans begin to co-exist and cooperatively share a variety of tasks, "natural" human-robot interaction that resembles human interaction is becoming increasingly important.

Reeves and Nass (1996) have shown that people's interactions with computers and similar machines/media are fundamentally social and natural. Human interactions are characterized by explicit as well as implicit channels of communication. While the explicit channel transmits overt messages, the implicit one transmits hidden messages about the communicator (his/her intention, attitude and like, dislike). Ensuring sensitivity to the other party's emotions or sensibility is one of the key tasks associated with the second, implicit channel (Cowie, Douglas-Cowie, Tsapatsoulis, Votsis, Kollias, Fellenz & Taylor 2001). In this context, research in (Mehrabian & Friar 1969) found that approximately 93% of the emotional meaning of a message is communicated implicitly through non-verbal channels. Therefore, endowing robots with an implicit communication channel and a degree of emotional intelligence should permit more meaningful and natural human-robot interaction (Picard 1997).

The potential applications of robots that can detect a person's affective states and interact with him/her based on such perception could be varied and numerous. Whether it is the domain of personal home aids that assist in cleaning and transportation, toy robots that engage and entertain children, interactive tutoring agents that help students learn better, professional service robots that act as assistants in offices, hospitals, and museums, or search, rescue and surveillance robots that accompany soldiers and firefighters – this novel aspect of human-robot interaction could impact them all.

For a robot to be emotionally intelligent it should clearly have a two-fold capability - the ability to display its own emotions (Fong, Nourbakhsh & Dautenhahn 2003, Kanda, Ishiguro, Ono, Imai & Nakatsu 2002) and the ability to understand human emotions and motivations (also referred to as affective states). There are several works that focus on making robot display emotions just like human beings – usually by using facial expressions and speech (Breazeal & Aryananda 2002, Haritaoglu, Cozzi, Koons, Flickner, Yacoob, Zotkin & Duriswami 2001, Hoffman & Breazeal 2004, Kanda, Ishiguro, Ono, Imai & Nakatsu 2002). We do not address this issue in this paper. Our work is complementary to this body of research. The focus of our work is to address the later capability, i.e., how to endow a robot with the ability to recognize human affective states. Specifically, in this work, we choose anxiety to be the target affective state that the robot must detect and be responsive to.

There are several modalities such as facial expression (Bartlett, Littlewort, Fasel & Movellan 2003), vocal intonation (Lee & Narayanan 2005), gestures and postures (Kapur, Kapur, Naznin, George & Peter 2005, Kleinsmith, Fushimi & Bianchi-Berthouze 2005, Kleinsmith, Ravindra De Silva & Bianchi-Berthouze 2005), and physiology (Picard, Vyzas & Healy 2001) that can be utilized to determine the underlying affective state of a person interacting with the robot. In this work, we choose physiology to infer affect due to several reasons. Physiological phenomena have been reliably correlated with affective states in the psychophysiology literature (Bradley 2000). Physiological modality circumvents several limitations of the vision and speech based methods. One of the chief advantages of using physiology is that physiological signals are continuously available and are not dependent on overt emotion expression. Hence, physiology-based affect detection could be very useful in situations where it is not possible to continuously monitor facial expressions of a person or in scenarios where a person's speech is not available to interpret his/her underlying emotion. Unobtrusive, small, wireless physiological sensors may be an ideal solution in these cases for real-time affect monitoring. Physiology is usually not under voluntary control and hence provides an undiluted assessment of the underlying affective state. It is also reasonably independent of cultural, gender and age related biases (Bradley 2000). This is, of course, not to say that physiological sensing does not have its own limitations. One main limitation is that one needs to wear physiological sensors. It may not be possible to do so under certain circumstances. In addition, the body part where such sensors are placed may not be fully functional. However, we believe that there are many situations where physiology-based implicit communication could be appropriate and useful.

While concepts from psychophysiology have been applied to human-computer interaction for many years now, the application of this technique in robotics domain is relatively new (Kulic & Croft 2003). Our preliminary work in (Rani, Sarkar, Smith & Kirby 2004) presented concepts and initial open loop results for a natural and intuitive human-robot interaction framework based on detection of human affective states from physiological signals.

In this paper, we choose anxiety to be the target affective state that the robot must detect and be responsive to. Anxiety is chosen for two primary reasons. First, anxiety plays an important role in various human-machine interaction tasks that can be related to task performance (Brown, Hall, Holtzer & Brown 1997, Hirata 1990). Second, the correlation of anxiety with physiology is well established in psychophysiology literature (Rohrmann, Hennig & Netter 1999) thus, providing a scientific basis to infer it.

The primary objective of this research is to investigate how human-robot interaction can be augmented by closed-loop implicit affective communication. In order to achieve this objective, we divide the research into several components: i) designing a human-robot task that can elicit anxiety in the participant; ii) implementing signal processing and machine learning techniques for anxiety detection; and iii) designing a robot control architecture that can respond to participant's anxiety. In this paper, we omit a discussion on signal processing and machine learning techniques for anxiety detection, which can be found in our previous work (Liu, Rani & Sarkar 2005, Rani, Liu, Sarkar & Vanman 2006). We also do not discuss details of the control architecture here. Instead, we present stateflow diagrams based on which the robot controller interacts with the human. We mostly focus on the human-robot interaction task design and its experimental evaluation in this paper. More specifically, we want to investigate whether such implicit communication facilitates the interactive learning (e.g., achieving the desired skill in a more affect-sensitive and individual-specific manner through closed-loop interactions). Emphasis has been placed on the design of computer-based interactive learning environment. To the best of our knowledge, learning with a robot that is capable of physiology-based emotion recognition and real-time behavior adjustment in closed interaction loop has not been done before.

The paper is organized as follows: the related literature survey is contained in Section 2. Section 3 briefly describes the physiological indices and learning algorithm used for detecting affective cues in human-robot interaction. In Section 4, the tasks that were utilized for model building are described. Section 5 presents in detail the human-robot closed-loop interaction task. This is followed by a detailed results and discussion section (Section 6). Finally, Section 7 summarizes the contributions and conclusions of the paper and provides the future directions of research.

2. Related Research

In this section, we give a brief overview of the research in psychophysiology aimed at detecting human mental state, as well as past work in interactive learning.

There is a rich history in the human factors and psychophysiology literature to understand occupational stress, operator workload (Kramer, Sirevaag & Braune 1987), operator mental effort (Vicente, Thornton & Moray 1987) and other similar mental states based on physiological measures such as those derived from electromyography (EMG), electroencephalography (EEG), and heart rate variability (HRV). Multiple psychophysiological measures such as HRV, EEG, blink rates and others have been used jointly in recent years to assess pilots’ and drivers' workload (Wilson 2002). Heart period variability (HPV) has been shown to be an important parameter for assessing mental workload relevant in human-computer interface (HCI) (Iszo, Mischinger & Lang 1999). In our previous work (Rani, Sarkar, Smith & Kirby 2004) we have shown the relationship between anxiety and several physiological parameters like HRV, facial EMG, skin conductance, blood pulse volume, and peripheral temperature. Prinzel et al. have studied the effect of an EEG based adaptive automation on tracking performance and workload (Prinzel, Freeman, Scerbo, Mikulka & Pope 2003). In another work by Nasoz et al. (Nasoz, Ozyer, Lisetti & Finkelstein 2002), physiological signals such as galvanic skin response, heartbeat, and temperature were utilized to create a multimodal Affective Driver Interface for the drivers of the future cars. Kulic et al. discuss their approach to estimate intent for human-robot interaction in (Kulic & Croft 2003). They focus on the two aspects of intent namely attention and approval, where attention was measured through gesture recognition and eye gaze tracking and approval was measured through facial expressions and physiological signals. Operator physiological response was also studied by Hanajima et al. who investigated the impact of robot motion on operator's HRV and electrodermal activity (Hanajima, Ohta, Hikita & Yamashita 2005).

Interactive learning has been the focus of research in recent years. Many works involve assessing a learner's affective state during automated tutoring sessions. In (Conati 2002), focus is on improving user interaction in educational computer games by achieving a trade-off between engagement and learning. In (Graesser, Wiemer-Hastings, Wiemer-Hastings & Kreuz 1999), AutoTutor - a computerized tutor has been developed which serves as a learning scaffold to assists students by simulating the discourse patterns and pedagogical strategies of a human tutor. Kapoor et al. in (Kapoor, Mota & Picard 2001) present preliminary work done in the area of developing a Learning Companion, a computer-based system that is responsive to the affective aspects of a learner.

3. Physiological Indices and Learning Algorithm

There is good evidence that the physiological activity associated with the affective state can be differentiated and systematically organized. The transition from one affective state to another, for instance, from relaxed to anxiety state is accompanied by dynamic shifts in indicators of Autonomic Nervous System (ANS) activity. In our work, we used this relationship between physiological response and underlying affective states to develop an affect-recognition system. The physiological signals we examined are: various features of cardiovascular activity, including interbeat interval, relative pulse volume, pulse transit time, heart sound, and pre-ejection period; electrodermal activity (tonic and phasic response from skin conductance); and electromyogram (EMG) activity (from Corrugator Supercilii, Zygomaticus, and upper Trapezius muscles). These signals were selected because they can be measured non-invasively and are relatively resistant to movement artifacts.

Multiple features (as shown in Table 1) were derived for each physiological measure. Some of these features are described in our previous work (Rani, Liu, Sarkar & Vanman 2006). "Sym" is the power associated with the sympathetic nervous system activity of the heart (in the frequency band 0.04-0.15 Hz.). "Para" is the power associated with the parasympathetic nervous system activity of the heart (in the frequency band 0.15-0.4 Hz.). "VLF" is the power associated with the Very Low Frequency band (less than 0.04 Hz.). InterBeat Interval (IBI) is the time interval in milliseconds between two "R" waves in the ECG waveform in millisecond. "IBI ECGmean" and "IBI ECGstd" are the mean and standard deviation of the IBI. Photoplethysmograph signal (PPG) measures changes in the volume of blood in the fingertip associated with the pulse cycle, and it provides an index of the relative constriction versus dilation of the blood vessels in the periphery. Pulse transit time (PTT) is the time it takes for the pulse pressure wave to travel from the heart to the periphery, and it is estimated by computing the time between systole at the heart (as indicated by the R-wave of the ECG) and the peak of the pulse wave reaching the peripheral site where PPG is being measured. Heart sound signal measures sounds generated during each heartbeat. These sounds are produced by blood turbulence primarily due to the closing of the valves within the heart. The features extracted from the heart sound signal consisted of the mean and standard deviation of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform. Bioelectrical impedance analysis (BIA) measures the impedance or opposition to the flow of an electric current through the body fluids contained mainly in the lean and fat tissue. A common variable in recent psychophysiology research, pre-ejection period (PEP) - derived from impedance cardiogram (ICG) and ECG, measures the latency between the onset of electromechanical systole, and the onset of left-ventricular ejection. PEP is most heavily influenced by sympathetic innervations of the heart. Electrodermal activity consists of two main components - tonic response and phasic response. Tonic skin conductance refers to the ongoing or the baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event related changes that occur, caused by a momentary increase in skin conductance (resembling a peak). The EMG signal from Corrugator Supercilii muscle (eyebrow) captures a person's frowns and detects the tension in that region. It is also a valuable source of blink information and helps us determine the blink rate. The EMG signal from the Zygomaticus Major muscle captures the muscle movements while smiling. upper Trapezius muscle activity measures the tension in the shoulders, one of the most common sites in the body for developing stress.

Table 1. Physiological Indices

Physiological Response	Features Derived	Label Used	Unit of Measurement
Cardiac activity	Sympathetic power (from ECG)	Sym	Unit/Square Second
	Parasympathetic power (from ECG)	Para	Unit/Square Second
	Very Low Frequency Power (from ECG)	VLF	Unit/Square Second
	Ratio of powers	Sym Para Para VLF Sym VLF	No unit
	Mean IBI	IBI ECG_{mean}	Milliseconds
	Std. of IBI	IBI ECG_{std}	Standard Deviation (no unit)
	Mean amplitude of the peak values of the PPG signal (Photoplethysmogram)	PPG $Peak_{mean}$	Micro Volts
	Standard deviation (Std.) of the peak values of the PPG signal	PPG $Peak_{std}$	Standard Deviation (no unit)
	Mean Pulse Transit Time	PTT_{mean}	Milliseconds
Heart Sound	Mean of the 3rd,4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Mean d3 Mean d4 Mean d5	No unit
	Standard deviation of the 3rd,4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Std d3 Std d4 Std d5	No unit
Bioimpedance	Mean Pre-Ejection Period	PEP_{mean}	Milliseconds
	Mean IBI	IBI ICG_{mean}	Milliseconds
Electrodermal activity	Mean tonic activity level	$Tonic_{mean}$	Micro-Siemens
	Slope of tonic activity	$Tonic_{slope}$	Micro-Siemens/Second
	Mean amplitude of skin conductance response (phasic activity)	$Phasic_{mean}$	Micro-Siemens
	Maximum amplitude of skin conductance response (phasic activity)	$Phasic_{max}$	Micro-Siemens
	Rate of phasic activity	$Phasic_{rate}$	Response peaks/Second
Electromyographic activity	Mean of Corrugator Supercilii activity	Cor_{mean}	Micro Volts
	Std. of Corrugator Supercilii activity	Cor_{std}	Standard Deviation (no unit)
	Slope. of Corrugator Supercilii activity	Cor_{slope}	Micro Volts/Second
	Mean Interbeat Interval of blink activity	IBI $Blink_{mean}$	Milliseconds
	Std. of Interbeat Interval of blink activity	IBI $Blink_{std}$	Standard Deviation (no unit)
	Mean amplitude of blink activity	Amp $Blink_{mean}$	Micro Volts
	Standard deviation of blink activity	$Blink_{std}$	Standard Deviation (no unit)
	Mean of Zygomaticus Major activity	Zyg_{mean}	Micro Volts
	Std. of Zygomaticus Major activity	Zyg_{std}	Standard Deviation (no unit)
	Slope. of Zygomaticus Major activity	Zyg_{slope}	Micro Volts/Second
	Mean of Upper Trapezius activity	$Trap_{mean}$	Micro Volts
	Std. of Upper Trapezius activity	$Trap_{std}$	Standard Deviation (no unit)
	Slope. of Upper Trapezius activity	$Trap_{slope}$	Micro Volts/Second
	Mean and Median frequency of Corrugator, Zygomaticus and Trapezius	$Zfreq_{mean}$ $Cfreq_{median}$ $Tfreq_{mean}$ etc.	Hertz
Temperature	Mean temperature	$Temp_{mean}$	Degree Centigrade
	Slope of temperature	$Temp_{slope}$	Degree Centigrade/Second
	Std. of temperature	$Temp_{std}$	Standard Deviation (no unit)

Various signal processing techniques such as Fourier transform, wavelet transform, adaptive thresholding, and peak detection, were used to derive relevant features from the physiological signals. All these features are powerful indicators of the underlying affective state of the person showing this response. We have exploited this dependence of a person's physiological response on affect to detect and identify affective states in real-time using advanced signal processing techniques. It should be noted that the phenomenon of person-stereotypy (different individuals expressing the same emotion differently under same contexts) makes it difficult to obtain universal patterns of emotions across individuals (Lacey & Lacey1958). For all the participants who took part in Phase I (see description in Section 4), it was observed that their physiological responses were highly correlated with their underlying affective stated. Any feature (derived from physiological signals) with an absolute correlation greater than equal to 0.3 with a given affective state was considered significant and was selected as inputs of the classifiers. In order to overcome person-stereotypy we adopted an individual-specific framework where we develop a model for each individual (e.g., we determine the physiological pattern of anxiety for each participant) in a Phase I study. In this work, regression tree methodology was employed to create a unique regression tree for each individual where the splitting threshold at each branch was specific to an individual's physiological response to the affective stressors. Later the same participants are invited to participate in a Phase II study where we verify the model developed in Phase I and evaluate the effectiveness of the affective feedback in the human-robot interaction task.

In our earlier work, we compared several machine learning algorithms for affect recognition (Rani, Liu, Sarkar & Vanman 2006) and found that regression tree technique was most efficient. In this work, we use regression tree for anxiety detection. Determining a person's probable anxiety level from his/her physiological response resembles a classification problem where the attributes are physiological features and the target function is anxiety level. The chief challenge in doing this was the complex nature of input physiological data sets. This complexity was primarily due to (i) high dimensionality of the input feature space (there are currently 46 features), (ii) mixture of data types, and (iii) nonstandard data structures. Additionally, a few physiological data sets were noisy where the biofeedback sensors had picked up movement artifacts. These data sets had to be discarded, resulting in the missing attributes. Tree structured classification (Breiman, Friedman, Olshen & Stone 1984) and regression techniques (also referred to as "decision tree techniques") have been frequently used to handle problem domains with the above-mentioned data complexity and discrepancies. Classification And Regression Trees (CARTs) have been extensively applied in the medical field (Kokol, Mernik, Završnik, Kancler & Malčic 1994). Important applications include diagnosing heart attacks, cancer diagnosis, and classification of age by gait measurement. Since the affect detection task resembles such classification problems, regression tree methodology was employed to develop a classification model for anxiety detection.

While creating such regression trees, two primary issues exist: (i) Choosing the best attribute to split the examples at each stage, and (ii) Avoiding data over fitting. Many different criteria could be defined for selecting the best split at each node. In this work, Gini Index function was used to evaluate the goodness of all the possible split points along all the attributes (Breiman, Friedman, Olshen & Stone 1984). Trees were pruned based on an optimal pruning scheme that first pruned branches that gave the least improvement in error cost. Pruning was performed to remove redundant nodes as bigger, overfitted trees have higher misclassification rates.

4. Phase I: Tasks for Model Building

In order to obtain physiological data to build models, 15 participants were presented with two computer tasks that elicited various affective states. These two tasks consisted of an anagram solving task and a Pong playing task. The anagram-solving task has been previously employed to explore relationships between both electrodermal and cardiovascular activity with mental anxiety (Pecchinenda, & Smith 1996). Emotional responses were manipulated in this task by presenting the participant with anagrams of varying difficulty levels, as established through pilot work. The Pong session consisted of a series of epochs each lasting up to 4 minutes, in which the participant played a variant of the early, classic video game “Pong”. This game has also been used in the past by researchers to study anxiety, performance, and gender differences Brown, Hall, Holtzer & Brown 1997). Various parameters of the game were manipulated to elicit required affective responses. These included ball speed and size, paddle speed and size, sluggish or over-responsive keyboard and random keyboard response.

Each session was subdivided into a series of discrete epochs that were bounded by self-reported affective state assessments. During the assessment, participants reported their perceived subjective emotional states. This information was collected using a battery of five self-report questions rated on a ten-point Likert scale. These questions inquired about the level of engagement, anxiety, anger, frustration and challenge perceived by the participant after each epoch. These psychological states play an important role in human-machine interaction. The affective states identified above were mainly chosen from the domain of negative affective states since they can be more closely related to performance and mental health of humans while working with machines. Discussion with Psychologists, review of research works done in psychophysiology and human factors, and preliminary piloting was instrumental in this phase of work. The important role played by these five affective states in human-robot interaction is discussed in (Rani 2005). Self-reports were used as reference points to link the objective physiological data to participants' subjective affective state. These assessments occurred every 3 minutes for anagram-solving and every 2-4 minutes for pong-playing. The participants reported their affective state on a scale of 0-9 where 0 indicated the lowest level and 9 indicated the maximum level.

Each participant took part in six sessions of the above two tasks – three one hour sessions of solving anagrams and three one hour sessions of playing Pong. These sessions spanned a period of one month. At the beginning of each session, baseline physiological signals were recorded in order to offset day-variability. At the end of Phase I study, we developed models for each participant that would predict a probable affective state (e.g., anxiety) based on their physiological markers.

The training datasets were formed by merging physiological data and self-reports of the participants. Figure 1 shows the procedure for merging the data to form an input-output set. The physiological data and self-reports were recorded continuously in two separate files at the time of acquisition. Later, the physiology data file was processed to extract the data pertaining to every epoch in a separate file. Then, each epoch file was again processed to extract the relevant features from the physiology data. Now, there was a set of physiological features pertaining to every epoch (say a vector of length p) and a vector of self-reported emotions pertaining to the same epoch (say a vector of length q).

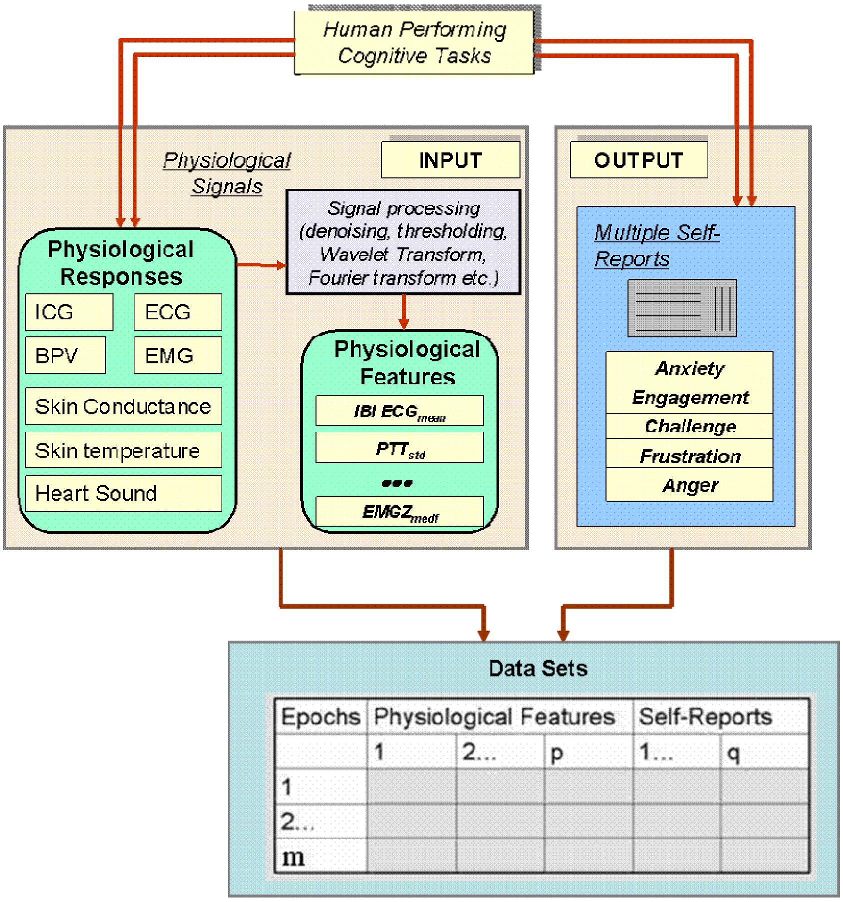


Figure 1. Formation of Data Input and Output Sets

The data-set formed by merging these two separate sets of information resembles a mxn matrix, where the first p columns represent the physiological features and the last q columns represent the self-reports ($n= p+q$). Each row corresponds to a separate epoch (m such epochs). The self-reports were normalized to [0, 1] and then discretized such that 0–0.33 was labeled low, 0.34–0.67 was medium and 0.68–1.0 was labeled high.

Four machine learning techniques – K Nearest neighbor (KNN), Regression Tree (RT), Bayesian Networks (BNT), and Support Vector Machines (SVM) were used to perform affect-recognition from the data sets that were formed above. The objective was: given a set physiological features, each labeled as an indicator of a particular level of arousal of a given affective state, determine the performance of the four learning techniques in predicting the class of unseen instances. It was found that SVM with a classification accuracy of 85.8% performed the best, closely followed by RT (83.5%), KNN (75.1%) and BNT (74.0%). Using informative features (the ones that were highly correlated with the affective states) improved the performance for KNN and BNT by almost 4%. In terms of space and time efficiency, RT ranked higher than the other three methods, hence was a natural choice as the machine learning technique used in this work.

5. Phase II: Closed Loop Human Robot Interaction

In this section, we describe in detail the various aspects of the human-robot interaction experiment that was designed to evaluate the effect of the implicit affective communication.

5.1 Participants and Basketball Task

5.1.1 Participants

Fourteen individuals (8 females, and 6 males) volunteered to participate in the experiments. Out of these, nine of them had also participated in the Phase I experiments. Their age ranged from 18 to 54 years of age. They were from diverse professional and ethnic backgrounds. Ideally, we would have liked to have all the participants from Phase I study participate in the Phase II experiments. However, some of the participants were not available during the Phase II study. Due to the nature of the tasks, the following were considered when choosing the participants: (i) fluency in English (to follow task instructions), and (ii) general health (the absence of any problem in hearing or sensing, and being able to use their arms and hands). 12 of them were right-handed and 2 of them were left-handed. Participants were solicited through phone, emails and flyers posted around the Vanderbilt University area. They were given monetary compensation for their voluntary participation.

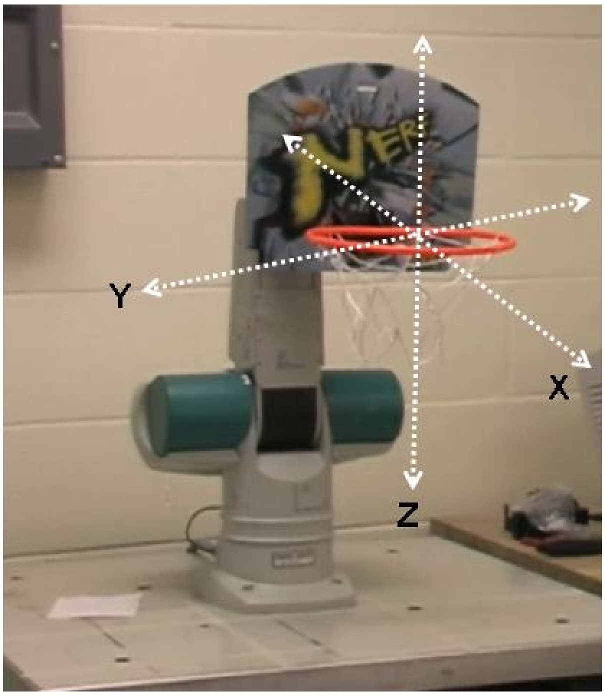


Figure 2. X, Y and Z direction motions of robot arm

5.1.2 Basketball Task

We designed a human-robot interaction task that could be used to evaluate the effect of implicit affective communication. We named this task as "robot-based basketball (RBB) task". In the RBB task, a basketball hoop was attached to a robotic manipulator that could move the hoop in different directions with different speeds. The manipulator is shown in Figure 2. There were three directions of motion for the robotic manipulator: X, Y, and Z directions. There were three speeds of motion – 100cm/sec, 50 cm/sec and 25 cm/sec. The player was required to shoot a required number of baskets into the moving hoop within a given time. The difficulty of the task could be varied by controlling parameters such as the manipulator speed and direction of motion. Based on a pilot study, different configurations of speed and motion were determined to vary the game difficulty. The three speeds of the robotic manipulator and the three directions of motion were permuted to get nine configurations. During piloting, participants were made to play fixed intervals of each of these configurations. After each epoch, they reported on the difficulty of the game as perceived by them (on a 10 point Likert scale). The numbers of baskets that they attempted, successfully made, and missed were also recorded. After the piloting was over, these results were compiled to determine the perceived difficulty level of each configuration. Then the configurations were sorted and grouped according to their difficulty ratings. It was found that there were three distinct clusters of configurations that were well separated along the difficulty scale. These clusters were named Levels I, II and III, in the increasing order of difficulty. The average number of baskets made by the participants for a given configuration was used to determine the threshold for that configuration. The threshold was 10% higher than the average of all configurations for a given level.

The game difficulty could be varied as a function of the player's affective states (in this case, anxiety). The main idea was two-fold: (i) to adapt the game difficulty in response to a player’s anxiety level such that each player could play the game at a low anxiety level; and (ii) to observe the effects of such implicit human-robot interaction aimed at reducing the player's anxiety and improving the player's performance. The relative difficulties of various trial configurations were established through pilot work. During the sessions, the participant's physiology was monitored with the help of wearable biofeedback sensors. After each epoch, (1.5 minutes duration) the participants reported their perceived subjective emotional states.

5.2 System Development

In order to develop the RBB task for real-time implementation we spent significant efforts in both hardware and software development. The system developed for the robot-based basketball game is shown in Figure 3.

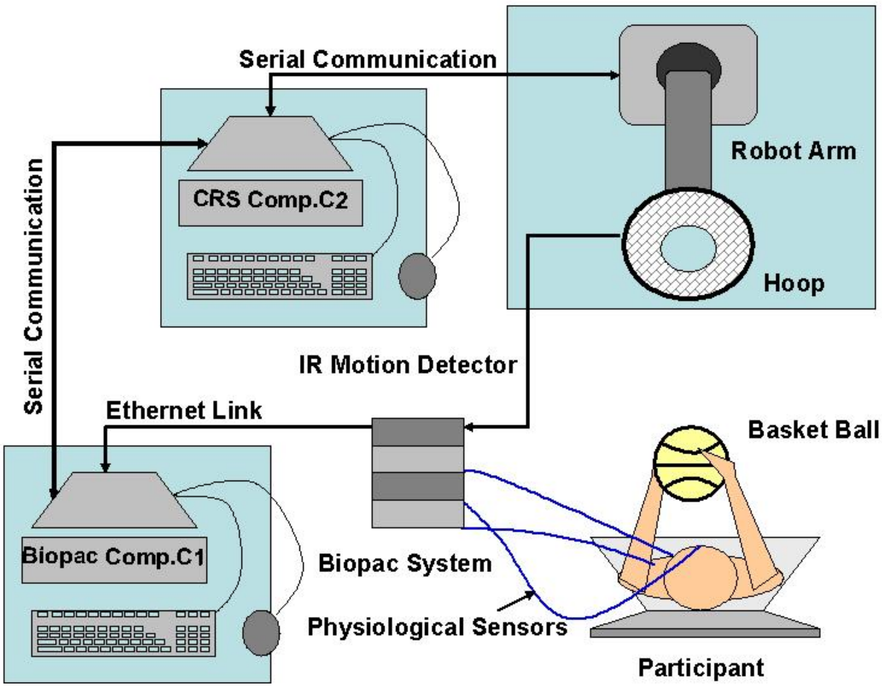


Figure 3. Experimental Set-up for Robot Basketball

The set-up included a 5 degrees-of-freedom robot manipulator (CRS Catalyst-5

System http://www.quanser.com) with a basketball hoop attached to its end-effector. Two sets of infrared (IR) transmitter and receiver pairs were attached to the hoop to detect balls going through the hoop. The set-up also included the biological feedback equipment (Biopac system http://www.biopac.com) that collected the physiological signals of the participant and the digital out from the IR sensors. The Biopac system was connected to a PC (C1) that: (i) acquired physiological signals from the medical equipment; (ii) acquired IR data through the analog input channels of the Biopac system; and (iii) ran signal processing and pattern recognition algorithms to process the physiological data to extract features and recognize anxiety level, and (iv) ran a program to fuse the affective information or the performance measure with current epoch configuration to determine the next epoch configuration.

C1 was connected serially to the CRS computer (C2) running Simulink. The task related triggers were transmitted from C1 to C2 via a RS232 link. The robot commands to control the various joints were transmitted from C2 to the robot. There was a communication protocol established between C1 and C2 that ensured that begin/end of the basketball epochs was appropriately synchronized with the physiological data acquisition on C1.

5.3 Experimental Design

Out of the 14 participants, 9 had taken part in Phase I experiments. As a result, the models to predict their probable affective states were already built based on the Phase I data. In the task design of robot-based basketball, adequate measures were taken to avoid physical effort from overwhelming the physiological response. The physiological sensors were placed on the non-dominant hand and side of the body of the participant. Test trials were conducted to check if the models build using training data from anagram solving task and a Pong playing task could be used to interpret and classify data from the basketball game. It was found that this could be done successfully. Similar conclusions were reached by Leon et al in (Leon, Clarke, Callaghan & Sepulveda 2007), where it was shown that physiological data collected from individuals with variable affect intensity or experiencing variable physical exertion could be successfully used to classify positive and negative affective states. The models for the other 5 participants had to be developed in Phase II experiments as described later. Each of these 14 participants took part in two robot basketball sessions (BB1 and BB2). In BB1 the robot changed the difficulty of the game based on performance without any regard to the anxiety level of the participant. In BB2, the game difficulty was changed based on the anxiety level of the participant without regard to the performance. Each basketball session was approximately 35 minutes long and consisted of 10 epochs of 1.5 minutes each. The remaining time was spent in setting-up, attaching sensors, self-reporting and taking breaks. During any given basketball epoch, the procedure was as follows:

- The participant was notified of the goal (number of baskets to be made and the time available) via speech on C1
- A start/stop command was played to instruct the player how to start and stop the game via speech on C1
- Once the epoch started, the participant was given feedback every 30 seconds regarding the number of baskets remaining and the time available via speech on C1

- After each epoch was over, the participant's performance was rated as excellent, above average or below average and informed to the participants via speech on C1
- This was followed by an interval of self-reporting. The self-reporting lasted for 30 seconds to 1 minute. After the self-reporting was completed, the next epoch would begin.

Three levels of difficulty were designed - Level I (easy), Level II (moderately difficult) and Level III (very difficult) - based on pilot study. Furthermore, three levels of performance, poor, good and excellent, were identified as well as three levels of anxiety were defined – low, medium and high. Figure 4 and Figure 5 show the stateflow models that were utilized to switch difficulty based on performance (P) and anxiety (A), respectively. It can be seen that when the switching between different difficulty levels was based on performance, excellent performance resulted in increase in the level of difficulty (except when the player was already at the highest level), good performance caused the level to remain constant at the current level, and poor performance resulted in decrease in difficulty level (except when the player was already at the lowest level). Similarly, it can be seen that when the switching between different difficulty levels was based on anxiety, low anxiety resulted in increase in the level of difficulty (except when the player was already at the highest level), medium anxiety caused the level to remain constant at the current level, and high anxiety resulted in decrease in difficulty level (except when the player was already at the lowest level).

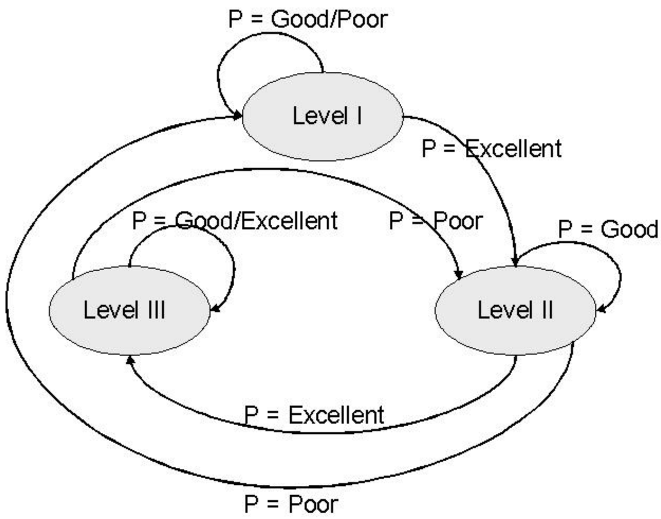


Figure 4. Stateflow Diagram for Performance-Based Modification of Game Difficulty

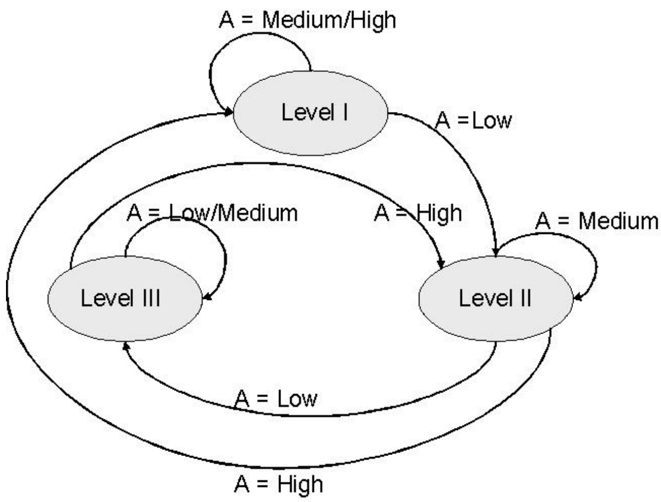


Figure 5. Stateflow Diagram for Anxiety-Based Modification of Game Difficulty

The following conditions were imposed to avoid bias in data due to habituation, session-order, and to deal with day-variability:

- In order to prevent habituation at least 10 days time interval between any two BB sessions was enforced.

- The sessions (performance-based and anxiety-based), were randomized to avoid any bias due to the order of sessions.
- All the other experimental conditions were kept constant over all sessions

In addition to these 9 participants from Phase I study, there were 5 new participants who engaged in six sessions each. Data from the first four sessions was utilized to build the affect-prediction model and the last two sessions were anxiety-based and performance-based sessions for these new participants. Each participant took part in four basketball sessions during which training data was collected. Each session was composed of 10 individual epochs of varying difficulty followed by an interval of self-reporting. The procedure was similar to the previous anagram and pong sessions (as described in Section 4). The participants' physiological data was linked with their self-reports and these datasets were utilized to build individual-specific affect-prediction models.

The Institutional Review Board (IRB) approval was sought and received for conducting these experiments. In the IRB application, all details of the experiment were reported and it was emphasized that the health and safety of the participants was by no means endangered by participating in these experiments. It was also mentioned that the maximum anxiety that the participants could experience was no greater than what they could experience while playing a difficult video game. A detailed consent form was also drafted that acquainted the participants with the experimental procedure and their role in it. They were also briefed by the experimenter about the physical as well as mental aspects of the experiment. Participants were allowed to participate in the experiment only after their consent had been obtained through a signed consent form.

5.4 Experimental Procedure

On arrival, the participant was taken to the experiment room where he/she would be seated in front of the robot where the tasks were presented. During the first session, the participants were told about the sensors – their purpose, the method of attaching sensors (non-invasive and unobtrusive) and their safety. Then, the sensors were attached to the participant's body. The signals being monitored included – electrocardiogram, phonocardiogram, bio-impedance, photoplethysmogram, skin conductance, peripheral temperature, and electromyogram (Corrugator Supercilii, Zygomaticus Major and upper Trapezius).

After the sensors were placed, the participants were given some leisure reading material and asked to relax for 5 minutes while their baseline reading was taken. After the baseline recording was completed, the experimenter started the session after briefing the participant regarding the rules of the task and general procedure. Once the session started, the participant was not disturbed for the rest of the session. After every epoch, the participant reported regarding his/her assessment of one's own emotion on a ten point Likert scale. There were a total of 5 questions, each of which inquired about the level engagement, anxiety, anger, frustration and challenge. At the end of each session, the participant answered questions pertaining to their overall experience during the entire session. This included their overall enjoyment, challenge and self-evaluation of their performance. These questions were asked to determine the aggregate player experience at the end of each session. These self-reports were used to compute the overall player satisfaction after playing for ten or so epochs. This helped us in determining if any one strategy scored over another in giving higher satisfaction to the player. After the session was over, the experimenter would unhook the sensors from the participant.

6. Results and Discussion

Fourteen participants took part in the robot basketball task. The results presented here are based on the validation sessions BB1 (performance-based) and BB2 (anxiety-based).

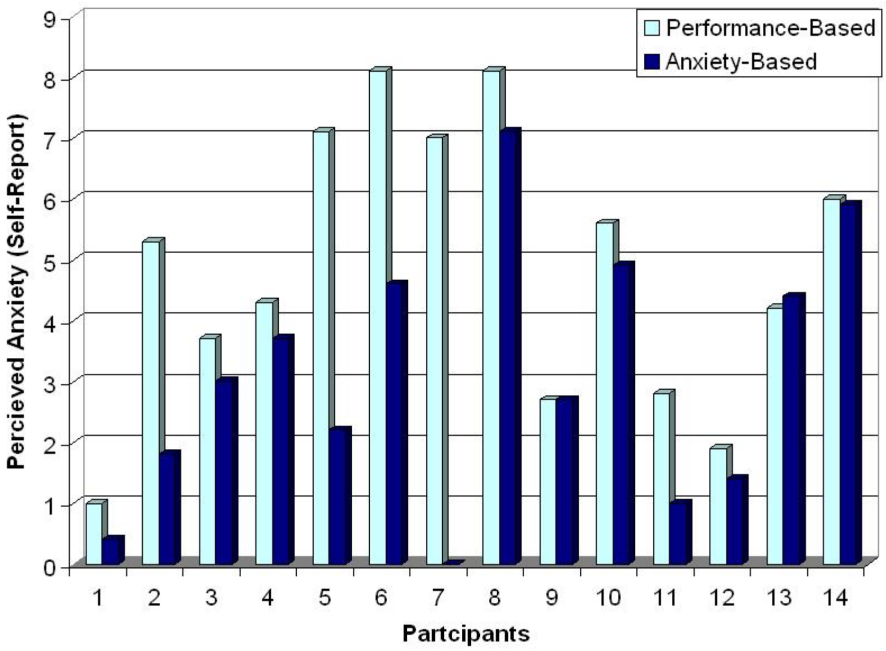


Figure 6. Subjective Anxiety as Reported by Participants

First, we present results to demonstrate that active monitoring of participants' anxiety and dynamic reconfiguration of epoch parameters allowed participant skill improvement while maintaining desired anxiety levels. Figure 6 shows the average anxiety of the participants as reported by them (perceived anxiety) during the two sessions. The lighter bars indicate the anxiety level during the performance-based sessions and the darker bars show the anxiety level during the anxiety-based sessions. It can be seen that out of 14 participants, 11 reported decrease in perceived anxiety, 1 reported an increase (P13) and 2 reported no change in anxiety (P9 and P14) during the anxiety-based session as compared to the performance-based session. The anxiety level was reported by the participants on a nine point Likert scale (0-9, where 0 indicated no anxiety and 9 indicated extremely high anxiety). This was a significant result as the anxiety-based sessions utilized the information regarding the probable anxiety level of the participant to continuously adapt the task difficulty to keep the participant in a lower anxiety state. The majority of the participants felt that they were less anxious when playing the anxiety-based game. In order to understand whether the difference in reported anxiety between the performance-based and the anxiety-based sessions are statistically significant we tested the null hypothesis that there was no change in anxiety between the performance-based and the anxiety-based sessions. Using Chi-square test, it was observed that the null hypothesis could be rejected at 99.99% confidence interval.

Anxiety was also calculated based on physiology for each epoch for each participant. The average real-time predictive accuracy across all the 14 participants was approximately 70%. Predictive accuracy represents how closely physiology-based quantitative measure of anxiety matched with that of the subjective rating of anxiety made by the participants.

Now we present results that show the performance of the participants in BB1 and BB2. Figure 7 shows the difference in performance between the performance-based and affect-based sessions. The lighter bars indicate performance in performance-based sessions and darker bars indicate performance in anxiety-based sessions. Performance during any given session was determined by computing the average performance across all the epochs in that session. For example if there were n epochs in a session, and the performance in the ith epoch was perfi, then the performance (P) in that session would be:

$$P = \frac{1}{n}\left(\sum_{i=1}^{n} perf_i\right) \tag{1}$$

Performance in a given epoch (*perfi*) was computed with respect to a threshold value that was determined during the pilot study. For instance, if in a given epoch i, a player playing at level k ($k \in \{1,2,3\}$), makes Bi number of baskets, when the threshold for that level is Tk, the performance of that player in that epoch i would be computed as shown below.

$$perf_i = \left(\frac{B_i - T_k}{T_k}\right) * 100 \tag{2}$$

Threshold level was different for each level and was determined by the pilot study conducted. The value of threshold for a given difficulty level was function of that particular difficulty level. Hence, a participant scoring *B* baskets in Level III would register a higher performance than when scoring *B* baskets in Level I. The aim was to allow participants to improve their skill and overall performance level while avoiding playing under high stress or anxiety. To avoid rushing the participants through the levels, the participants were made to start playing at lower levels and encouraged to steadily climb the difficulty level by playing better.

As seen in Figure 7, it was observed that out of fourteen participants, nine showed better performance during the affect-based session, while four had degradation in performance (P1, P8, P10, and P13) and one did not show any improvement (P14) during the anxiety-based session. In order to determine whether this change in performance was statistically significant we performed Chi-square test on the null hypothesis that there was no change in performance between performance-based and anxiety-based sessions. It was found that the null hypothesis could be rejected at the 99.99% confidence level.

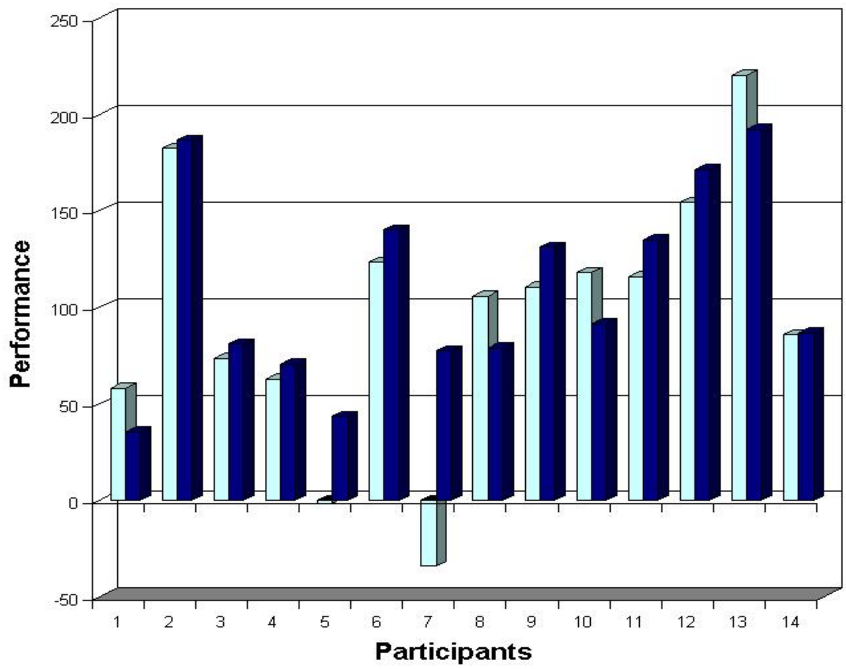


Figure 7. Difference in Performance between the Performance-Based and Affect-Based Sessions

Out of the nine participants that showed better performance during affect-based session, five participants played at a similar or higher difficulty level during the affect-based session than in the performance-based session. In order to understand whether the performance improvement was statistically significant, we performed Wilcoxon rank sum test on the two performance vectors – one from the performance-based sessions and the other from anxiety-based sessions. This test performs a two-sided rank sum test of the hypothesis that two independent samples come from distributions with equal medians. It was found that the null hypothesis could be rejected at the 99.99% confidence level. The remaining four participants showed better performance while playing at a lower difficulty level during the affect-based sessions. Of the four participants that showed lower performances in anxiety-based sessions, three played at higher difficulty level and one played at a lower difficulty level. The one participant who showed no change in performance played at a lower level of difficulty level during the anxiety-based session than in the performance-based session.

At the end of each session, the participants reported the level of challenge that they had experienced. It was seen that participants did not necessarily perceive the anxiety-based session to be more challenging than the performance-based one. Figure 8 shows challenge as reported by participants after performance-based and anxiety-based sessions. Six participants reported an increase in challenge. Eight participants reported either a decrease or constant level of challenge across both the sessions. The exact percentages of participants reporting increased challenge was 43%, decreased challenge was 28% and constant challenge was 29%.

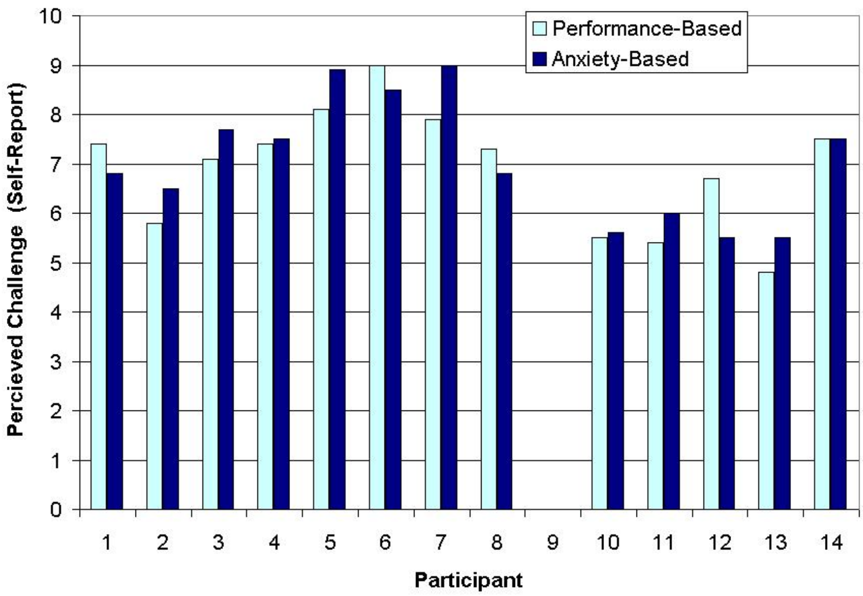


Figure 8. Subjective Challenge as Reported by Participants

An index called Satisfaction Index (*SI*) was defined by combining the values of challenge (*C*), enjoyment (*E*) and performance appraisal (*P*) reported by the participants at the end of each session.

$$SI = C + E + P \tag{3}$$

The SI could be a possible measure of the overall satisfaction of the participant during a given session controlled by either performance-based or anxiety-based mechanism. There have been many efforts in the past to develop metrics for measuring enjoyment in games, but no formal standards have been yet developed for evaluating fun, enjoyment, or satisfaction. Echoing similar opinion, Wiberg, in her technical report states "research into the aspect of user satisfaction has so far been neglected in the research discipline of HCI …When discussing fun and entertainment in the context of usability, the most closely related notion is 'user satisfaction'" (Wiberg 2005). In a work by Sweetser and Wyeth (2005), the authors present a model of enjoyment based on eight elements - concentration, challenge, skills, control, clear goals, feedback, immersion, and social interaction. They claim that each of these elements contributes to achieving enjoyment in games.

We used challenge and skill (as indicated by performance) along with a direct report on enjoyment to compute the Satisfaction Index. Figure 9 shows the values of the SI during the two sessions for each participant. It can be observed that 10 out of 14 participants reported an increase in the SI during the anxiety-based session. Four participants reported a decrease in SI (P8, P10, P12, and P13). It should be noted that out of the nine participants who showed better performance in anxiety-based session as opposed to performance-based session, eight participants also reported higher satisfaction index in the anxiety-based session.

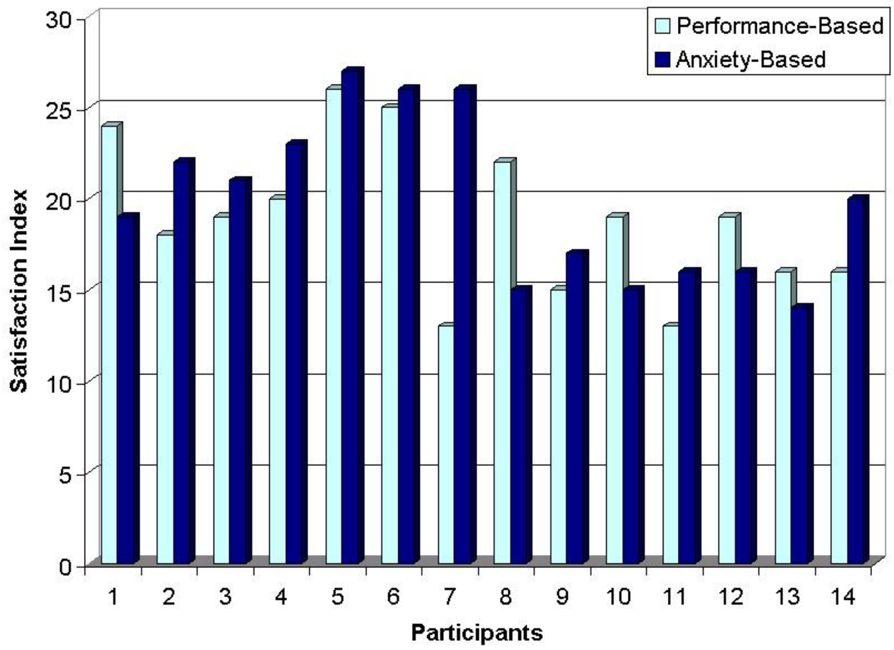


Figure 9. Satisfaction Index for all the Participants

On the other hand, out of the four participants who showed better performance in the performance-based session, three of them reported higher satisfaction in the performance-based session. This brings forth the possibility that the participants who performed better in performance-based session preferred performance-based task adaptation and experienced higher levels of challenge, enjoyment and confidence in their performance during these sessions than in the anxiety-based sessions. It was observed that the null hypothesis stating that there was no change in satisfaction index between performance-based and anxiety-based sessions could be rejected at 99% confidence interval using Chi-square test.

In summary, the following could be concluded from the robot-based basketball task:

- The anxiety-based sessions were successful in reducing the anxiety of the participants. 79% of the participants reported a decrease in anxiety, while 7% reported an increase in anxiety and 14% reported no change in anxiety level.
- Anxiety-based task modification also resulted in an improvement in performance for the majority of the participants. 64% of the participants showed an increase in performance. There was a degradation of performance for 29% of the participants and it remained constant for 7% of the participants.
- The average accuracy of the anxiety prediction system was 70% for the Robot-Based Basketball task. 42 % of the participants reported an increase in challenge in the anxiety-based session and 71% reported an increase in the overall satisfaction during the anxiety-based session.

7. Conclusion and Future Work

In this paper, we have demonstrated a novel approach to closed loop human-robot interaction based on implicit affective communication. The presented approach is based on physiology-based affect recognition. We have shown that it is possible for a robot to detect human anxiety in real-time as well as appropriately respond to it during an interaction task. We have designed a new human-robot interaction task, called robot-based basketball task, and developed an experimental system for its real-time implementation and verification. Experiments with 14 participants demonstrated that the robot could influence (lower) human anxiety for 11 of these participants during the course of task execution, which we believe is a significant result. In addition, we have found that such anxiety-based game led to higher improvement of performance for 9 of these participants. It suggests that the implicit affective communication facilitates this learning task when the participants interact with a robot in a closed-loop manner. Other results that are discussed in the paper also seem to indicate the benefit and importance of such communication in human-robot interaction.

Applications of physiology-based affective communication in human-robot interaction could be invaluable in providing implicit human information to the robot. While physiology can be integrated with other modalities of affect-recognition such as facial expressions, vocal intonation, gestures and postures, it could also be employed independently to give reliable affect information. Physiology has distinct advantages over other modalities, as it is involuntary and continuously available. While wearing multiple wired physiological sensors may be a limiting factor, it can be safely assumed that this limitation will soon vanish given the fast pace of advancement in wireless and sensor technology.

Future work would consist of expanding the range of tasks and contexts to which this framework can be applied to and increasing the reliability and robustness of affect recognition. We would also like to work towards increasing the range of affective states detected and discriminated beyond anxiety.

Acknowledgements

We gratefully acknowledge the help of Professor Eric Vanman of the Psychology department of Georgia State University during this work.

References

Backs, R. W., Lenneman, J. K., Wetzel, J. M., & Green, P. (2003). Cardiac measures of driver workload during simulated driving with and without visual occlusion. *Human Factors, 45*(4), 525-538.

Bartlett, M. S., Littlewort, G., Fasel, I., & Movellan, J. R. (2003). Real time face detection and expression recognition: Development and application to human-computer interaction, *Conference on Computer Vision and Pattern Recognition Workshop* (Vol. 5). Wisconsin, USA.

Bradley, M. M. (2000). Emotion and motivation. Handbook of Psychophysiology. In J. T. Cacioppo, L. G. Tassinary & Gary G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 602-642). New York: Cambridge University Press.

Breazeal, C., & Aryananda, L. (2002). Recognition of affective communicative intent in robot-directed speech. *Autonomous Robots, 12*(1), 83-104.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.

Brown, R. M., Hall, L. R., Holtzer, R., Brown, S. L., & Brown, N. L. (1997). Gender and video game performance. *Sex Roles, 36*(11-12), 793 – 812.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence, 16*(7-8), 555-575.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *Ieee Signal Processing Magazine, 18*(1), 32-80.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*(3-4), 143-166.

Graesser, A., C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R. (1999). AutoTutor: A simulation of a human tutor. *Journal of Cognitive Systems Research*, 35-51.

Hanajima, N., Ohta, Y., Hikita, H., & Yamashita, M. (2005). Investigation of impressions on approach motion of a mobile robot based on psychophysiological analysis, *IEEE International Workshop on Robot and Human Interactive Communication* (pp. 79-84). Nashville, USA.

Haritaoglu, I., Cozzi, A., Koons, D., Flickner, M., Yacoob, Y., Zotkin, D., & Duriswami, R. (2001). Attentive toys, *International Conference on Multimedia and Expo* (Vol. 22-25, pp. 917-920). Tokyo, Japan.

Hirata, K. (1990). The concept of computer anxiety and measurement of it. *Bulletin of Aichi University of Education, 39*, 203-212.

Hoffman G., & Breazeal C. (2004). Robots that Work in Collaboration with People, *AAAI Fall Symposium on the Intersection of Cognitive Science and Robotics: From*

Interfaces to Intelligence. Washington DC.

Iszo, L., Mischinger, G., & Lang. E. (1999). Validating a new method for ergonomic evaluation of human computer interfaces. *Periodica Polytechnica Ser. Soc. Man. Sci., 7*(2), 119-134.

Kanda, T., Ishiguro, H., Ono, T., Imai, M., & Nakatsu, R. (2002). Development and evaluation of an interactive humanoid robot "Robovie", *IEEE International Conference on Robotics and Automation* (pp. 1848-1855).

Kapoor, A., Mota, S., & Picard, R. (2001). Towards a Learning Companion that Recognizes Affect, *Proceedings of the AAAI Fall Symposium*, North Falmouth, USA.

Kapur, A., Kapur, A., Naznin, V., George, T., & Peter, D. (2005). Gesture-based affective computing on motion capture data, *International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 1-7). Beijing, China.

Kleinsmith, A., Fushimi, T., & Bianchi-Berthouze, N. (2005). An incremental and interactive affective posture recognition system, *UM 2005 Workshop: Adapting the Interaction Style to Affective Factors*. Edinburgh, UK.

Kleinsmith, A., Ravindra De Silva, P., & Bianchi-Berthouze, N. (2005). Recognizing emotion from postures: cross-cultural differences in user modeling, *User Modeling 2005* (pp. 50-59). Edinburgh, Scotland, UK.

Kokol, P., Mernik, M., Završnik, J., Kancler, K., & Malčic, I. (1994). Decision trees based on automatic learning and their use in cardiology. *Journal of Medical Systems, 9*(4), 201-206.

Kramer, A. F., Sirevaag, E. J., & Braune, R. (1987). A Psychophysiological Assessment of Operator Workload during Simulated Flight Missions. *Human Factors, 29*(2), 145-160.

Kulic, D., & Croft, E. (2003). Estimating Intent for Human-Robot Interaction. *Proceedings of International Conference on Advanced Robotics*, (pp. 810-815). Portugal.

Kulic, D., & Croft, E. (2003). Intent-Based Planning and Control for Human Robot Interaction. *Proceedings of International Conference on Advanced Robotics*. Portugal.

Lacey, J. I., & Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy. *The American journal of psychology, 71*(1), 50-73.

Lee, C. M., & Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing, 13*(2), 293-303.

Leon, E., Clarke, G., Callaghan, V., & Sepulveda, F. (2007). A user-independent real-time emotion recognition system for software agents in domestic environments. *Engineering Applications of Artificial Intelligence, 20*(3), 337-345.

Liu, C., Rani, P., & Sarkar, N. (2005). Comparison of Machine Learning Techniques for Affect Detection in Human Robot Interaction, *IEEE/RSJ International Conference on Intelligent Robots and Systems,* (pp. 2662-2667) Alberta, Canada.

Mehrabian A., & Friar, J. T. (1969). Encoding of attitude by a seated communicator via postures and position cues. *Journal of Consulting and Clinical Psychology*, 330-336.

Nasoz, F., Ozyer, O., Lisetti, C. L., & Finkelstein, N. (2002). Multimodal Affective Driver Interfaces for Future Cars, *Proceedings of the tenth ACM international conference on Multimedia,* (pp. 319-322). France.

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

Pecchinenda, A., & Smith, C. A. (1996). The affective significance of skin conductance activity during a difficult problem-solving task. *Cognition and Emotion, 10*(5), 481-504.

Picard, R. W. (1997). *Affective Computing*. Cambridge: The MIT Press.

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

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component. Human Factors, 45(4), 601-613.

Rani, P. (2005). *Psychophysiology based Affective Communication for Implicit human-Robot Interaction.* Doctoral Dissertation, Vanderbilt University.

Rani, P., Liu, C. C., Sarkar, N., & Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis and Applications, 9*(1), 58-69.

Rani, P., Sarkar, N., Smith, C. A., & Kirby, L. D. (2004). Anxiety detecting robotic system - towards implicit human-robot collaboration. *Robotica, 22*, 85-95.

Reeves, B., & Nass, C. I. (1996). *The media equation : how people treat computers, televisions, and new media as real people and places*. New York: Cambridge University Press.

Rohrmann, S., Hennig, J., & Netter, P. (1999). Changing psychobiological stress reactions by manipulating cognitive processes. *International Journal of Psychophysiology, 33*(2), 149-161.

Sweetser, P., & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *ACM Computers in Entertainment, 3*(3), 1-24.

Vicente, K. J., Thornton, D. C., & Moray, N. (1987). Spectral-Analysis of Sinus Arrhythmia - a Measure of Mental Effort. *Human Factors, 29*(2), 171-182.

Wiberg, C. (2005). Affective computing vs. usability? Insights of using traditional usability evaluation methods, *Computer Human Interaction, Workshop on Innovative Approaches to Evaluating Affective Interfaces*.

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple Psychophysiological measures. *The International Journal of Aviation Psychology, 12*(1), 3-18.

CHAPTER V: MANUSCRIPT 4

PHYSIOLOGY-BASED AFFECT RECOGNITION FOR COMPUTER ASSISTED INTERVENTION OF CHILDREN WITH AUTISM SPECTRUM DISORDER

Changchun Liu, Karla Conn, Nilanjan Sarkar, Wendy Stone

(Published by International Journal of Human-Computer Studies)

Abstract

Generally, an experienced therapist continuously monitors the affective cues of the children with Autism Spectrum Disorders (ASD) and adjusts the course of the intervention accordingly. In this work, we address the problem of how to make the computer-based ASD intervention tools affect-sensitive by designing therapist-like affective models of the children with ASD based on their physiological responses. Two computer-based cognitive tasks are designed to elicit the affective states of liking, anxiety, and engagement that are considered important in autism intervention. A large set of physiological indices are investigated that may correlate with the above affective states of children with ASD. In order to have reliable reference points to link the physiological data to the affective states, the subjective reports of the affective states from a therapist, a parent, and the child himself/herself were collected and analyzed. A Support Vector Machines (SVM) based affective model yields reliable prediction with approximately 82.9% success when using the therapist's reports. This is the first time, to our knowledge, that the affective states of children with ASD have been experimentally detected via physiology-based affect recognition technique.

Key words: Human-computer Interaction, Autism Intervention, Physiological Sensing, Support Vector Machines, Affect Recognition.

1. Introduction

Autism is a neurodevelopmental disorder characterized by core deficits in social interaction, social communication, and imagination. These characteristics often vary significantly in combination and severity, within and across individuals, as well as over time (DSM-IV-TR, American Psychiatric Association, 2000). Emerging research suggests prevalence rates as high as approximately 1 in 150 for the broad autism spectrum (CDC, 2007). While, at present, there is no single universally accepted intervention, treatment, or known cure for Autism Spectrum Disorders (ASD) (NRC, 2001; Sherer and Schreibman, 2005), there is an increasing consensus that intensive behavioral and educational intervention programs can significantly improve long term outcomes for individuals and their families (Cohen et al., 2006; Rogers, 1998). Despite the urgent need and societal import of intensive treatment (Rutter, 2006), appropriate intervention resources for children with ASD and their families are often difficult to access and highly expensive (Tarkan, 2002). Therefore, an important new direction for research on ASD is the identification and development of assistive therapeutic tools that can make application of intensive treatment more readily accessible.

In response to this need, a growing number of studies have been investigating the application of advanced interactive technologies in intervention of children with ASD, namely computer technology, virtual reality (VR) environments, and robotic systems. It has been shown that computer and VR based intervention may provide a simplified but exploratory interaction environment for children with ASD (Moore et al., 2000; Parsons and Mitchell, 2002). Various software packages have been developed and applied to address specific deficits associated with autism, e.g., understanding of false belief (Swettenham, 1996), attention (Trepagnier et al., 2006), expression recognition (Silver and Oakes, 2001), and social communication (Bernard-Opitz et al., 2001; Parsons et al., 2005). Different from using computer software or VR environments, the interaction between children with ASD and physical robots during the intervention contributes important real-time and embodied characteristics of face-to-face social interaction among humans. Robots have been used to teach basic social interaction skills using turn-taking and imitation games, and the use of robots as social mediators and as objects of shared attention can encourage interaction with peers and adults (Dautenhahn and Werry, 2004; Kozima et al., 2005; Pioggia et al., 2005; Robins et al., 2004). In the rest of the paper, we will use the term computer to imply both computer and robot assisted ASD interventions.

Even though there is increasing research in interactive intervention, we found no published studies that specifically addressed the automatic detection of affective cues exhibited by children with ASD. This gap could be important since research suggests that people tend to interact with computers as they might relate to other people, provided that the technology behaves in a socially competent manner (Reeves and Nass, 1996). Human interactions are characterized by explicit as well as implicit channels of communication. While the explicit channel transmits overt messages, the implicit one transmits hidden messages about the communicator (e.g., his/her intention and attitude). There is a growing consensus that endowing a computer with an ability to understand implicit affective cues should permit more meaningful and natural human-computer interaction (Picard, 1997). Several computer-based automated tutoring systems have been successfully developed that assess and utilize the affective information of typical people

(Conati, 2002; Prendinger et al., 2005). It is common in autism therapy that therapists who work with children with ASD continuously monitor the children's various affective information or cues in order to adapt their intervention strategies. For example, a “likes and dislikes chart” is recommended during interventions to record the children’s preferred activities and/or sensory stimuli that could be used later as reinforcers and/or “alternative behavior” (Seip, 1996). Children with autism are particularly vulnerable to anxiety and intolerance of feelings of frustration, which requires a therapist to plan tasks at an appropriate level of difficulty (Ernsperger, 2003). The engagement of children with ASD is the ground basis for the “floor-time therapy” to help them develop relationships and social/communication skills (Wieder and Greenspan, 2005). The existing computer assisted therapeutic tools for ASD do not possess the ability of deciphering affective information of the children. We believe that such ability could be critical, given that the affective factors of children with ASD have significant impacts on the intervention practice.

A computer that can detect the affective states of a child with ASD and interact with him/her based on such perception could have a wide range of potential impacts. Interesting activities likely to retain the child's attention could be chosen when a low level of engagement is detected. Complex social stimuli, sophisticated interactions, and unpredictable situations could be gradually, but automatically, introduced when the computer has the knowledge that the child is comfortable or not anxious at a certain level of interaction dynamics for a reasonably long period of time. A therapist could use the history of the child’s affective information to analyze the effects of the therapeutic approach. With the record of the activities and the consequent emotional changes in a child, a computer could learn individual preferences and affective characteristics over time and thus could alter the manner in which it responds to the needs of different children.

The relative paucity of research that addresses affect detection of children with ASD, combined with the import of affect-sensitive intervention assistive systems, underscores the need for empirical studies in this area. The primary objective of the present research is to develop affective models for children with ASD via a physiology-based affect recognition technique while they interact with a computer. In order to achieve this objective, we divide the research into several components: (i) to design computer-based cognitive tasks for affect elicitation; (ii) to derive physiological features via signal processing; (iii) to investigate multiple subjective reports; and (iv) to develop affective models by using machine learning techniques.

The rest of the paper is organized as follows: The scope and research rationale of this paper is presented in Section 2. Section 3 describes the physiological indices used for affect recognition. In Section 4, we describe the computer tasks designed for affect-elicitation and the experimental setup. This part is followed by a section with the detailed results and discussions (Section 5). Finally, Section 6 summarizes the contributions of the paper and outlines the future directions of this research. In addition, the machine learning algorithm employed in this study is presented in the Appendix.

2. Scope and Rationale

There are several modalities such as facial expression (Bartlett et al., 2003), vocal intonation (Lee and Narayanan, 2005), gestures and postures (Asha et al., 2005; Kleinsmith et al., 2005), and physiology (Kulic and Croft, 2007; Liu et al., 2006;

Mandryk and Atkins, 2007; Nasoz et al., 2003; Picard et al., 2001; Rani et al., 2004) that can be utilized to evaluate the affective states of individuals interacting with a computer. In this work we chose to create affective models based on physiological data for several reasons. Children with ASD often have communicative impairments (both nonverbal and verbal), particularly regarding expression of affective states (DSM-IV-TR, American Psychiatric Association, 2000; Green et al., 2002; Schultz, 2005). While these vulnerabilities place limits on traditional conversational and observational methodologies, physiological signals are continuously available and are arguably not directly impacted by these difficulties (Ben Shalom et al., 2006; Groden et al., 2005; Toichi and Kamio, 2003). As such, physiological modeling may represent a methodology for gathering rich data despite potential communicative impairments. In addition, physiological data may offer an avenue for recognizing aspects of affect that may be less obvious for humans but more suitable for computers. Specifically, a computer system may be able to quickly implement signal processing and pattern recognition tools to infer underlying affective states that a human could not. Furthermore, there is evidence that the transition from one affective state to another is accompanied by dynamic shifts in indicators of Autonomic Nervous System (ANS) activity (Bradley, 2000). The physiological signals that have been used in this research consist of various cardiovascular, electrodermal, electromyographic, and body temperature signals, all of which have been extensively investigated in the psychophysiology literature (Bradley, 2000). Several researchers in the human–machine interaction community have focused on physiology-based affect-recognition for typical adults. Picard et al. (2001) have employed a combination of Sequential Floating Forward Search and Fisher Projection methods to classify eight emotions with 81% accuracy. K-Nearest Neighbors (KNN), Discriminant Function Analysis, and Marquardt Backpropagation algorithms were applied to differentiate among six emotions by Nasoz et al. (2003), and the correct classification accuracies – 71%, 74%, and 83%, respectively – were achieved for the three algorithms. In our earlier work (Rani et al., 2006), we compared several machine learning algorithms: namely, KNN, Bayesian Network, Support Vector Machines (SVM), and Regression Tree for determining the intensity of the affective states, and the best prediction accuracy rate 85.8% was achieved using SVM.

An important question when estimating human affective response is how to operationalize the affective state. Although much existing research categorizes physiological signal data into "basic emotions," there is no agreement on a set of basic emotions (Cowie et al., 2001). This fact implies that it requires pragmatic choices to select target affective states to be recognized (Cowie et al., 2001). In this research we chose anxiety, engagement, and liking as the target affective states. Anxiety was chosen for two primary reasons. First, anxiety plays an important role in various human-machine interaction tasks that can be related to task performance (Brown et al., 1997). Second, anxiety frequently co-occurs with ASD and plays an important role in the behavior difficulties of children with autism (Gillott et al., 2001). Engagement, defined as "sustained attention to an activity or person" (NRC, 2001), has been regarded as one of the key factors for children with ASD to make substantial gains in academic, communication, and social domains (Ruble and Robson, 2006). With "playful" activities in the learning environments, the liking of the children (i.e., the enjoyment they experience when interacting with the computer) may create urges to explore and allow prolonged interaction for the children with ASD, who are susceptible to being withdrawn (Dautenhahn and Werry, 2004; Papert, 1993).

Furthermore there is evidence that the affective state could be an aggregate of various affective categories at different arousal levels (Vansteelandt et al., 2005), and within a given context, different individuals express the same emotion with different characteristic response patterns (i.e., phenomenon of person stereotypy) (Lacey and Lacey, 1958). The novelty of the presented affective model is that it is individual-specific in order to accommodate the phenomenon of person stereotypy and the spectrum nature of autism (DSM-IV-TR, American Psychiatric Association, 2000), and it consists of an array of recognizers, each of which determines the intensity (e.g., high/low level) of one target affective state for each individual. Even though physiology has been successfully employed to build affect recognizers for typical individuals in several research groups (Kulic and Croft, 2007; Mandryk and Atkins, 2007; Picard et al., 2001; Rani et al., 2006), the studies of the correlation of the physiological signals and the affective states of people with ASD are relatively few (Ben Shalom et al., 2006; Groden et al., 2005) and no quantitative modeling results (e.g., affective model with reliable prediction capability) have been reported. To our knowledge affect recognition for children with ASD by using a large set of physiological indices has not been researched.

The primary objective of this preliminary study is to investigate the feasibility of affective modeling for children with ASD via a physiology-based affect recognition technique. Currently, children with ASD are recommended to undergo at least 25 hours-per-week of year-round intensive autism intervention (i.e., one-on-one therapy with a trained therapist) outside of school and extracurricular activities (NRC, 2001; Tarkan, 2002). The developed affective model can be used in the computer-assisted autism interventions to detect the children's affective states on-line, move them toward the intervention goals in an affective manner, and make the treatment more accessible (e.g., possibly allowing intensive intervention to be conducted at home). This work included an autism therapist who has 5 years of experience in therapeutic and diagnostic interventions for children with ASD and each participant’s parent. The therapist and the parent observed the experiments (as described in section 4.3) and provided subjective reports based on their expertise/experience in inferring the presumable underlying affective states from the observable behaviors of a child with ASD. The therapist and the parent did not use the participant's physiological signals to recognize affective states, but these signals were recorded for eventual affective modeling (i.e., a mapping between the objective physiological signals and the subjective reports) as described in section 4.4. In this study, the therapist’s reports on perceived intensity of the affective states of a participating child and the extracted physiological indices (as described in section 3) are employed to build therapist-like affect recognizers. In autism interventions, a therapist continuously monitors the affective cues of children with ASD based on behavioral observations. In this work, the “therapist-like affect recognizers” were developed to emulate the therapist's affect-recognition capability, however, based on the children’s physiological signals. With the incorporation of the therapist’s reports, the recognizers will be capable of autonomously delivering similar assessments of the affective states of the children with ASD in real time even when the therapist is not available. Ultimately, integrating the affective models with current interactive intervention approaches may allow for automating the intensive, repetitive aspects of the existing behavioral therapy techniques and possibly steer the individual towards the intervention goal in an affect-sensitive manner.

3. Physiological Signal Acquisition and Indices

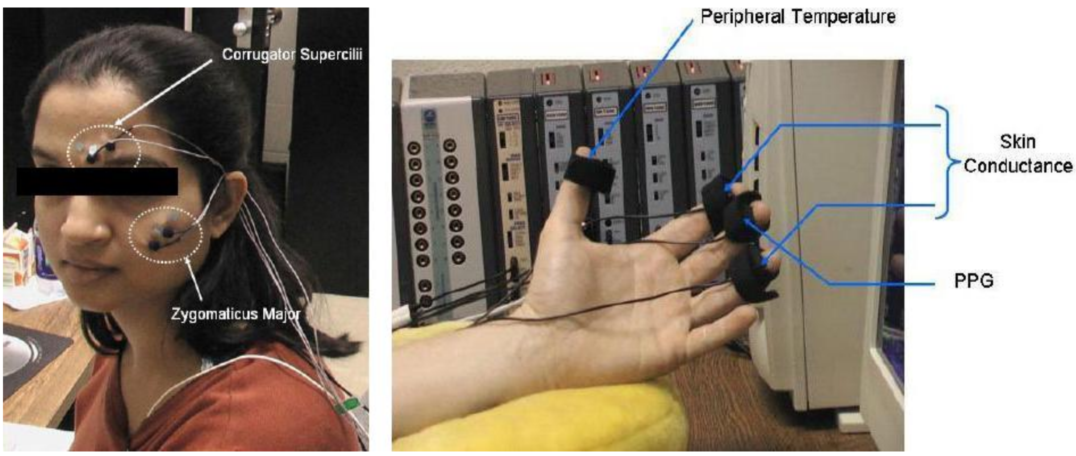
There is good evidence that the physiological activity associated with affective states can be differentiated and systematically organized (Bradley, 2000). The cardiovascular and electromyogram activities have been used to examine the positive and negative affective states of people (Cacioppo et al., 2000; Papillo and Shapiro, 1990). Electrodermal activities have been shown to be associated with task engagement (Pecchinenda and Smith, 1996). The variation of peripheral temperature due to emotional stimuli was studied by (Kataoka et al., 1998). In this work, we exploited the dependence of physiological responses on underlying affective states to develop affective models for children with ASD by using the machine learning method as described in section 4.4 and the Appendix. The physiological signals we examined were: various cardiovascular activities including electrocardiogram (ECG), impedance cardiogram (ICG), photoplethysmogram (PPG), and phonocardiogram (PCG)/heart sound; electrodermal activities (EDA) including tonic and phasic responses from skin conductance; electromyogram (EMG) activities from corrugator supercilii, zygomaticus major, and upper trapezius muscles; and peripheral temperature. Relevant features were derived from the physiological signals using various signal-processing techniques such as Fourier transform, wavelet transform, thresholding, and peak detection. The physiological signals that were examined in this work along with the features derived from each signal are described in Table 1.

Table 1. Physiological Indices

Physiological Response	Features Derived	Label Used	Unit of Measurement
Electrocardiogram	Sympathetic power (from ECG)	Sym	Unit/Square Second
	Parasympathetic power (from ECG)	Para	Unit/Square Second
	Very Low Frequency Power (from ECG)	VLF	Unit/Square Second
	Ratio of powers	Sym Para Para VLF Sym VLF	No unit
	Mean IBI	IBI ECGmean	Milliseconds
	Std. of IBI	IBI ECGstd	Standard Deviation (no unit)
Photoplethysmogram	Mean amplitude of the peak values of the PPG signal (Photoplethysmogram)	PPG Peakmean	Micro Volts
	Standard deviation (Std.) of the peak values of the PPG signal	PPG Peakstd	Standard Deviation (no unit)
	Mean Pulse Transit Time	PTTmean	Milliseconds
Heart Sound	Mean of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound	Mean d3 Mean d4 Mean d5	No unit
	Standard deviation of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Std d3 Std d4 Std d5	No unit
Bioimpedance	Mean Pre-Ejection Period	PEPmean	Milliseconds
	Mean IBI	IBI ICGmean	Milliseconds
Electrodermal activity	Mean tonic activity level	Tonicmean	Micro-Siemens
	Slope of tonic activity	Tonicslope	Micro-Siemens /Second
	Mean amplitude of skin conductance response (phasic activity)	Phasicmean	Micro-Siemens
	Maximum amplitude of skin conductance response (phasic activity)	Phasicmax	Micro-Siemens
	Rate of phasic activity	Phasicrate	Response peaks/Second
Electromyographic activity	Mean of Corrugator Supercilii activity	Cormean	Micro Volts
	Std. of Corrugator Supercilii activity	Corstd	Standard Deviation (no unit)
	Slope of Corrugator Supercilii activity	Corslope	Micro Volts/Second
	Mean Interbeat Interval of blink activity	IBI Blinkmean	Milliseconds
	Std. of Interbeat Interval of blink activity	IBI Blinkstd	Standard Deviation (no unit)
	Mean amplitude of blink activity	Amp Blinkmean	Micro Volts
	Standard deviation of blink activity	Blinkstd	Standard Deviation (no unit)
	Mean of Zygomaticus Major activity	Zygmean	Micro Volts
	Std. of Zygomaticus Major activity	Zygstd	Standard Deviation (no unit)
	Slope of Zygomaticus Major activity	Zygslope	Micro Volts/Second
	Mean of Upper Trapezius activity	Trapmean	Micro Volts
	Std. of Upper Trapezius activity	Trapstd	Standard Deviation (no unit)
	Slope of Upper Trapezius activity	Trapslope	Micro Volts/Second
	Mean and Median frequency of Corrugator, Zygomaticus, and Trapezius	Zfreqmean Cfreqmedian Tfreqmean	Hertz
Temperature	Mean temperature	Tempmean	Degree Centigrade
	Slope of temperature	Tempslope	Degree Centigrade/Second
	Std. of temperature	Tempstd	Standard Deviation (no unit)

3.1 Acquisition of Physiological Signals

The physiological signals were acquired using the Biopac MP150 physiological data acquisition system (www.biopac.com). ECG was measured from the chest using the standard two-electrode configuration. ICG describes the changes of thorax impedance due to cardiac contractility and was measured by four pairs of surface electrodes that were longitudinally configured on both sides of the body. The top pair of ICG electrodes was placed on the neck parallel to and about 3 cm above the second pair, located at the base of the neck; the bottom electrodes were placed parallel to and about 5 cm below the third ones, which were placed on the sides of the chest at the level of the xiphisternal junction. A microphone specially designed to detect heart sound waves was placed on the chest to measure PCG. PPG, peripheral temperature, and EDA were measured from the middle finger, the thumb, and the index and ring fingers of the non-dominant hand, respectively, using surface electrodes sewn in stretchy Velcro straps. EMG was measured by placing surface electrodes on two facial muscles (corrugator supercilii and zygomaticus major) and an upper back muscle (upper trapezius). All the physiological sensors were extensions of the Biopac physiological data acquisition system. The sampling rate was fixed at 1000 Hz for all the channels. Appropriate amplification and band-pass filtering were performed. Before each session, a three-minute baseline recording was done that was later used to offset day-variability. During the baseline recording, participants were asked to relax in a seated position and read age-appropriate leisure material. Subsequently, emotional stimuli induced by cognitive tasks were applied in epochs of up to four minutes in length (as described in section 4.2). Previous research (Pecchinenda and Smith, 1996; Rani et al., 2006) has shown that physiological signals (e.g. electrodermal activity, electromyographic activity, and cardiovascular activity) of 2-4 minutes in length were adequate for detecting affective states (e.g., anxiety, anger, engagement, etc.) from similar computer-based tasks. Each child with ASD took part in six one-hour sessions containing 13-15 epochs each. Each session took place on a different day to avoid bias in data due to habituation. Figure 1 shows the sensor setup.



(a) (b)

Figure 1. Sensor Setup. (a) shows the position of facial EMG sensors and (b) shows the placement of sensors on non-dominant hand

<u>3.2 Cardiovascular Activity</u>

ECG measures the heart activity through the electrical signal of the heart muscle. The number of beats per minute (bpm) is called the heart rate and is typically 70-80 bpm at rest. Inter beat interval (IBI) is the time interval in milliseconds between two "R" peaks in the ECG waveform. The R-peak detection algorithm performed band-pass filtering on the raw ECG signal and the signal was then smoothed by a 10 ms moving average window. Peaks were then detected in the resulting signal, and detection heuristic rules were applied to avoid missing R peaks or detecting multiple peaks for a single heart beat.

These rules included obtaining the amplitude threshold (the difference between a peak and the corresponding inflection point) at which a peak should be considered a beat, enforcing a minimum interval of 300 ms and maximum interval of 1500 ms between peaks, checking for both positive and negative slopes in a peak to ensure that baseline drift is not misclassified as a peak, and backtracking with reexamination/interpolation when peak missing was detected. Generally, the average change for heart rate is expected to be within the range of 2-15 bpm (Bradley, 2000). The chosen interval threshold between peaks was well above the rate of change of heart rate due to genuine heart acceleration. Time-domain features of IBI, such as the mean and standard deviation (SD), can be computed from the detected R peaks. IBI variability was explored by performing power spectral analysis on the IBI data to localize the sympathetic and parasympathetic nervous system activities associated with different frequency bands. "Sym" was the power associated with the sympathetic nervous system activity of the heart (in frequency band 0.04-0.15 Hz). "Para" was the power associated with the parasympathetic nervous system activity of the heart (in frequency band 0.15-0.4 Hz). "VLF" was the power associated with the frequency band less than 0.04 Hz. The ratios of different frequency components were also computed as the input features for affective modeling.

The PPG signal measures changes in the volume of blood in the fingertip associated with the blood volume pulse (BVP) cycle, and it provides an index of the relative constriction versus dilation of the blood vessels in the periphery. The raw PPG signal was smoothed by a 10 ms moving average window, and the baseline drift was accounted for by subtracting the average value of the signal. Pulse transit time (PTT) is the time it takes for the pulse pressure wave to travel from the heart to the periphery, and it was estimated by computing the time between systole at the heart (as indicated by the R-wave of the ECG) and the peak of the BVP wave reaching the peripheral site where PPG was measured. Besides PPT, the mean and SD values of BVP peak amplitudes were also extracted as features.

ICG analysis measures the impedance or opposition to the flow of an electric current through the body fluids. The ICG signal was first filtered by a 5th order Butterworth filter (low-pass: 10 Hz) to clean up any residual noise on the waveform and was then differentiated. A common variable in recent psychophysiology research, pre-ejection period (PEP) derived from ICG and ECG measures the latency between the onset of electromechanical systole and the onset of left-ventricular ejection and is most heavily influenced by sympathetic innervation of the heart. The time intervals between the successive peaks of ICG time-derivative and "R" peaks of ECG were calculated to obtain the value of PEP. The indices obtained were the mean of PEP and the average time interval between two peaks of the ICG time-derivative. The peak detection mechanisms used to determine the peaks of BVP and ICG time-derivative were similar to the ECG R-peak detection algorithm, while additional heuristic rules were added to reduce the degradation of the signal quality due to motion artifacts and avoid spurious peak detection with backtracking. Unlike ECG signals, the peak amplitudes of PPG and ICG showed a larger deviation over a given period of time. An adaptive thresholding rule was integrated in the peak detection algorithm to address this deviation, which continuously changed/updated the threshold value to determine whether candidates for peaks qualified as the valid peaks.

The heart sound signal measured sounds generated during each heartbeat. These sounds are produced by blood turbulence primarily due to the closing of the valves within the heart. The features extracted from the heart sound signal consisted of the mean and SD of the 3rd (138-275 Hz), 4th (69-138 Hz), and 5th (34-69 Hz) level coefficients of the Daubechies wavelet transform.

3.3 Electrodermal Activity

Electrodermal activity consists of two main components - tonic response and phasic response. Tonic skin conductance refers to the ongoing or the baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event related changes that are caused by a momentary increase in skin conductance (resembling a peak superimposed on tonic skin conductance). The raw EDA signal was smoothed by a 25 ms moving average window and then down-sampled by 10 to remove the high frequency measurement noise. The phasic skin conductance detection algorithm used the following heuristics for considering a particular peak as a valid skin conductance response: (i) the slope of the rise to the peak should be greater than 0.05 microsiemens/minute; (ii) the amplitude should be greater than 0.05 microsiemens; and (iii) the rise time should be greater than 0.25 seconds. Once the phasic responses were identified, we determined the rate of the responses and the mean and maximum phasic amplitude. All the signal points that were not included in the response constituted the tonic part of the skin conductance signal. The slope of tonic activity was obtained using linear regression. Another feature derived from tonic response was the mean tonic amplitude.

3.4 Electromyogram Activity

EMG measures the electrical activity in the muscle during contraction. The EMG signal from corrugator supercilii muscle (eyebrow) captures a person's frown and detects the tension in that region, and the EMG signal from the zygomaticus major muscle captures the muscle movements while smiling. Upper trapezius muscle EMG activity measures the tension in the shoulders, one of the most common sites in the body for developing stress. Time-domain features, such as the mean, SD, and slope were calculated from the EMG signals after performing a band-pass filtering operation (10-500 Hz). The analysis of the EMG activities in the frequency domain involved applying Fast Fourier Transform (FFT) on a given EMG signal, integrating the EMG spectrum, and normalizing it to [0,1] to calculate the two features of interest - the median frequency and mean frequency for each EMG signal. The blink-related features were determined from the corrugator supercilii EMG signals after being preprocessed by a low-pass filter (10 Hz).

3.5 Peripheral Temperature

Variations in the peripheral temperature mainly come from localized changes in blood flow caused by vascular resistance or arterial blood pressure and reflect the autonomic nervous system activity. The signal was down-sampled by 10 and filtered to remove high-frequency noise, from which the mean, SD, and the slope were calculated as features.

4. Experimental Investigation

4.1 Participants

Due to the fact that autism is a spectrum disorder (DSM-IV-TR, American Psychiatric Association, 2000), no one intervention technique will work for the entire population (NRC, 2001; Sherer and Schreibman, 2005). The research on autism intervention assistive tools is generally guided by the individual characteristics, needs, and preferences of the children (i.e., individual-specific approach) and focus on one sector of the population to develop a method with the flexibility to allow future modifications for a wider part of the population (Pioggia et al., 2005; Robins et al., 2005; Robins et al., 2004; Werry et al., 2001). The spectrum nature of autism and the phenomenon of person stereotypy (Lacey and Lacey, 1958) led us to choose an individual-specific approach to work on a long-term basis with a small group of children with autism in order to evaluate the affect recognition tool to be used in computer-assisted autism intervention.

Six participants in the age range of 13 to 16 years volunteered to participate in the experiments with the consent of their parents. Each had a diagnosis on the autism spectrum, either autistic disorder, Asperger's Syndrome, or pervasive developmental disorder not otherwise specified (PDD-NOS), according to their medical records. Participants were recruited using standard referral procedures that included (i) newsletters distributed through the Vanderbilt Treatment and Research Institute for Autism Spectrum Disorders, (ii) flyers placed in the Vanderbilt Center for Child Development, and (iii) website advertisements through the Vanderbilt Kennedy Center and the Autism Society of Middle Tennessee. The Institutional Review Board (IRB) approval was sought and received for conducting the experiment. Interested parents throughout middle Tennessee contacted the research office by phone or e-mail to set up an initial telephone screening. Monetary compensation (a $10 gift card per session) was given for the children's voluntary participation. Due to the nature of the designed cognitive tasks (as described in section 4.2), the following criteria were considered when choosing the participants: (i) a minimum competency level of age-appropriate language and cognitive skills (i.e., "high functioning") and (ii) no history of mental retardation. Each child with ASD underwent the Peabody Picture Vocabulary Test III (PPVT-III) to assess cognitive function (Dunn and Dunn, 1997). The PPVT-III is a measure of single-word receptive vocabulary that is often used as a proxy for intelligence quotient (IQ) testing because of its high correlations with standardized tests such as the Wechsler Intelligence Scale for Children (Bee and Boyd, 2004). It provides standard scores with a mean of 100 and a standard deviation of 15, and DSM-IV-TR (2000) classifies full scale IQ's above 70 as non-retarded. Participants in our study obtained a standard score of 80 or above on the PPVT-III measure. Table 2 shows the characteristics of the six children who participated in the experiments.

Table 2. The characteristics of the participants

Child ID	Gender	Age	Diagnosis	PPVT-III Score
A	Male	15	Autistic Disorder	99
B	Male	15	Asperger's Syndrome	80
C	Male	13	Autistic Disorder	81
D	Male	14	PDD-NOS	92
E	Male	16	PDD-NOS	93
F	Female	14	PDD-NOS	83

Several conditions posed challenges in recruitment of participants who matched the inclusion-exclusion criteria (e.g., cognitive skills and age range) and in coordination of schedules between the autism therapist and the parent of the participating child who were also involved in the experiment. First, autism may often co-occur with varying levels of mental retardation (DSM-IV-TR, American Psychiatric Association, 2000), which reduces the possible participant pool. Second, the IRB stipulates cutoffs between participants in different age ranges (e.g., 7-12 years, 13-17 years, 18 years and above, etc.), and autism intervention studies usually focus on one sector of the population within a certain age range (Gaylord-Ross et al., 1984; IRB, 2004; Parsons et al., 2005). Third, the responsibilities of raising a child with ASD are vast; therefore, willing parents often had to bring their child to the laboratory on weekends or after school on days without conflicts with other activities (e.g., social skills therapy) or family obligations. The group sizes and the cardinality of participant age range of many studies on computer-assisted autism intervention are commensurate with our work when an individual-specific approach was used (Pioggia et al., 2005; Robins et al., 2005; Robins et al., 2004; Werry et al., 2001). It is worth noting that this individual-specific study was based on a large sample size of observations for each child with ASD, which is comparatively more favorable than many other works (Groden et al., 2005; Pioggia et al., 2005; Robins et al., 2004). Each child completed approximately 85 epochs over 6 sessions, which represents 6 different days and yields 6 hours of data for each child as described in section 4.4. This preliminary study focused on high-functioning children with ASD between 13 and 16 years old. Future work may include a reduction of the verbal components in the cognitive tasks which would allow application to the broader ASD population.

4.2 Cognitive Tasks for Affect Elicitation

Two computer-based cognitive tasks were designed and implemented to invoke varying intensities of the following three affective states: anxiety, engagement, and liking, in the participants. Physiological data from participants were collected (as described in section 3) during the experiment. The two tasks consisted of an anagram solving task and a Pong playing task. The anagram solving task has been previously employed to explore relationships between physiology and anxiety (Pecchinenda and Smith, 1996). Emotional responses were manipulated in this task by presenting the participant with anagrams of varying difficulty levels, as established through pilot work. A long series of trivially easy anagrams caused less engagement. An optimal mix of solvable and difficult anagrams caused liking and engagement at times. Unsolvable or extremely difficult anagrams and giving time deadlines generated anxiety.

The Pong task consisted of a series of trials/epochs each lasting up to four minutes, in which the participant played a variant of the early, classic video game "Pong." This game has been used previously by researchers to study anxiety, performance, and gender differences (Brown et al., 1997). Various parameters of the game were manipulated to elicit the required affective responses. These included: ball speed and size, paddle speed and size, sluggish or over-responsive keyboard, random keyboard response, and the level of the computer opponent player. Very low speeds and large sizes of the ball and the paddle made games less engaging after a while; whereas high ball and paddle speeds along with smaller sizes of the two made the game engaging. Very high ball speeds and sluggish or over-responsive keyboard caused anxiety at times. Games with a moderate-level computer opponent player usually generated liking. The game configurations were established through pilot work.

Each task sequence was subdivided into a series of discrete epochs that were bounded by the subjective affective state assessments. These assessments were collected using a battery of questions about the target affective states and perceived task difficulty level rated on an eight-point Likert scale, where 1 indicated the lowest rating and 8 indicated the maximum rating. Each participant took part in six sessions – three one-hour sessions of solving anagrams and three one-hour sessions of playing Pong – on six different days. No more than one, one-hour session with an individual participant took place per day.

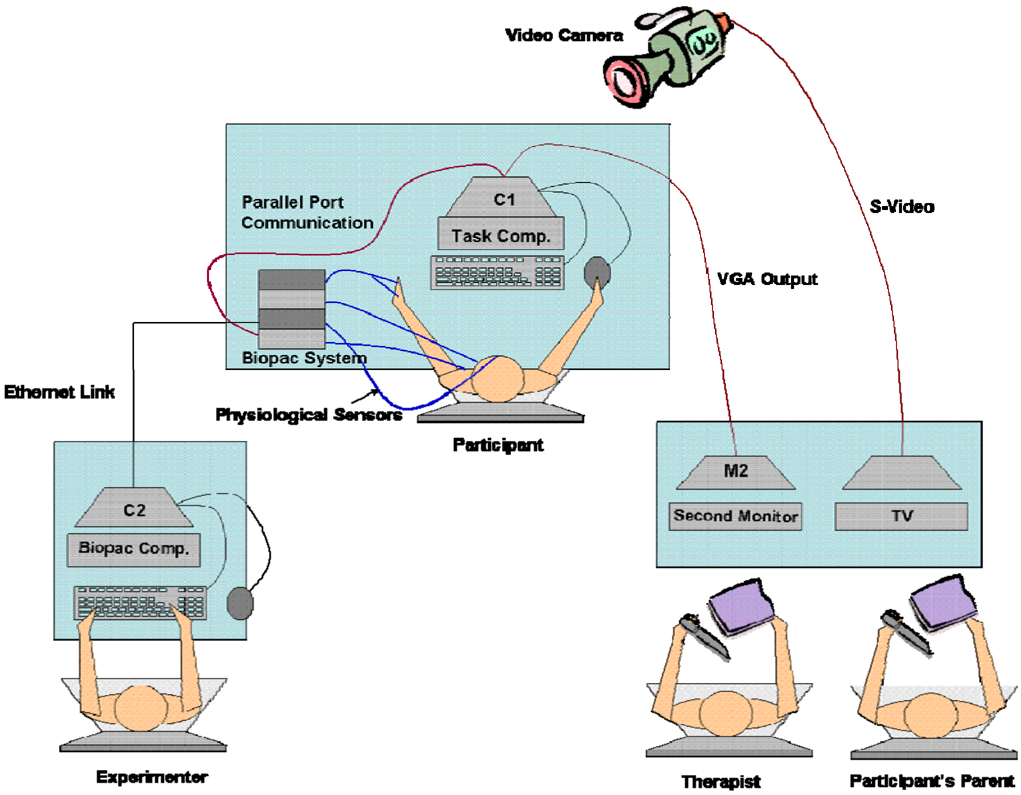


Figure 2. Experimental setup for collecting physiological data and subjective reports in the computer-based tasks

4.3 Experimental Setup

Figure 2 shows the setup for the experiment. The child with ASD was involved in the cognitive tasks on computer C1 while his/her physiological data was acquired via the Biopac system (www.biopac.com). Physiological signals were transferred from the Biopac transducers to C2 through an Ethernet link at 1000 Hz after being amplified, digitized, and stored. C1 was also connected to the Biopac system via a parallel port, through which the task-related markers were recorded along with the physiological data in a time-synchronized manner. Different markers were defined to indicate the following events: start/end of game, performance events (right/wrong answer in anagram, hitting/missing ball in Pong), start/end of each epoch, and self-report logging.

To gain perspective from different sources and enhance the reliability of the subjective reports on the target affective states, a therapist with experience in autism intervention for children with ASD and each participant's parent were also involved in the study, who may best know the participant. We video recorded the sessions to cross-reference observations made during the experiment. The signal from the video camera was routed to a television, and the signal from the participant's computer screen where the task was presented was routed to a separate computer monitor M2. The therapist and the participant's parent were seated at the back of the experiment room, watched the experiment on the TV from the view of the video camera, and observed how the task (anagrams or Pong) progressed on the separate monitor.

4.4 Procedure

On the first visit, participants completed the PPVT-III measurement to determine a standardized measure of receptive vocabulary and eligibility for the experiments. After initial briefing regarding the tasks, physiological sensors from a Biopac system were attached to the participant's body and a three-minute baseline recording was performed. Each session lasted about an hour and consisted of a set (13-15) of either 3-minute epochs for anagram tasks or up to 4-minute epochs for Pong tasks. Each epoch was followed by subjective report questions rated on an eight-point Likert scale. The participants reported their perceived subjective affective states through a pop-up dialog window presented on C1. The therapist and the participant's parent also answered the questions about how they thought the participant was feeling during the finished epoch on an eight-point Likert scale based on their audio/visual observations from the viewing monitors (TV and M2). These three sets of subjective reports related to the target affective states, from the therapist, the participant's parent, and the participant, were used as the possible reference points to link the concurrently collected objective physiological data to the participant's affective states.

For developing affective models, we built mappings to determine the intensity (i.e., high/low) of a particular affective state from the physiological features. It resembles a binary classification problem where the attributes are the physiological features (listed in Table 1) and the target function is the degree of arousal. In this work we employed SVM to determine the underlying affective state of a child with ASD given a set of physiological indices, based on our previous work (Rani et al., 2006) which showed SVM gave the best classification accuracy compared to KNN, Bayesian Network, and Regression Tree as applied to the domain of affect recognition using physiological signals for typical adults. Details of the theory and learning method of SVM can be found in (Vapnik, 1998) and are briefly described in the Appendix. Each participant had a data set that was comprised of both the objective physiological features and corresponding subjective reports on intensity of target affective states from the therapist, the participant's parent, and the participant. The subjective report forms instructed that 1-4 indicates the low level, 5-8 indicates the high level, and the different values represent the variation within each level. Each participant's data set contained approximately 85 epochs. Multiple subjective reports were analyzed, and one was chosen as the possible reference points to link the physiological measures to the participant's affective state. As illustrated in Figure 3, a therapist-like affect recognizer (i.e. a recognizer that captures the therapist's ability to assess affective states) can be developed when the therapist’s reports are used. Current therapeutic settings do not retain quantitative records of the affective states of the children with ASD. A therapist generally uses qualitative affective evaluations suitable for binary (high/low) assessments to make intervention adjustments (e.g., using likes/dislikes charts (Seip, 1996)). This study of differentiating high/low levels of the target affective states from physiological signals attempts to emulate the present autism intervention practice and to experimentally demonstrate the feasibility of affective modeling for these children with ASD via psychophysiological analysis. Further segmentation of the 8-point subjective reports and building multi-class recognizers will be used in a future study of computer-assisted autism intervention for a finer-grained analysis but such an analysis is beyond the scope of this paper.

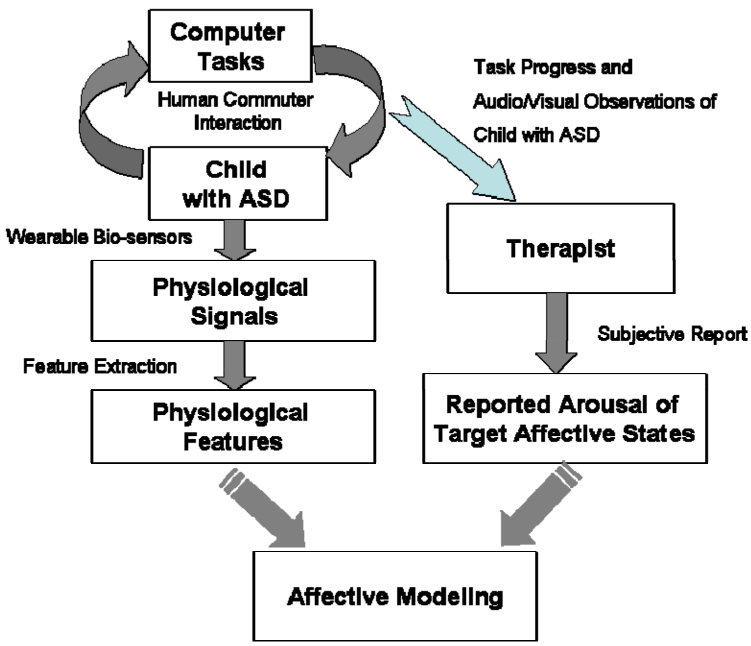


Figure 3. Affective modeling overview when the therapist’s subjective reports are used

A SVM-based recognizer was trained on each participant's data set for each target affective state. In this work, in order to deal with the nonlinearly separable data, soft-margin classifiers with slack variables were used to find a hyperplane with less restriction (Eqn. 1, Appendix). RBF (Radial Basis Function) was selected as the kennel function because it often delivers better performance (Vapnik, 1998). A ten-fold cross-validation was used to determine the kernel parameter and regularization parameter (Eqn. 2, Appendix) of the classifier.

5. Results and Discussion

One of the prime challenges of this work is attaining reliable subjective reports. While there have been reports that adolescents could be better sources of information than adults when it comes to measuring some psychiatric symptoms (Cantwell et al., 1997), researchers are reluctant to trust the responses of adolescents on self-reports (Barkley, 1998). In this study, one should be especially wary of the dependability of self-reports from children with ASD, who may have deficits in processing (i.e., identifying and describing) their own emotions (Hill et al., 2004). In order to overcome this difficulty, a therapist and a parent of each participant observed the experiment and provided subjective reports based on their expertise/experience in inferring presumable underlying affective states from the observable behaviors of children with ASD. Their reports about how they thought the participant was feeling were collected after each epoch.

To measure the amount of agreement among the different reporters, the kappa statistic was used (Siegel and Castellan, 1988). The kappa coefficient (K) measures pair-wise agreement among a set of reporters making category judgments, correcting for expected chance agreement:

$$K = \frac{P(A) - P(E)}{1 - P(E)} \tag{3}$$

where $P(A)$ is the proportion of times that the reporters agree and $P(E)$ is the proportion of times that we would expect the reporters to agree by chance. When there is complete agreement, then $K = 1$; whereas, when there is no agreement other than that which would be expected by chance, then $K = 0$.

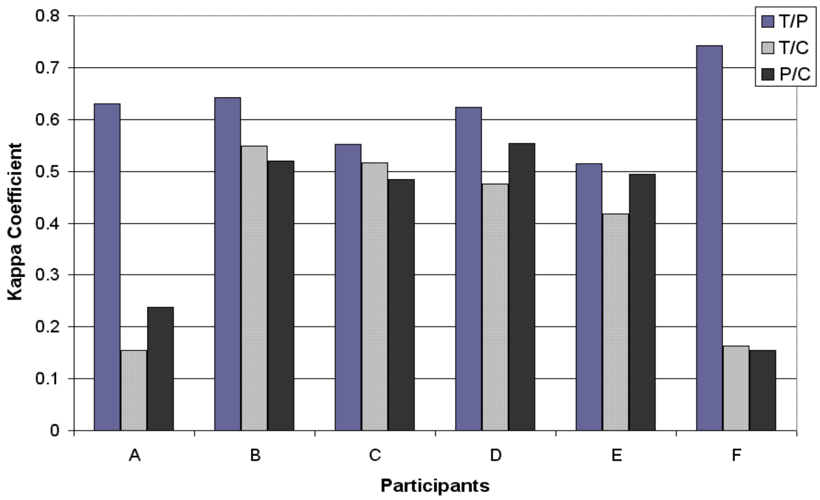


Figure 4. Average Kappa Statistics between Reporters for Affective States. Kappa coefficients averaged across affective states measure the agreement between the different subjective reports (T-Therapist, P-Parent, C-Child) corresponding to each participant (Child ID A-F)

The results of the values of kappa coefficient (K), averaged across three target affective states, are shown in Figure 4. From the results, we can see that among the three possible pairs for each child (Therapist-Parent (T/P), Therapist-Child (T/C), and Parent-Child (P/C)) the agreement between the therapist and each participant's parent (T/P) shows the largest mean of the kappa statistic values. The data were submitted to two related-samples t-tests, which were both significant ($t(17) = 4.28$, $p < 0.001$ for the

Therapist-Parent pair; $t(17) = 3.70$, $p < 0.01$ for the Therapist-Child pair). Note that the Kappa agreement between therapist and parent is substantial for Child A, Child B, Child D, and Child F and moderate for Child C and Child E. Such results might stem from the fact that it could be difficult for the therapist or parent to distinguish certain emotions for a particular child with ASD. For example, the agreement between therapist and parent for the anxiety level of Child C and Child E (Kappa coefficient: 0.352 and 0.372, respectively) are considerably less than the average level (mean Kappa coefficient of T/P: 0.617). In the experiment, Child A and Child F's self-ratings for liking, anxiety, and engagement were almost constant which resulted in lower kappa statistic values for the therapist and child pair (T/C) and the parent and child pair (P/C) than those of the other participants. This may be due to the fact that the spectrum developmental disorder for children with autism manifests in different abilities to recognize and report their own emotions. A lack of agreement with adults does not necessarily mean that the self-report of children with ASD is not dependable. However, given the fact that therapists' judgment based on their expertise is the state-of-the-art in most autism intervention approaches and the reasonably high agreement between the therapist and the parents for all of the six children, the subjective report from the therapist was used as the reference points linking the objective physiological data to the children's affective states. In order to enhance the consistency of the subjective reports, the same therapist was involved in all of the sessions. This choice allows for building a *therapist-like* affective model. Once the affect modeling is completed, the recognizers will be capable of autonomously inferring the affective states of the child with ASD from the physiological signals in real time even when the therapist is not available.

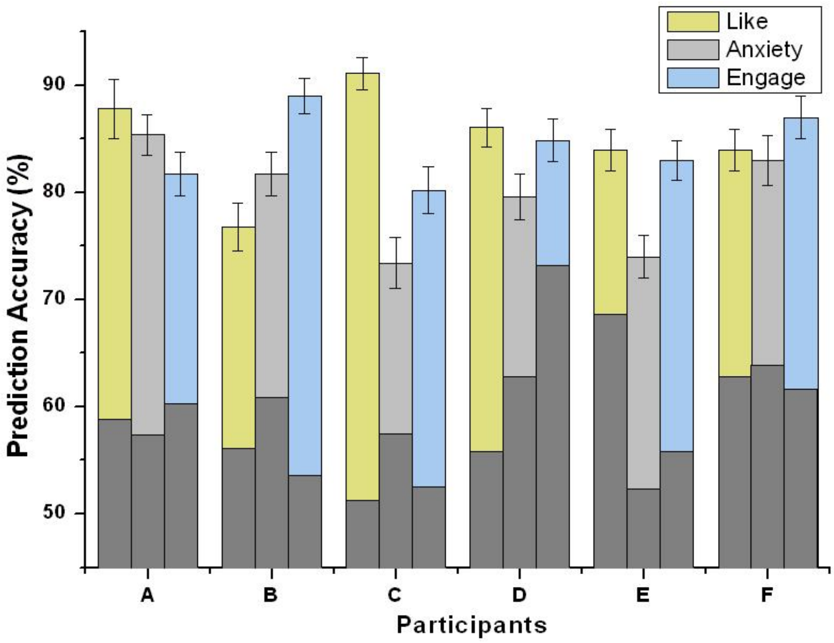


Figure 5. Prediction Accuracy of the Affective Model

The performance of the developed affective models based on the therapist's reports for each child (i.e., individual-specific approach) is shown in Figure 5. The cross-validation method, "leave-one-out," was used. The affective model produced high recognition accuracies for each target affective state of each participant. The average correct prediction accuracies across all participants with ASD were: 85.0% for liking, 79.5% for anxiety, and 84.3% for engagement, which are comparable to the best results achieved for typical adults (Nasoz et al., 2003; Picard et al., 2001; Rani et al., 2006). Figure 5 shows that for Child C and Child E, the prediction accuracy for anxiety is lower; moreover, as mentioned previously for these two participants there is also considerably less agreement between the therapist and the parent (T/P) on the subjective reports with respect to the anxiety level. The comparatively low (approximately 5% less) average prediction accuracy of anxiety may be due to the fact that the intensity of anxiety of a particular child with ASD (e.g., Child C and Child E) could be more difficult for the therapist to distinguish based on the observations than the other two affective states (i.e., liking and engagement).

We also compared the performance of affective modeling to a control method that represents random chance. Suppose we had an equal number of subjective reports that rated a particular affective state level (low/high) for a participant, then the chance probability would be 50%. However, the prevalence of each level could be different. For example, in 48 out of 86 epochs the engagement of Child E was rated as low, where a random classification could assign all test epochs to this category and make accurate classifications $(48/86)\times100 = 55.8\%$ of the time. We thus considered the level with a majority of epochs and used the average of these higher numbers (across the participants' affective states) to represent the chance condition, which is denoted by dark grey bars in Figure 5. While the physiology-based affective modeling alone did not provide perfect classification (i.e., 100%) of affective states of children with ASD, they did yield reliable matches with the subjective rating and significantly outperformed a random classifier (averaging 82.9% vs. 59.2%). This was promising considering that this task was challenging in two respects: (i) the reports were collected from the therapist who was observing the children with ASD as opposed to having typical adults capable of differentiating and reporting their own affective states and (ii) varying levels of arousal of any given affective state (e.g., low/high anxiety) were identified instead of determining between two discrete affective states.

Table 3. Prediction Accuracy of the Affective Modeling based on Different Physiological Signals (%)*

Physiological Signals	Liking	Anxiety	Engage	Mean
Cardiovascular	75.7	68.5	76.2	73.5
Electrodermal	73.4	72.3	73.3	73.0
Electromyographic	73.1	65.8	70.1	69.7
Electrodermal + Electromyographic	75.0	69.4	71.4	71.9
Cardiovascular + Electromyographic	79.6	70.2	79.9	76.6
Cardiovascular + Electrodermal	79.9	74.3	81.9	78.7
All	85.0	79.5	84.3	82.9

In order to explore the effects of reducing the number of physiological signals and the possibility of achieving more economical modeling (i.e., reducing the set of signals to be measured), we examined the performance of the affect recognizers when cardiovascular, electrodermal, and electromyographic activities and their combinations were used. As shown in Table 3, all the recognizers delivered better predication than random guess (mean prediction rate: 52.9%), and with more information from physiological activities the performance of the affective models tends to improve (except the combination of electrodermal and electromyographic activities). This may be due to the fact that the inherent kernel representation and soft-margin optimization endow SVM the capability to work effectively in the high-dimensional feature space (Burges, 1998). While electromyographic (EMG) signals have been used as indicators of affective response for typical individuals (Kulic and Croft, 2007; Rani et al., 2006), in this study

* Peripheral temperature has relatively few features derived as shown in Table 1 and was not examined independently. Instead, it was studied conjunctively with the electrodermal activity, both of which were acquired from the non-dominant hand of a participant.

we observed that it is less discriminatory than the cardiovascular and electrodermal activities. As suggested in (DSM-IV-TR, American Psychiatric Association, 2000; Green et al., 2002), children with ASD often have nonverbal communicative impairments regarding expression of affective states (e.g., abnormal body postures and gestures and absence of facial expression), which might reduce the discriminatory capability of EMG signals (e.g., muscle activities from both the corrugator supercilii and the zygomaticus major) to reveal affective cues of the participants. While no combination of physiological activity surpassed the percent accuracy achieved when all signals were used, the results in Table 3 suggested that it may be possible to selectively reduce the set of signals and obtain nearly-as-good performance (e.g., using a combination of cardiovascular and electrodermal signals).

With post-hoc analysis, we found the prediction accuracy generally tends to be higher when the therapist and the participant's parent agree more on the subjective reports about how they thought the participant was feeling during the finished epoch. As shown in Table 4, the Kappa statistic of the therapist and parent is positively correlated with the prediction accuracy of the developed affect recognizer ($r = 0.71$, $p < 0.001$). In this experiment, the Kappa statistic could indicate whether it is relatively easy or difficult to differentiate the affective states of a child by observation. The autism therapist used in this work had no previous interaction with the participants. The prediction accuracy could likely improve if the therapist interacts with a particular child with ASD for a significant amount of time and gains more knowledge of his/her affective expression before making the reports regarding the presented interaction tasks, which is generally the case for ASD intervention.

Table 4. Therapist-Parent (T/P) Kappa Statistics and Prediction Accuracy

Child ID		Liking	Anxiety	Engage
A	Kappa Statistics (T/P)	0.566	0.831	0.494
	Prediction Accuracy (%)	87.8%	85.4%	81.7%
B	Kappa Statistics (T/P)	0.585	0.634	0.708
	Prediction Accuracy (%)	76.8%	81.7%	89.0%
C	Kappa Statistics (T/P)	0.753	0.352	0.551
	Prediction Accuracy (%)	91.1%	73.4%	80.2%
D	Kappa Statistics (T/P)	0.698	0.562	0.611
	Prediction Accuracy (%)	86.1%	79.6%	84.9%
E	Kappa Statistics (T/P)	0.721	0.372	0.449
	Prediction Accuracy (%)	83.7%	74.1%	83.2%
F	Kappa Statistics (T/P)	0.884	0.528	0.814
	Prediction Accuracy (%)	84.8%	82.6%	87.2%

6. Conclusions and Future Work

There is increasing consensus that development of assistive therapeutic tools can make application of intensive intervention for children with ASD more readily accessible. In recent years, various applications of advanced interactive technologies have been investigated in order to facilitate and/or partially automate the existing behavioral intervention that addresses specific deficits associated with autism. However, the current computer-assisted therapeutic tools for children with ASD do not possess the ability of deciphering the affective cues of the children, which could be critical given that the affective factors of children with ASD have significant impacts on the intervention practice. In this work, we presented a physiology-based affect modeling framework for children with ASD. The developed model could allow the recognition of affective states of the child with ASD from the physiological signals in real time and provide the basis for computer-based affect-sensitive interactive autism intervention.

We have designed and implemented two computer-based cognitive tasks – solving anagrams and playing Pong – to elicit the affective states of liking, anxiety, and engagement for children with ASD that are considered important in autism intervention. In order to have reliable reference points to link the physiological data to the affective states, the reports from the child, the therapist, and the child's parent were collected and analyzed. We have investigated a large set of physiological indices that may correlate with the affective states of children with ASD. A SVM-based affective model yielded reliable prediction with approximately 82.9% success when using the therapist's reports. This is the first time, to our knowledge, that the affective states of children with ASD have been experimentally detected via a physiology-based affect recognition technique.

It should be noted that due to the phenomenon of person stereotypy and the spectrum nature of autism, an individual-specific approach has been employed for affective modeling based on a large sample size of observations (as described in section 4.4) of each of the six participating children with ASD. The methodology for inducing, gathering, and modeling the experimental data in this paper is not dependent on the participants. The group sizes and the cardinality of participant age range of many related studies are commensurate with our work and the sample size of observations in this work is comparatively extensive. The consistently reliable prediction accuracy for each participant demonstrated that it was feasible to model the affective states of these children with ASD via psychophysiological analysis.

Note that the presented work requires physiological sensing that has its own limitations. For example, one needs to wear physiological sensors, and use of such sensors could be restrictive under certain circumstances. However, none of the participants in our study had any objection in wearing the physiological sensors. Similar observations were achieved in (Conati et al., 2003) that suggested concerns for intrusiveness of physiological sensors could be lessened for children in a game-like environment. Given the rapid progress in wearable computing with small, non-invasive sensors and wireless communication, physiological sensors can be worn in a wireless manner such as in physiological sensing clothing and accessories (Picard, 1997; Wijesiriwardana et al., 2004), which could alleviate possible constraints on experimental design. We believe that physiology-based affect recognition can be appropriate and useful for the application of interactive autism intervention and could be used conjunctively with other modalities (e.g., facial expression, vocal intonation, etc.) to allow flexible and robust affective modeling for children with ASD.

Future work includes a reduction of the verbal components in the cognitive tasks which would allow application to a broader part of the ASD population. Computer-based intervention tools that address the social communication deficits of children with ASD will be developed. We will also investigate how to augment the interactive autism intervention by having a computer respond appropriately to the inferred affects based on the affective model described here.

Appendix. Pattern Recognition using Support Vector Machines

Here we briefly describe the principle of classification using SVM. SVM is a linear machine working in a high k-dimensional feature space formed by an implicit embedding of n-dimensional input data X (e.g., a vector of derived physiology features as described in section 3) into a k-dimensional feature space ($k > n$) through the use of a nonlinear mapping $\phi(X)$. This allows using linear algebra and geometry to separate the data normally only separable with nonlinear rules in the input space. The problem of finding a linear classifier for given data points with known class labels can be described as finding a separating hyperplane $W^T\varphi(X)$ that satisfies

$$y_i\left(W^T\varphi(X_i)\right)=y_i\left(\sum_{j=1}^{k}w_j\phi_j(X_i)+w_0\right)\geq 1-\xi_i \text{ for } i=1,2,...,N \tag{1}$$

where $y_i\in\{+1,-1\}$ represents the class label (e.g., high/low intensity of a target affective state); N is the number of training data pairs (X_i,y_i); $\varphi(X)=\left[\phi_0(X),\phi_1(X),...,\phi_k(X)\right]^T$ is the mapped feature vector ($\phi_0(X)=1$); and $W=\left[w_0,w_1,...,w_k\right]$ is the weight vector of the network. The nonnegative slack variable ξ_i generalizes the linear classifier with soft margin to deal with nonlinearly separable problems, where margin is defined as the sum of the distances of the hyperplane from the nearest data points of each of the two classes. Good generalization performance can be achieved by maximizing the margin while allowing for some misclassifications on the training set to avoid over-fitting (Burges, 1998).

To allow efficient computation of inner products directly in the feature space and circumvent the difficulty of specifying the non-linear mapping explicitly, all operations in learning and testing modes are done in SVM using so-called kernel functions satisfying Mercer conditions defined as $K(X_i,X)=\varphi^T(X_i)\varphi(X)$ (Vapnik, 1998). The most distinctive fact about SVM is that the learning task is reduced to a dual quadratic programming problem by introducing the Lagrange multipliers α_i (Burges, 1998; Vapnik, 1998):

Maximize

$$Q(\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K\left(X_i,X_j\right)$$

Given the constraints (2)

$$\sum_{i=1}^{N}\alpha_i y_i=0 \qquad 0\le\alpha_i\le C$$

where C is a user-defined regularization parameter that determines the balance between the complexity of the network characterized by the weight vector W and the error of classification of data. The corresponding α_i's are non-zero only for the support vectors, those training points nearest to the hyperplane, which induces solution sparseness. The quadratic programming of SVM leads in all cases to the global minimum of the cost function. With the kernel representation and soft margin mechanism, SVM provides an efficient technique that can tackle the difficult, high dimensional affect recognition problem even when physiological data are noisy.

References

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders: DSM-IV-TR (4th ed.). Washington, DC: American Psychiatric Association.

Asha, K., Ajay, K., Naznin, V., George, T., & Peter, F. D. (2005). Gesture-based affective computing on motion capture data. Paper presented at the Int. Conf. on Affective Computing and Intelligent Interaction (ACII), Beijing, China.

Barkley, R. A. (1998). Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment (2 ed.). New York: Guilford Press.

Bartlett, M. S., Littlewort, G., Fasel, I., & Movellan, J. R. (2003). Real time face detection and facial expression recognition: development and applications to human computer interaction. Paper presented at the Computer Vision and Pattern Recognition Workshop, Madison, Wisconsin.

Bee, H. & Boyd, D. (2004). The Developing Child. (10th ed.). Boston: Pearson.

Ben Shalom, D., Mostofsky, S. H., Hazlett, R. L., Goldberg, M. C., Landa, R. J., Faran, Y., McLeod, D. R., & Hoehn-Saric, R. (2006). Normal physiological emotions but differences in expression of conscious feelings in children with high-functioning autism. J Autism Dev Disord, 36(3), 395-400.

Bernard-Opitz, V., Sriram, N., & Nakhoda-Sapuan, S. (2001). Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. J Autism Dev Disord, 31(4), 377-384.

Bradley, M. M. (2000). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary & G. Berntson (Eds.), Handbook of Psychophysiology (pp. 602-642). New York: Cambridge University Press.

Brown, R. M., Hall, L. R., Holtzer, R., Brown, S. L., & Brown, N. L. (1997). Gender and video game performance. Sex Roles, 36(11-12), 793 – 812.

Burges, C. J. C. (1998). A tutorial on Support Vector Machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2), 121-167.

Cacioppo, J.T., Berntson, G.G., Larsen, J.T., Poehlmann, K.M., & Ito, T.A., 2000. The psychophysiology of emotion. In: Lewis, M., & Haviland-Jones, J.M. (Eds.), Handbook of Emotions. The Guilford Press, New York, pp. 173–191.

Cantwell, D. P., Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1997). Correspondence between adolescent report and parent report of psychiatric diagnostic data. Journal of the American Academy of Child and Adolescent Psychiatry, 36, 610-619.

CDC. (2007). Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, 14 sites, United States, 2002. MMWR Surveill Summ, 56(1), 12-28.

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: replication of the UCLA model in a community setting. J Dev Behav Pediatr, 27(2 Suppl), S145-155.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. Applied Artificial Intelligence, 16(7-8), 555-575.

Conati, C., Chabbal, R., & Maclaren, H. (2003). A Study on Using Biometric Sensors for Detecting User Emotions in Educational Games. Workshop on Assessing and Adapting to User Attitude and Affects: Why, When and How, Pittsburgh, Pennsylvania.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. (2001). Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine, 18(1), 32-80.

Dautenhahn, K., & Werry, I. (2004). Towards interactive robots in autism therapy: background, motivation and challenges. Pragmatics & Cognition, 12(1), 1-35.

Dunn, L. M., & Dunn, L. M. (1997). PPVT-III: Peabody Picture Vocabulary Test-Third Edition Circle Pines, Minnesota: American Guidance Service.

Ernsperger, L. (2003). Keys to Success for Teaching Students with Autism: Future Horizons.

Gaylord-Ross, R. J., Haring, T. G., Breen, C., & Pitts-Conway, V. (1984). The training and generalization of social interaction skills with autistic youth. Journal of Applied Behavior Analysis, 17, 229–247.

Gillott, A., Furniss, F., & Walter, A. (2001). Anxiety in high-functioning children with autism. Autism, 5(3), 277-286.

Green, D., Baird, G., Barnett, A. L., Henderson, L., Huber, J., & Henderson, S. E. (2002). The severity and nature of motor impairment in Asperger's syndrome: a comparison with Specific Developmental Disorder of Motor Function. Journal of Child Psychology and Psychiatry and Allied Disciplines, 43(5), 655-668.

Groden, J., Goodwin, M. S., Baron, M. G., Groden, G., Velicer, W. F., Lipsitt, L. P., Hofmann, S. G., & Plummer, B. (2005). Assessing Cardiovascular Responses to Stressors in Individuals with Autism Spectrum Disorders. Focus on Autism and Other Developmental Disabilities, 20(4), 244-252.

Hill, E. Berthoz, S. and Frith, U. (2004). Brief report: cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. Journal of Autism and Developmental Disabilities, 34(2), 229-235.

IRB. (February 9, 2004). IRB Policy IV.B.1: Procedure for Documentation of Informed Consent for Human Subjects Research. Vanderbilt University Medical Center, http://mcapps01.mc.vanderbilt.edu/IRB/policy&procedures.nsf.

Kataoka, H., Kano, H., Yoshida, H., Saijo, A., Yasuda, M., & Osumi, M. (1998). Development of a skin temperature measuring system for non-contact stress evaluation. IEEE Ann. Conf. Engineering Medicine Biology Society, pp. 940–943.

Kleinsmith, A., Fushimi, T., & Bianchi-Berthouze, N. (2005). An incremental and interactive affective posture recognition system. Paper presented at the UM 2005 Workshop: Adapting the Interaction Style to Affective Factors, Edinburgh, UK.

Kozima, H., Nakagawa, C., & Yasuda, Y. (2005). Interactive robots for communication-care: A case-study in autism therapy. Paper presented at the IEEE International Workshop on Robot and Human Interactive Communication, Nashville, Tennessee.

Kulic, D., & Croft, E. (2007). Physiological and subjective responses to articulated robot motion. Robotica, 25, 13-27.

Lacey, J. I., & Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy. Am J Psychol, 71(1), 50-73.

Lee, C. M., & Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. IEEE Transactions on Speech and Audio Processing, 13(2), 293-303.

Liu, C., Rani, P., & Sarkar, N. (2006). Human-Robot interaction using affective cues. Paper presented at the International Symposium on Robot and Human Interactive Communication, Hatfield, United Kingdom.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. International Journal of Human-Computer Studies, 65(4), 329-347.

Moore, D. J., McGrath, P., & Thorpe, J. (2000). Computer aided learning for people with autism - A framework for research and development. Innovations in Education and Training International, 37(3), 218-228.

Nasoz, F., Alvarez, K., Lisetti, C., & Finkelstein, N. (2003). Emotion recognition from physiological signals for presence technologies. Int J Cogn Technol Work Spec Issue Presence, 6(1).

NRC (National Research Council). (2001). Educating children with autism. Washington, DC: National Academy Press.

Papert, S. (1993). Mindstorms: Children, Computers, and Powerful Ideas (2 ed.). New York: Basic Books.

Papillo, J.F., & Shapiro, D., 1990. The cardiovascular system. In: Cacioppo, J.T., & Tassinary, L.G. (Eds.), Principles of Psychophysiology: Physical, Social, and Inferential Elements. Cambridge University Press, Cambridge, pp. 456–512.

Parsons, S., & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. J Intellect Disabil Res, 46(Pt 5), 430-443.

Parsons, S., Mitchell, P., & Leonard, A. (2005). Do adolescents with autistic spectrum disorders adhere to social conventions in virtual environments? Autism, 9(1), 95-117.

Pecchinenda, A., & Smith, C. A. (1996). The affective significance of skin conductance activity during a difficult problem-solving task. Cognition and Emotion, 10(5), 481-504.

Picard, R. W. (1997). Affective Computing. Cambridge: The MIT Press.

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

Pioggia, G., Igliozzi, R., Ferro, M., Ahluwalia, A., Muratori, F., & De Rossi, D. (2005). An android for enhancing social skills and emotion recognition in people with autism. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13(4), 507-515.

Prendinger, H., Mori, J., & 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.

Rani, P., Liu, C. C., Sarkar, N., & Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human-robot interaction. Pattern Analysis and Applications, 9(1), 58-69.

Rani, P., Sarkar, N., Smith, C. A., & Kirby, L. D. (2004). Anxiety detecting robotic system - towards implicit human-robot collaboration. Robotica, 22, 85-95.

Reeves, B., & Nass, C. I. (1996). The media equation: how people treat computers, televisions, and new media as real people and places. New York: Cambridge University Press.

Robins, B., Dickerson, P., & Dautenhahn, K. (2005). Robots as embodied beings – Interactionally sensitive body movements in interactions among autistic children and a robot. Proc. 14th IEEE International Workshop on Robot and Human Interactive Communication, Nashville, Tennessee.

Robins, B., Dickerson, P., Stribling, P., & Dautenhahn, K. (2004). Robot-mediated joint attention in children with autism: A case study in robot-human interaction. Interaction Studies, 5(2), 161–198.

Rogers, S. J. (1998). Empirically supported comprehensive treatments for young children with autism. J Clin Child Psychol, 27(2), 168-179.

Ruble, L. A., & Robson, D. M. (2006). Individual and Environmental Determinants of Engagement in Autism. J Autism Dev Disord.

Rutter, M. (2006). Autism: its recognition, early diagnosis, and service implications. J Dev Behav Pediatr, 27(2 Suppl), S54-58.

Schultz, R.T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. Int J Dev Neurosci, vol. 23, pp. 125-41.

Seip, J. A. (1996). Teaching the autistic and developmentally delayed: A guide for staff training and development: Delta, British Columbia.

Sherer, M. R., & Schreibman, L. (2005). Individual Behavioral Profiles and Predictors of Treatment Effectiveness for Children with Autism. Journal of Consulting and Clinical Psychology, 73(3), 525-538.

Siegel, S., & Castellan, J. N. J. (1988). Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill.

Silver, M., & Oakes, P. (2001). Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others. Autism, 5(3), 299-316.

Swettenham, J. (1996). Can children with autism be taught to understand false belief using computers? J Child Psychol Psychiatry, 37(2), 157-165.

Tarkan, L. (October 21, 2002). Autism therapy is called effective, but rare. New York Times.

Toichi, M., & Kamio, Y. (2003). Paradoxical autonomic response to mental tasks in autism. J Autism Dev Disord, 33(4), 417-426.

Trepagnier, C. Y., Sebrechts, M. M., Finkelmeyer, A., Stewart, W., Woodford, J., & Coleman, M. (2006). Simulating social interaction to address deficits of autistic spectrum disorder in children. Cyberpsychol Behav, 9(2), 213-217.

Vansteelandt, K., Van Mechelen, I., & Nezlek, J. B. (2005). The co-occurrence of emotions in daily life: A multilevel approach. Journal of Research in Personality, 39(3), 325-335.

Vapnik, V. N. (1998). Statistical Learning Theory. New York: Wiley-Interscience.

Werry, I., Dautenhahn, K., & Harwin, W. (2001). Investigating a robot as a therapy partner for children with autism. In: Proceedings of the 6th European conference for the advancement of assistive technology, Ljubljana, Slovenia.

Wieder, S., & Greenspan, S. (2005). Can Children with Autism Master the Core Deficits and Become Empathetic, Creative, and Reflective? The Journal of Developmental and Learning Disorders, 9.

Wijesiriwardana, R., Mitcham, K., & Dias, T. (2004). Fibre-meshed transducers based real time wearable physiological information monitoring system. Paper presented at the International Symposium on Wearable Computers, Washington, DC.

CHAPTER VI: MANUSCRIPT 5

ONLINE AFFECT DETECTION AND ROBOT BEHAVIOR ADAPTATION FOR INTERVENTION OF CHILDREN WITH AUTISM

Changchun Liu, Karla Conn, Nilanjan Sarkar, *Member, IEEE,* and Wendy Stone

(Published as Regular Paper in IEEE Transactions on Robotics)

Abstract

Investigation into robot-assisted intervention for children with autism spectrum disorder (ASD) has gained momentum in recent years. Therapists involved in interventions must overcome the communication impairments generally exhibited by children with ASD by adeptly inferring the affective cues of the children to adjust the intervention accordingly. Similarly, a robot must also be able to understand the affective needs of these children - an ability that the current robot-assisted ASD intervention systems lack - to achieve effective interaction that addresses the role of affective states in human-robot interaction and intervention practice. In this paper we present a physiology-based affect-inference mechanism for robot-assisted intervention where the robot can detect the affective states of a child with ASD as discerned by a therapist and adapt its behaviors accordingly. This work is the first step towards developing "understanding" robots for use in future ASD intervention. Experimental results with 6 children with ASD from a proof-of-concept experiment (i.e., a robot-based basketball game) are presented. The robot learned the individual liking level of each child with regard to the game configuration and selected appropriate behaviors to present the task at his/her preferred

liking level. Results show the robot automatically predicted individual liking level in real time with 81.1% accuracy. This is the first time, to our knowledge, that the affective states of children with ASD have been detected via a physiology-based affect recognition technique in real time. This is also the first time that the impact of affect-sensitive closed-loop interaction between a robot and a child with ASD has been demonstrated experimentally.

Index Terms: Autism intervention, closed-loop human-robot interaction, and physiological sensing

I. Introduction

Autism is a neurodevelopmental disorder characterized by core deficits in social interaction, social communication, and imagination [1]. Emerging research suggests prevalence rates as high as approximately 1 in 150 for the broad autism spectrum [2]. While there is at present no single accepted intervention, treatment, or known cure for autism spectrum disorders (ASD), there is growing consensus that intensive behavioral and educational intervention programs can significantly improve long term outcomes for individuals and their families [3]. Despite the urgent need and societal import of intensive treatment [4], appropriate intervention resources for children with ASD and their families are often extremely costly when accessible [5]. Therefore, an important new direction for research on ASD is the identification and development of assistive intervention tools that can make application of intensive treatment more readily accessible.

In response to this need, a growing number of studies have been investigating the application of advanced interactive technologies to address core deficits related to autism, namely computer technology [6], virtual reality environments [7], and robotic systems

[8]–[13]. Initial results indicate that robots may hold promise for rehabilitation of children with ASD. Dautenhahn and Werry [8] have explored how a robot can become a playmate that might serve a therapeutic role for children with autism in the Aurora project. Research suggested that robots can allow simplified but embodied social interaction that is less intimidating or confusing for children with ASD [8]. Michaud and Theberge-Turmel [9] investigated the impact of robot design on the interactions with children and emphasized that systems need to be versatile enough to adapt to the varying needs of different children. Pioggia et al. [10] developed an interactive life-like facial display system for enhancing emotion recognition in individuals with ASD. Robots have also been used to teach basic social interaction skills using turn-taking and imitation games, and the use of robots as social mediators and as objects of shared attention can encourage interaction with peers and adults [8][11][12]. Robotic technology poses the advantage of furnishing robust systems that can support multimodal interaction and provide a repeatable, standardized stimulus while quantitatively recording and monitoring the performance progress of the children with ASD to facilitate autism intervention assessment and diagnosis [13]. By employing human-robot interaction (HRI) technologies, robot-based therapeutic tools can partially automate the time-consuming, routine behavioral therapy sessions and may allow intensive intervention to be conducted at home [8].

Even though there is increasing research in robot-assisted autism intervention, the authors found no published studies that specifically addressed how to automatically detect and respond to affective cues of children with ASD. We believe that such ability could be critical given the importance of human affective information in HRI [14][15] and the significant impacts of the affective factors of children with ASD on the intervention practice [16]. Common in autism intervention, therapists who work with children with ASD continuously monitor affective cues of the children in order to make appropriate decisions about adaptations to their intervention strategies. For example, 'likes and dislikes chart' is recommended to record the children's preferred activities and/or sensory stimuli during interventions that could be used as reinforcers and/or 'alternative behaviors' [16]. Children with autism are particularly vulnerable to anxiety and intolerant of feelings of frustration, which requires a therapist to plan tasks at an appropriate level of difficulty [17]. The engagement of children with ASD is the ground basis for the 'floor-time therapy' to help them develop relationships and improve their social and communication skills [18].

The potential impacts brought by a robot that can detect the affective states of a child with ASD and interact with him/her based on such perception could be various. Interesting activities could be chosen to retain the child's attention when the detected engagement level is low. Complex social stimuli, sophisticated interactions, and unpredictable situations could be gradually but automatically introduced when the robot recognizes that the child is comfortable or not anxious at a certain level of interaction dynamics for a reasonably long period of time. A therapist could use the child's affective records to analyze the therapeutic approach. With the record of the activities and the consequent emotional changes in a child, a robot could learn individual affective characteristics over time and thus could adapt the ways it responds to the needs of different children.

The primary objective of the current research is to investigate how to augment human-robot interaction to be used in autism intervention by endowing the robot with the ability to recognize and respond to the affective states of a child with ASD. In order to achieve this objective, the research is divided into two phases. Phase I represents the development of affective models through psychophysiological analysis, which includes designing cognitive tasks for affect-elicitation, deriving physiological features via signal processing, and developing affective models using machine learning techniques. Phase II is characterized by the investigation of affect sensitivity during the closed-loop interaction between a child with ASD and the robot. A proof-of-concept experiment was designed wherein a robot learns individual preferences based on the predicted liking level of the children with ASD as discerned by the therapist and selects an appropriate behavior accordingly.

The paper is organized as follows: The scope and rationale of this work is presented in Section II. Section III describes our proposed framework for automatically detecting and responding to affective cues of children with ASD in the human-robot interaction, as well as the experimental design. This description is followed by the detailed results and discussion section (Section IV). Finally, Section V summarizes the contributions of the paper and outlines possible future directions of this research.

II. Scope and Rationale

The overview of the affect-sensitive closed-loop interaction between a child with ASD and a robot is presented in Fig. 1. The physiological signals from the children with ASD are recorded when they are interacting with the robot. These signals are processed in real time to extract features, which are fed as input into the models developed in Phase I. The models determine the perceived affective cues and return this information as an output. The affective information, along with other environmental inputs, is used by a controller to decide the next course of action for the robot. The child who engages with the robot is then influenced by the robot's behavior, and the closed-loop interaction cycle begins anew.

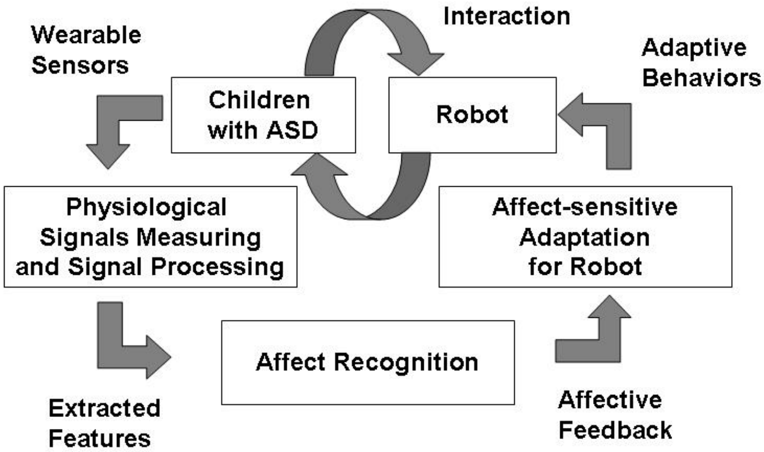


Fig. 1. Framework overview

Human-robot interactions are characterized by explicit as well as implicit channels of communication with presumed underlying affective states [15]. There are several modalities such as facial expression [19], vocal intonation [20], gestures [21], and physiology [22]–[24] that can be utilized to evaluate the affective states. In this work we chose to create affective models based on physiological data for several reasons. Children with ASD often have communicative impairments (both nonverbal and verbal), particularly regarding expression of affective states [1]. These vulnerabilities place limits on traditional conversational and observational methodologies; however, physiological signals are continuously available and are not necessarily directly impacted by these difficulties [25]. As such, physiological modeling may represent a methodology for gathering rich data despite the potential communicative impairments of children with ASD. In addition, physiological data may offer an avenue for recognizing aspects of affect that may be less obvious for humans but more suitable for computers by using signal processing and pattern recognition tools. Furthermore, evidence shows that the transition from one affective state to another state is accompanied by dynamic shifts in indicators of Autonomic Nervous System (ANS) activity [26]. In our previous work, we successfully developed affective models from physiological signals for typical adults with reliable prediction performance [27]–[29][30]. Even though in recent years physiology has been successfully employed to build affect recognizers for typical individuals in several research groups [22]–[24], the studies on the correlation of the physiological signals and the affective states of people with ASD are relatively few [25][30]. To our knowledge real-time physiology-based affect recognition for children with ASD has not been known.

An important question when estimating human affective response is how to operationalize the affective states. Although much existing research on affective modeling categorizes affective states into "basic emotions," there is no consensus on a set of basic emotions among the researchers [31]. This fact implies that pragmatic choices are required to select target affective states for a given application [31]. In this work we chose anxiety, engagement, and liking to be the target affective states. Anxiety was chosen for two primary reasons. First, anxiety plays an important role in various human-machine interaction tasks that can be related to task performance [32]. Second, anxiety is not simply a frequently co-occurring disorder; in some ways it may also be a hallmark of autism [25][33]. Engagement, defined as "sustained attention to an activity or person," has been regarded as one of the key factors for children with ASD to make substantial gains in academic, communication, and social domains [34]. With 'playful' activities during the intervention, the liking of the children (i.e., the enjoyment they experience when interacting with the robot) may create the urge to explore and allow prolonged interaction for the children with ASD, who are susceptible to being withdrawn [8].

Notably, evidence shows that several affective states could co-occur at different arousal levels [35], and different individuals could express the same emotion with different characteristic response patterns under the same contexts (i.e. phenomenon of person stereotypy) [36]. The novelty of the presented affective modeling is that it is individual-specific to accommodate the differences encountered in emotional expression, and it consists of an array of recognizers – each of which determines the intensity of one target affective state for each individual. In this work, a therapist observed the experiments (described in Section III (B-2)) and provided subjective reports based on expertise in inferring presumable underlying affective states from the observable behaviors of children with ASD. The therapist's reports on perceived intensity of the affective states of a child and the extracted physiological indices (described in Section III (B-4)) were employed to develop *therapist-like* affect recognizers that predict high/low levels of anxiety, engagement, and liking for each child with ASD.

Once affective modeling was completed in Phase I, the therapist-like recognizers equipped the robot with the capability to detect the affective states of the children with ASD in real time from on-line extracted physiological features, which could be utilized in future interventions even when a therapist is not available. As stated in [37], it is important to have robots maintain characteristics of adaptability when applied to autism intervention. In Phase II, we designed and implemented a proof-of-concept experiment (robot-based basketball) wherein a robot adapts its behaviors in real time according to the preference of a child with ASD, inferred from the interaction experience and the predicted consequent liking level. This work is the first time, to our knowledge, that the feasibility and the impact of affect-sensitive closed-loop interaction between a robot and a child with ASD have been demonstrated experimentally. While the results are achieved in a non-social interaction task, it is expected that the real-time affect recognition and response system described in this work will provide a basis for future research into developing robot-assisted intervention tools to help children with ASD explore social interaction dynamics in an affect-sensitive and adaptive manner.

III. Experimental Investigation

A. Participants

Six participants within the age range of 13 to 16 years old volunteered to partake in the experiments with the consent of their parents. Each of the participants had a diagnosis on the autism spectrum, either autistic disorder, Asperger's Syndrome, or pervasive developmental disorder not otherwise specified (PDD-NOS), according to their medical records. Due to the nature of the tasks, the following were considered when choosing the participants: (i) having a minimum competency level of age-appropriate language and cognitive skills and (ii) not having any history of mental retardation. Each child with ASD was given the Peabody Picture Vocabulary Test III (PPVT-III) [38] to screen cognitive function. Inclusion in our study was characterized as obtaining a standard score of 80 or above on the PPVT-III measure. Institutional Review Board (IRB) approval was sought and received for conducting experiments. Table I shows the characteristics of the participants in the experiments.

Table I Characteristics of Participants

Child ID	Gender	Age	Diagnosis	PPVT-III Score
A	Male	15	Autistic Disorder	99
B	Male	15	Asperger's Syndrome	80
C	Male	13	Autistic Disorder	81
D	Male	14	PDD-NOS	92
E	Male	16	PDD-NOS	93
F	Female	14	PDD-NOS	83

B. Phase I- Affective Modeling

While the eventual goal is to develop affect-sensitive human-robot interaction, we built the affective models using physiological data gathered from two human-computer interaction tasks. Our previous work [29] showed that affective models built through human-computer interaction tasks could be successfully employed to achieve affect recognition in human-robot interaction for typical individuals. This observation suggests that it is possible to broaden the domain of tasks for affective modeling, thus reducing the habituation effect due to continuous exposure to the same robotic system.

1) Task Design

Two computer-based cognitive tasks were designed to evoke varying intensities of the following three affective states: anxiety, engagement, and liking, from the participants. Physiological data from participants were collected during the experiment. The two tasks consisted of an anagram-solving task and a Pong-playing task. The anagram-solving task has been previously employed to explore relationships between both electrodermal and cardiovascular activity with anxiety [39]. Affective responses were manipulated in this task by presenting the participant with anagrams of varying difficulty levels, as established through pilot work. A long series of trivially easy anagrams caused less engagement. An optimal mix of solvable and difficult anagrams caused liking and engagement at times. Unsolvable or extremely difficult anagrams and giving time deadlines generated anxiety.

The Pong task involved the participant playing a variant of the classic video game "Pong." This game has been used previously by researchers to study anxiety, performance, and gender differences [32]. Various parameters of the game were manipulated to elicit the required affective responses. These parameters included: ball speed and size, paddle speed and size, sluggish or over-responsive keyboard, random keyboard response, and the level of the computer opponent player. Low speeds and large sizes of the ball and paddle made games less engaging after a while; whereas ball and paddle movements at high speeds along with smaller sizes of the two made the game engaging. Very high speeds caused anxiety at times. Playing against a moderate-level computer player usually generated liking. The task configurations were established through pilot work.

Each task sequence was subdivided into a series of discrete trials/epochs that were bounded by the subjective affective state assessments. These assessments were collected using a battery of five questions regarding the three target affective states and the perceived difficulty and performance rated on an eight-point Likert scale where 1 indicated the lowest level and 8 indicated the maximum level. Each participant took part in six sessions – three one-hour sessions of solving anagrams and three one-hour sessions of playing Pong – on six different days.

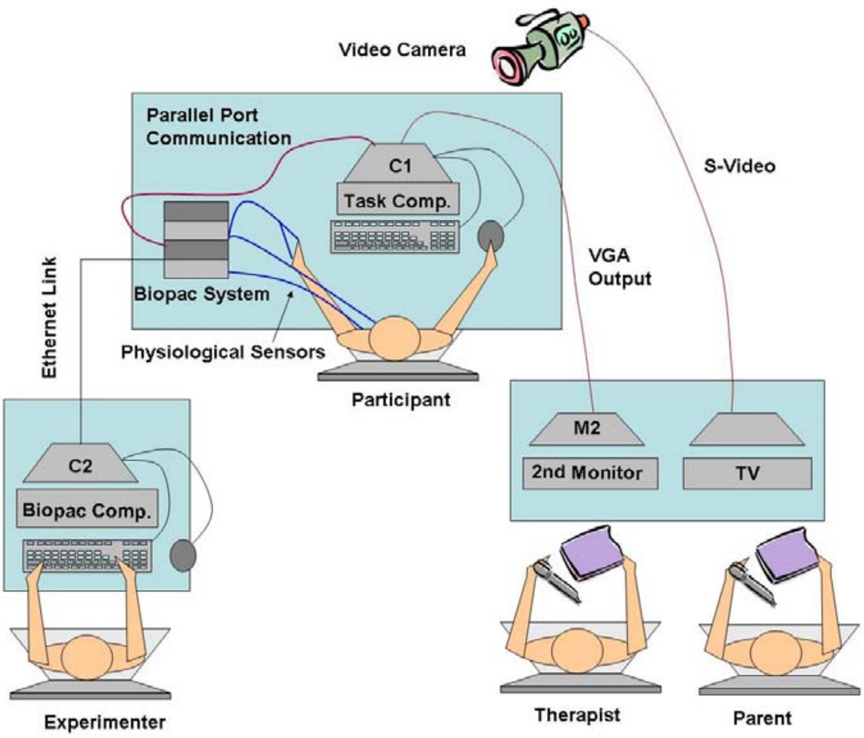


Fig. 2. Experimental set-up for affective modeling tasks

<u>2) Experimental Setup</u>

Fig. 2 shows the set-up for the experiment. A child with ASD was involved in the cognitive tasks on computer C1 while his/her physiological data was acquired via wearable biofeedback sensors and the Biopac system (www.biopac.com). After being amplified and digitized, physiological signals were transferred from the Biopac transducers to C2 through an Ethernet link at 1000Hz and stored. Because of the suspected unreliability of the subjective self-reports from children with ASD, a therapist with experience in working with children with ASD and a parent of the participant were also involved in the study. The signal from the video camera was routed to a television, and the signal from the participant's computer screen where the task was presented was routed to a separate computer monitor M2. The therapist and a parent were seated at the back of the experiment room, watching the experiment from the view of the video camera and observing how the task progressed on the separate monitor.

<u>3) Experimental Procedure</u>

On the first visit, participants completed the PPVT-III measurement to determine eligibility for the experiments. After initial briefing regarding the tasks, physiological sensors from a Biopac system were attached to the participant's body. Participants were asked to relax in a seated position and read age-appropriate leisure material while a three-minute baseline recording was performed, which was later used to offset day-variability. Each session lasted about an hour and consisted of a set (13-15) of either 3-minute epochs for anagram tasks or up to 4-minute epochs for Pong tasks. Each epoch was followed by subjective report questions rated on an eight-point Likert scale. After each epoch, the therapist and the parent also answered the questions about how they thought the participant was feeling during the finished epoch on an eight-point Likert scale. These three sets of reports were used as the possible reference points to link the objective physiological measures to the participant's affective state.

<u>4) Physiological Indices for Affective Modeling</u>

There is good evidence that the physiological activity associated with affective states can be differentiated and systematically organized [26]. Cardiovascular and electromyogram activities have been used to examine positive and negative affective states of people [40][41]. Blood pulse volume amplitude and sympathetic activity have been shown to be associated with task engagement [42]. The relationships between both electrodermal and cardiovascular activities with anxiety were investigated in [39][43]. This correlation between physiological responses and underlying affective states was employed in this work to develop affective models for children with ASD. The physiological signals examined were: features of cardiovascular activity (including inter-beat interval, relative pulse volume, pulse transit time, heart sound, and pre-ejection period), electrodermal activity (tonic and phasic response from skin conductance), and electromyogram (EMG) activity (from corrugator supercilii, zygomaticus, and upper trapezius muscles). These signals were selected because they are likely to demonstrate variability as a function of the targeted affective states, as well as they can be measured non-invasively, and are relatively resistant to movement artifacts [36].

The physiological signals examined in this work along with the features derived from each signal are described in Table II. Signal processing techniques such as Fourier transform, wavelet transform, thresholding, and peak detection, were used to derive the relevant features from the physiological signals. Inter beat interval (IBI) is the time interval between two “R” waves in the electrocardiogram (ECG) waveform. Power spectral analysis is performed on the IBI data to localize the sympathetic and parasympathetic nervous system activities associated with two frequency bands. The high frequency (HF) component (0.15-0.4 Hz; which corresponds to the rate of normal respiration) measures the influence of the vagus nerve in modulating the sinoatrial node and is associated with parasympathetic nervous system activity. The low frequency (LF) component (0.04-0.15 Hz) provides an index of sympathetic effects on the heart. Photoplethysmograph (PPG) signal measures changes in the volume of blood in the finger tip associated with the pulse cycle and provides an index of the relative constriction versus dilation of the blood vessels in the periphery. Pulse transit time (PTT) is the time it takes for the pulse pressure wave to travel from the heart to the periphery. PPT is estimated by computing the time between systole at the heart (as indicated by the R-wave of the ECG) and the peak of the pulse wave reaching the peripheral site where PPG is being measured. The heart sound signal measures sounds generated during each heartbeat. These sounds are produced by blood turbulence primarily due to the closing of the valves within the heart. The features extracted from the heart sound signal consist of the mean and standard deviation of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform. Bioelectrical impedance analysis (BIA) measures the impedance or opposition to the flow of an electric current through the body fluids contained mainly in the lean and fat tissue. A common variable in recent psychophysiology research, pre-ejection period (PEP) measures the latency between the onset of electromechanical systole and the onset of left-ventricular ejection. PEP is derived from impedance cardiogram (ICG) and ECG and is most heavily influenced by sympathetic innervation of the heart. Electrodermal activity consists of two main components - Tonic response and Phasic response. Tonic skin conductance refers to the ongoing or the baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event related changes that occur, caused by a momentary increase in skin conductance (resembling a peak). The EMG signal from Corrugator Supercilii muscle (eyebrow) captures a person's frown and detects the tension in that region. This EMG signal is also a valuable source of blink information and helps determine the blink rate. The EMG signal from the Zygomaticus Major muscle captures the muscle movements while smiling. Upper Trapezius muscle activity measures the tension in the shoulders, one of the most common sites in the body for developing stress.

Table II Physiological Indices

Physiological Signals	Features Derived	Label Used	Unit of Measurement
Cardiac activity	Sympathetic power (from ECG)	Sym	Unit/Square Second
	Parasympathetic power (from ECG)	Para	Unit/Square Second
	Very Low Frequency Power (from ECG)	VLF	Unit/Square Second
	Ratio of powers	Sym Para Para VLF Sym VLF	No unit
	Mean IBI	IBI ECGmean	Milliseconds
	Std. of IBI	IBI ECGstd	Standard Deviation
	Mean amplitude of the peak values of the PPG signal (Photoplethysmogram)	PPG Peakmean	Micro Volts
	Standard deviation (Std.) of the peak values of the PPG signal	PPG Peakstd	Standard Deviation
	Mean Pulse Transit Time	PTTmean	Milliseconds
Heart Sound	Mean of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Mean d3 Mean d4 Mean d5	No unit
	Standard deviation of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Std d3 Std d4 Std d5	No unit
Bioimpedance	Mean Pre-Ejection Period	PEPmean	Milliseconds
	Mean IBI	IBI ICGmean	Milliseconds
Electrodermal activity	Mean tonic activity level	Tonicmean	Micro-Siemens
	Slope of tonic activity	Tonicslope	Micro-Siemens /Second
	Mean amplitude of skin conductance response (phasic activity)	Phasicmean	Micro-Siemens
	Maximum amplitude of skin conductance response (phasic activity)	Phasicmax	Micro-Siemens
	Rate of phasic activity	Phasicrate	Response peaks/Second
Electromyographic activity	Mean of Corrugator Supercilii activity	Cormean	Micro Volts
	Std. of Corrugator Supercilii activity	Corstd	Standard Deviation
	Slope of Corrugator Supercilii activity	Corslope	Micro Volts/Second
	Mean Interbeat Interval of blink activity	IBI Blinkmean	Milliseconds
	Std. of Interbeat Interval of blink activity	IBI Blinkstd	Standard Deviation
	Mean amplitude of blink activity	Amp Blinkmean	Micro Volts
	Standard deviation of blink activity	Blinkstd	Standard Deviation
	Mean of Zygomaticus Major activity	Zygmean	Micro Volts
	Std. of Zygomaticus Major activity	Zygstd	Standard Deviation
	Slope of Zygomaticus Major activity	Zygslope	Micro Volts/Second
	Mean of Upper Trapezius activity	Trapmean	Micro Volts
	Std. of Upper Trapezius activity	Trapstd	Standard Deviation
	Slope of Upper Trapezius activity	Trapslope	Micro Volts/Second
	Mean and Median frequency of Corrugator, Zygomaticus, and Trapezius	Zfreqmean Cfreqmedian Tfreqmean	Hertz
Temperature	Mean temperature	Tempmean	Degree Centigrade
	Slope of temperature	Tempslope	Degree Centigrade/Second
	Std. of temperature	Tempstd	Standard Deviation

5) SVM-based Affective Modeling

Determining the intensity (e.g., high/low) of a particular affective state from the physiological response resembles a classification problem where the attributes are the physiological features and the target function is the degree of arousal. Our earlier work [27] compared the efficacy of several machine learning algorithms to recognize the affective states from the physiological signals of typical individuals and found that Support Vector Machines (SVM) gave the highest classification accuracy. In this work, SVM was employed to determine the underlying affective state of a child with ASD given a set of physiological features. Details of the theory and learning methods of SVM can be found in [44] and are briefly described in Appendix A.

As illustrated in Fig. 3, each participant had a data set comprised of both the objective physiological features and corresponding subjective reports on arousal level of target affective states from the therapist, the parent, and the participant. The physiological features were extracted by using the approaches described in Section III (B-4). Each subjective report was normalized to [0, 1] and then discretized such that 0–0.50 was labeled as low level and 0.51–1 was labeled as high level. All three affective states were partitioned separately so that there were two levels for each affective state. Each data set contained approximately 85 epochs. The multiple subjective reports were analyzed, and one was chosen as the possible reference points to link the physiological measures to the participant's affective state. For example, a therapist-like affect recognizer can be developed when the therapist's reports are used. A SVM-based recognizer was trained on each individual's data set for each target affective state. In this work, in order to deal with the nonlinearly separable data, soft margin classifiers with slack variables were used to find a hyperplane with less restriction [44]. RBF (Radial Basis Function) was selected as the kennel function because it often delivers better performance [45]. A ten-fold cross-validation was used to determine the kernel parameter and regularization parameter of the recognizer.

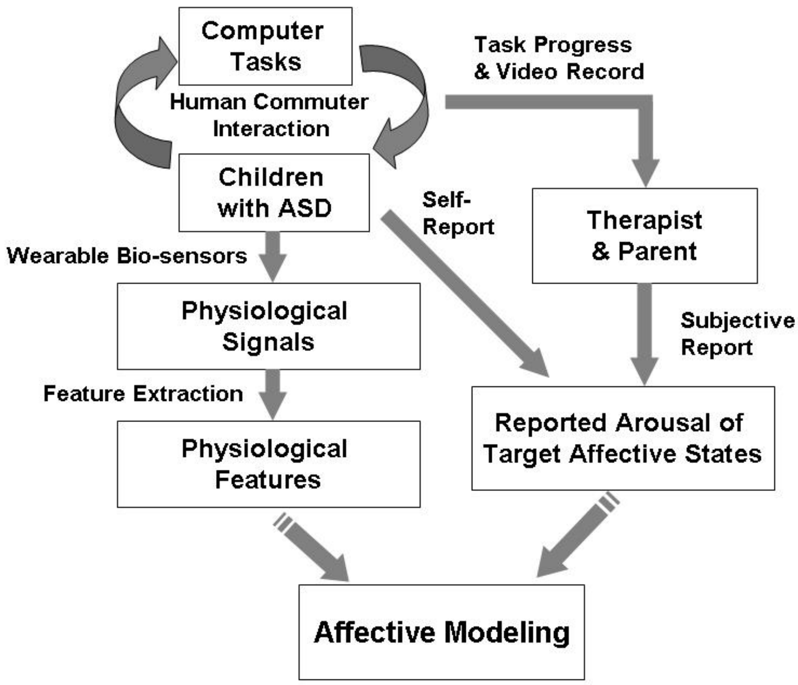


Fig. 3. Overview of affective modeling

Once affective modeling is accomplished, the affect recognizers can accept as input the physiological features extracted on-line and produce as output the probable level of the target affective state of a child with ASD while interacting with a robot. In the design for the human-robot interaction task in Phase II, adequate measures were taken to avoid physical effort from overwhelming the physiological response.

C. Phase II- Closed-Loop Human Robot Interaction

1) Task Design

A closed-loop human robot interaction task, “robot-based basketball (RBB),” was designed. The main objective was two-fold: (i) to enable the robot to learn the preference of the children with ASD implicitly using physiology-based affective models as well as select appropriate behaviors accordingly; and (ii) to observe the effects of such affective-sensitivity in the closed-loop interaction between the children with ASD and the robot.

The affective model developed in Phase I is capable of predicting the intensity of liking, anxiety, and engagement simultaneously. However to designate a specific objective for the experiment in Phase II without compromising its proof-of-concept purpose, one of the three target affective states was chosen to be detected and responded to by the robot in real time. As has been emphasized in [8], the liking of the children (i.e., the enjoyment they experience when interacting with a robot) is a goal as desirable as skill learning for autism intervention. Therefore, liking was chosen as the affective state around which to modify the robot’s behaviors in Phase II.

In the RBB task, an undersized basketball hoop was attached to the end-effector of a robotic manipulator, which could move the hoop in different directions (as shown in Fig. 4) with different speeds. The children were instructed to shoot a required number of baskets into the moving hoop within a given time. Three robot behaviors were designed as shown in Table III. For example, in behavior 1 the robot moves towards and away from the participant (i.e., in the X direction) at a slow speed with soft background music, and the shooting requirement for successful baskets is relatively low. The parameter configurations were determined based on a pilot study to attain varied impacts on affective experience for different behaviors. From this pilot study, the averaged performance of participants for a given behavior was compiled and analyzed. The threshold of shooting requirement (*TSR*) was defined as 10% lower than the average performance. At the end of each epoch, the participant's performance was rated as excellent ($baskets \geq \lfloor 1.2TSR \rfloor$), above average ($\lfloor 0.8TSR \rfloor \leq baskets < \lfloor 1.2TSR \rfloor$), or below average ($baskets < \lfloor 0.8TSR \rfloor$). Behavior transitions occurred between but not within epochs. As such, each robot behavior extended for the length of an epoch (1.5 minutes in duration) to have the participant fully exposed to the impact of that behavior.

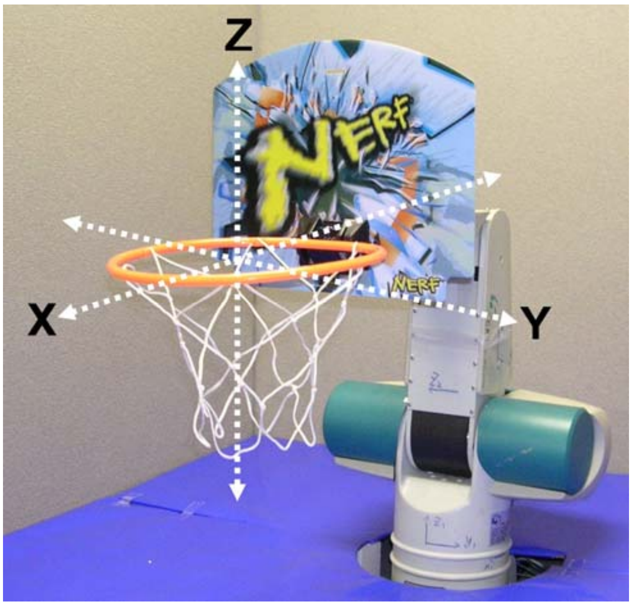


Fig. 4. X, Y, and Z directions for behaviors used in RBB

Table III Robot Behaviors

Behavior ID	Motion Direction	Speed (sec/period)	Threshold (shots/epoch)	Background Music
1	X	8	12	Serene
2	Y	4	20	Lively
3	Z	2	30	Irregular

Each of the six participants took part in two robot basketball sessions (RBB1 and RBB2). In RBB1 (non-affect based) the robot selected its behavior randomly (i.e., without any regard to the liking information of the participant), and the presentation of each type of behavior was evenly distributed. This session was designed for two purposes: (i) to explore the state space and action space of the QV-learning algorithm used in RBB2 for behavior adaptation (described in Section III (C-4)); and (ii) to validate that the different robot behaviors have distinguishable impact on the child's level of liking. In RBB2 (liking-based), the robot continues to learn the child's individual preference and selects the desirable behavior based on interaction experiences (i.e., records of robot behavior and the consequent liking level of a participant predicted by the affective model). The idea is to investigate whether the robot can automatically choose the most-liked behavior of each participant as observed from RBB1 by means of physiology-based affective model and QV-learning.

2) Experimental Setup

The real-time implementation of the robot-based basketball system is shown in Fig. 5. The set-up included a 5 degrees-of-freedom robot manipulator (CRS Catalyst-5 System) with a small basketball hoop attached to its end-effector. Two sets of infrared (IR) transmitter and receiver pairs were attached to the hoop to detect small, soft foam balls going through the hoop. The set-up also included the biological feedback equipment (Biopac system) that collected the participant's physiological signals and the digital output from the IR sensors. The Biopac system was connected to a PC (C1) that: (i) acquired physiological signals from the Biopac system and extracted physiological features on-line, (ii) predicted the probable liking level by using the affective model developed in Phase I, (iii) acquired IR data through the analog input channels of the Biopac system, (iv) ran a QV-learning algorithm that learns the participant's preference and chooses the robot's next behavior accordingly. Computer C1 was connected serially to the CRS computer (C2), which ran Simulink software. The behavior switch triggers were transmitted from C1 to C2 via a RS232 link. The commands to control the robot's various joints were transmitted from C2 to the robot. There was a communication protocol established between C1 and C2 that ensured the beginning and end of the basketball task was appropriately synchronized with the physiological data acquisition on C1. As in Phase I tasks, the therapist and a parent were also involved, watching the experiment from the TV that was connected to a video camera.

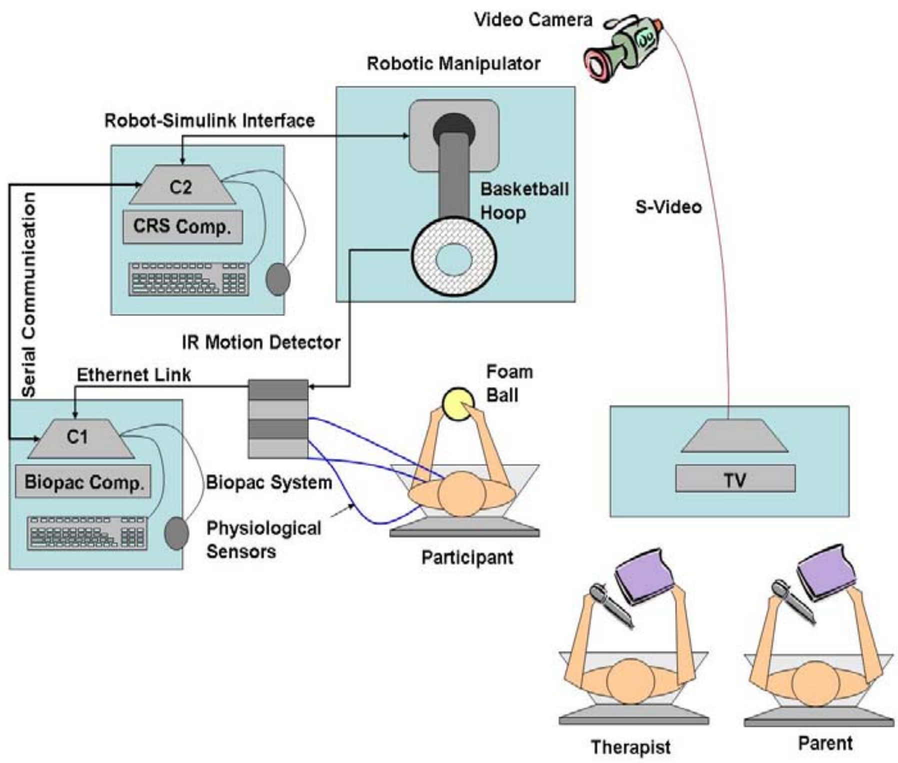


Fig. 5. Experimental set-up for robot basketball

3) Experimental Procedure

Each basketball session (RBB1 or RBB2) was approximately 1 hour long and included 27 minutes of active human-robot interaction (i.e., 18 epochs of 1.5 minutes each). The remaining time was spent attaching sensors, guiding a short practice, taking a baseline recording, collecting subjective reports, and pausing for scheduled breaks. During the experiment, the participant was asked to take a break after every four epochs and the participant could request a break whenever he/she desired one. During each basketball epoch, the participant received commands and performance assessments from pre-recorded dialogue via a speech program running on C1 and the interaction proceeded as follows:

1. The participant was notified of the goal (i.e., *TSR*).
2. A start command instructed the participant to start shooting baskets.
3. Once the epoch started, the participant was given voice feedback every 30 seconds regarding the number of baskets remaining and the time available.
4. A stop command instructed the participant to stop shooting baskets, which ended the epoch.
5. At the end of each epoch, the participant's performance was rated and relayed to him/her as excellent, above average, or below average.

Each epoch was followed by subjective reports that took 30-60 seconds to collect. The subjective assessment procedure was the same as the protocol used in the affective modeling tasks in Phase I. After the subjective report was complete, the next epoch would begin. To prevent habituation, a time interval of at least 7 days between any two RBB sessions was enforced.

4) Affect-sensitive Behavior Adaptation in Closed-Loop Human Robot Interaction

We defined the state, action, state transition, and reward functions so that the affect-sensitive robot behavior adaptation problem could be solved using the QV-learning algorithm as described in [46] and Appendix B.

The set of states consisted of three robot behaviors as described in Table III. In every state, the robot has three possible actions (1/2/3) that correspond to choosing behavior 1, 2, or 3, respectively, for the next time step (i.e., next epoch). Each robot behavior persists for one full epoch and the state/behavior transition occurs only at the end of an epoch. The detection of consequent affective cues (i.e., the real-time prediction of the liking level for the next epoch) was used to evaluate the desirability of a certain action. To have the robot adapt to a child's individual preference, a reward function was defined based on the predicted liking level. If the consequent liking level was recognized as high, the contributing action was interpreted as positive and a reward was granted ($r = 1$); otherwise the robot received a punishment ($r = -1$). QV-learning uses this reward function to have the robot learn how to select the behavior that was expected to result in a high liking level and therefore positively influenced the actual affective (e.g., liking) experience of the child.

RBB1 enables state and action exploration where the behavior-switching actions are chosen randomly, with the number of visits to each state evenly distributed. The V-function and Q-function are updated using Eq. (3) and Eq. (4) from Appendix B. After RBB1, the subjective reports are analyzed to examine the impacts of different behaviors on each participant's preference. In RBB2 the robot starts from a non-preferred behavior/state and continues the learning process by using Eq. (3) and Eq. (4). A greedy action selection mechanism is used to choose the behavior-switching action with the highest Q-value.

Because of the limited number of states and actions in this proof-of-concept experiment, tabular representation is used for the V-function and the Q-function. To prevent a certain action and/or state from being overly dominant and to counteract the habituation effect, the values of $Q(s, a)$ and $V(s)$ are bounded by using the reward or punishment encountered in the interaction. The parameters in Eq. (3) and Eq. (4) are chosen as $\alpha = 0.8$ and $\gamma = 0.9$. Before RBB1 begins, the initial values in the V-table and the Q-table are set to 0.

IV. Results and Discussion

In this section we present both the Phase I results of physiology-based affective modeling for children with ASD and Phase II results of the affect-sensitive closed-loop interaction between children with ASD and the robot.

A. Affect Detection

Due to the unresolved debate on the definition of emotion (e.g., objective entities or socially constructed labels), researchers in affective computing often face difficulties obtaining the ground truth to label the natural emotion data accordingly. As suggested in [31][47], the immediate implication of such a controversy is that pragmatic choices (e.g., application- and user-profiled choices) must be made to develop an automatic affect recognizer. While there have been some criticisms on the use of subjective report (i.e., self-assessment or the reports collected from observers) and its effect on possibly forcing the determination of emotions, the subjective report is generally regarded as an effective way to evaluate the affective responses. As a result, subjective report is widely used for affective modeling and for endowing an intelligent system with the recognition abilities similar to those of the reporters [15][21]. One of the prime challenges of this work is to attain reliable subjective reports. Researchers are generally reluctant to trust the response of adolescents on self-report [48]. In this study, one should be especially wary of the dependability of self-reports from children with ASD, who may have deficits in processing (i.e., identifying and describing) their own emotions [49]. Therefore, in order to overcome this difficulty, reports on how a therapist and a parent thought the participant was feeling based on his/her observed behaviors were collected after each epoch.

To measure the amount of agreement among the different reporters, the kappa statistic was used [50]. The kappa coefficient (K) measures pair-wise agreement among a set of reporters making category judgments, correcting for expected chance agreement. When there is complete agreement, then K=1; whereas, when there is no agreement other than that which would be expected by chance, then $K = 0$.

It was observed that the agreement between the therapist and parent showed the largest kappa statistic values (mean = 0.62) among the three possible pairs for each child ($p < 0.05$, paired t-test). The mean of the kappa statistic values between the children and either the therapist or the parent were relatively small (0.37 and 0.40, respectively). Lack of agreement with adults does not necessarily mean that the self-reports of children with ASD are not dependable; however, given the objective of this study is to develop an affect-sensitive robotic system for autism intervention where the therapists' judgement based on their expertise is the state-of-the-art and the fact that there is a reasonably high agreement between the therapist and the parents for all of the six children, the subjective reports of the therapist were used as the reference points linking the objective physiological data to the children's affective state. To make the subjective reports more consistent, the same therapist was involved in all of the experiments. This choice allowed for building a *therapist-like* affective model. In the rest of the paper, unless otherwise specified, we will use the term liking, anxiety, and engagement to imply the target affective states as discerned by the therapist.

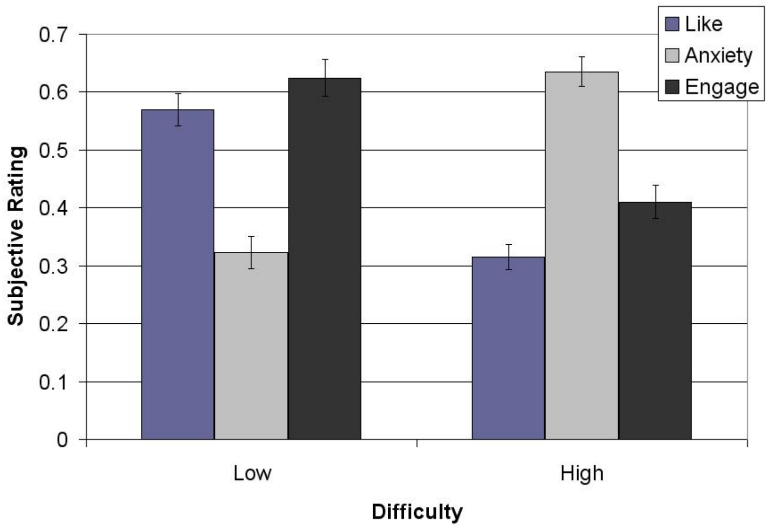


Fig. 6. Rated average affect response from therapist's reports

Fig. 6 shows a comparison of the therapist's average ratings for liking, anxiety, and engagement when the children with ASD played easy or difficult epochs in the Phase I computer games. When averaged across all participants, liking decreased, anxiety increased, and engagement decreased with increasing task difficulty. Table IV shows the correlation analysis between the reported affective states and the task difficulty. For each set of variables, the probability value (p-value) was computed from a two side t-test. Due to the large sample size (approximately 85 epochs for each participant), the p-value for all correlations was less than 0.005. Through point biserial correlation analysis, it was found that difficulty is strongly positively correlated with anxiety and negatively correlated with liking and engagement. By examining Pearson correlation coefficients, it was observed that there is strong positive correlation between liking and engagement and negative correlation between liking and anxiety, and there also exists a weak correlation between the reported anxiety and engagement. The results in Fig. 6 and Table IV present the results across all the children. However, when each child is examined individually, different trends could arise. For example, for Child A, anxiety is positively correlated with engagement (Pearson correlation =0.45); for Child F, no significant correlation is observed (Pearson correlation = -0.15, $p > 0.05$); while for the four other children (B, C, D, and E) anxiety negatively correlated with engagement (Pearson correlation equals -0.50, -0.39, -0.61, and -0.58, respectively), which revealed diverse affective characteristics of the children with ASD.

Table IV Results of Correlation Analysis from Therapist's reports

	Anxiety	Engage	Difficulty
Liking	-0.521	0.885	-0.616
Anxiety		-0.401	0.731
Engage			-0.486

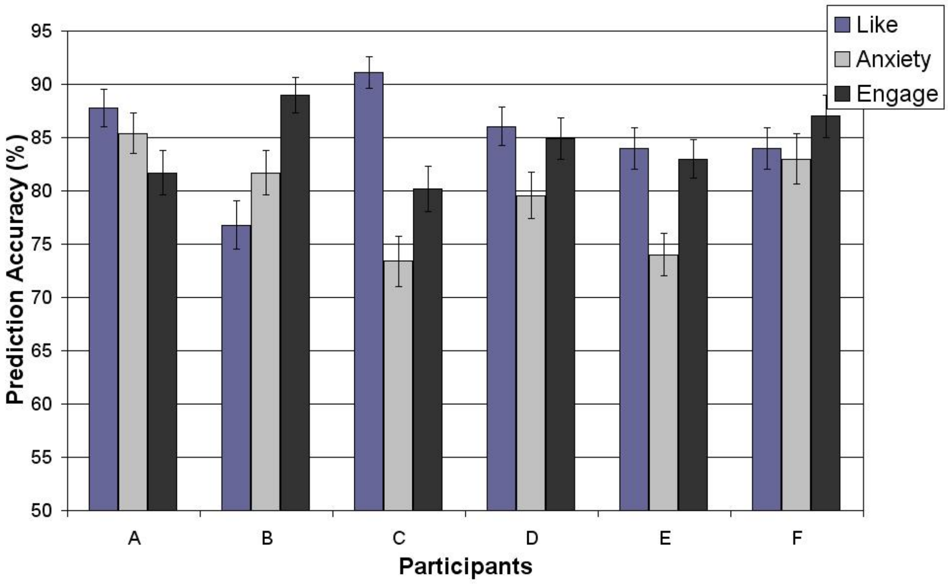


Fig. 7. Classification Accuracy of the Affect Recognizer

The performance of the developed affective model for each child is shown in Fig. 7. The cross-validation method, 'leave-one-out', was used. The affective model produced high recognition accuracies for each target affective state of each participant. The average correct prediction accuracies across all participants were: 85.0% for liking, 79.5% for anxiety, and 84.3% for engagement. This was promising considering that this task was challenging in two respects: (i) the reports were collected from the therapist who was observing the children with ASD as opposed to having typical adults capable of differentiating and reporting their own affective states, and (ii) varying levels of arousal of any given affective state (e.g., low/high anxiety) were identified instead of determining discrete emotions (e.g., anger, joy, sadness, etc.). Determining the difference in arousal level in one affective state is more subtle than distinguishing between two discrete affective states. In order to explore the effects of reducing the number of physiological signals and the possibility of achieving more economical modeling, we examined the performance of the affect recognizers when cardiovascular, electrodermal [2], and electromyographic (EMG) activities and their combinations were used for affective modelling separately. It was observed that EMG signal is less discriminatory than cardiovascular and electrodermal activities (with prediction accuracy of 69.7%, 73.5%, and 73.0%, respectively). While no combination surpassed the prediction accuracy achieved when all signals were used (82.9%), the results suggested it may be possible to selectively reduce the set of signals and obtain nearly-as-good performance (e.g., using a combination of cardiovascular and electrodermal signals yielded 78.7% prediction accuracy).

B. Affect Adaptation in Robot-based Basketball Task

The six children with ASD who completed the Phase I experiments also took part in the robot basketball task. The results described here are based on the RBB1 (non-affect based) and RBB2 (liking-based) tasks.

First, we present results to validate that different behaviors of the robot had distinguishable impacts on the liking level of the children with ASD. To reduce the bias of validation, in RBB1 the robot selects behaviors randomly and the occurrence of each behavior is evenly distributed. Fig. 8 shows the average labeled liking level for each behavior as reported by the therapist in RBB1. The difference of the impact is significant for five children (participants A, B, D, E, and F) and moderate for participant C. By performing two-way ANOVA analysis on the behavior (i.e., most-preferred, moderately-

[2] Peripheral temperature has relatively few features derived as shown in Table II and was not examined independently. Instead, it was studied conjunctively with the electrodermal activity, both of which were acquired from the non-dominant hand of a participant.

preferred, and least-preferred behavior) and participant, it was found that the differences of reported liking for different behaviors are statistically significant ($p < 0.05$), whereas no significant effect due to different participants was observed. Furthermore, it was also observed that different children with ASD may have different preferences for the robot's behaviors. These results demonstrated that it is important to have a robot learn the individual's preference and adapt to it automatically, which may allow a more tailored and affect-sensitive interaction between children with ASD and the robot. For example, when a robot learns that a certain behavior is liked more by a particular child, it can choose that behavior as his/her 'social feedback' or 'reinforcer' in robot-assisted autism intervention. Playful interaction will be more likely to emerge by addressing a child's preference.

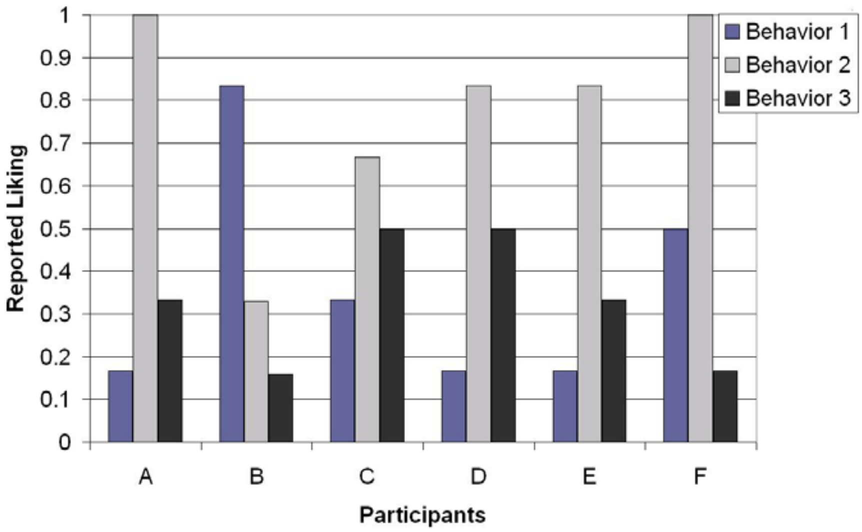


Fig. 8. Mean liking level for different behaviors in RBB1

Second, the predictive accuracy of how closely the real-time physiology-based quantitative measures of liking, as obtained from affective models developed in Phase I, matched with that of the subjective rating of liking made by the therapist during Phase II is presented in Fig. 9. The average predictive accuracy across all the participants was approximately 81.1%. The highest was 86.1% for Child D, and the lowest was 77.8% for Child B and Child E. Note that the affective model was evaluated based on physiological data obtained on-line from a real-time application for children with ASD. However, this prediction accuracy is comparable to the results achieved through off-line analysis for typical individuals [22][27].

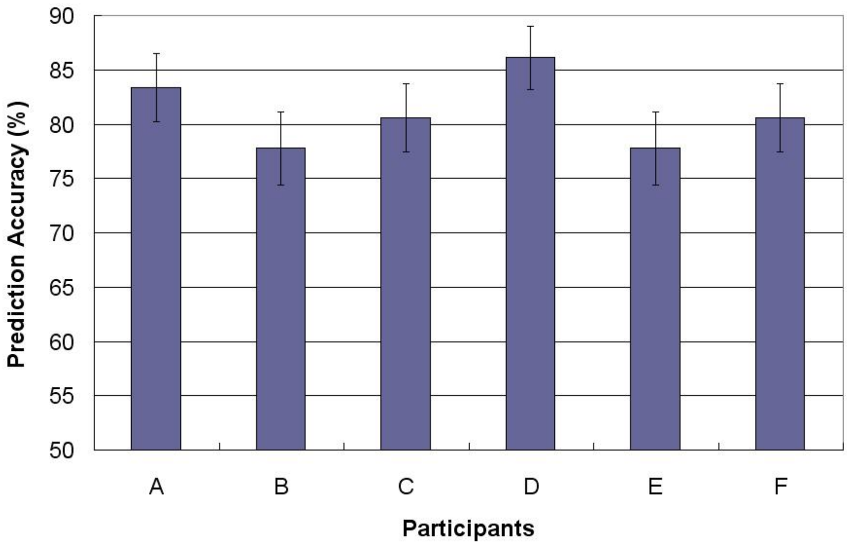


Fig. 9. Real-time classification accuracy of liking

Third, we present results about robot behavior adaptation and investigate its impact on the interaction between the children with ASD and the robot. Table V shows the percentages of different behaviors that were chosen in the RBB2 session for each participant. The robot learned the individual's preference and selected the most-preferred behavior with high probability for all the participants. Averaged across all participants, the most-preferred, moderately-preferred, and least-preferred behaviors were chosen

72.5%, 16.7%, and 10.8% of the time, respectively. The preference of a behavior was defined by the reported liking level in RBB1 as shown in Fig. 8. To understand these results more clearly we describe an individual case. Fig. 10 shows the affect-sensitive behavior adaptation in RBB2 for Child A, who prefers behavior 2 most (refer to Fig. 8). The real-time predicted liking level (i.e., high/low) is denoted by 'H' or 'L'. The robot starts in a non-preferred behavior (behavior 1) and then explores other behaviors before settling on the most-preferred behavior (behavior 2) where the liking level is the high (as confirmed by the affective model prediction as well as the therapist's subjective report). After a considerable time interacting with behavior 2 (e.g., epoch 7), the participant appears to not enjoy this behavior as much as before. The affective model detects this change and returns negative rewards. The QV-learning algorithm updates its state/action space and directs the robot to switch behaviors. However, after exploring other behaviors, the robot eventually finds that behavior 2 is the most-preferred by Child A (e.g., epoch 11) and continues the interaction using this behavior. At epoch 16, even though the predicted liking level is low, due to high frequent positive rewards received for behavior 2, the robot checks the updated Q function and remains at this behavior. There could be several reasons why less-preferred behaviors were chosen in RBB2. The learned behavior selection policy might not have been optimal after the exploration in RBB1, and the QV-learning algorithm continued the learning process in RBB2. Another reason could be that the affective model is not 100% accurate and may return false reward/punishment, which may have given the robot imperfect instruction for behavior switches. Habituation to the most-preferred behavior during RBB2 could also be a factor that might have contributed to temporary changes in preference which led the robot to choose other behaviors.

Table V Proportion of Different Behaviors Performed in RBB2

Child ID	Most-Liked Behavior		Moderate-Liked Behavior		Least-Liked Behavior	
	ID	Proportion	ID	Proportion	ID	Proportion
A	2	82.4%	3	11.8%	1	5.8%
B	1	70.6%	2	17.7%	3	11.7%
C	2	58.8%	3	23.5%	1	17.7%
D	2	76.5%	3	11.8%	1	11.7%
E	2	76.5%	3	17.6%	1	5.9%
F	2	70.6%	1	17.7%	3	11.7%

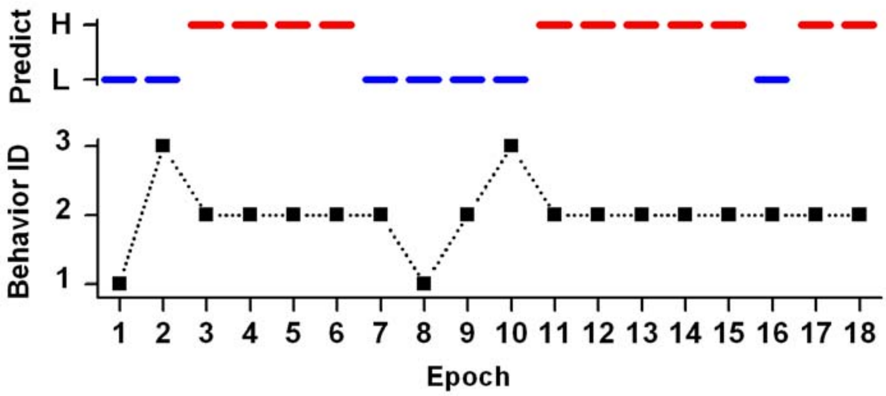


Fig. 10. Behavior selected by affect-sensitive robot in RBB2 for Child A

In Table V, the robot chose a less-preferred behavior more often for Child C than for other participants. As can be seen in Fig. 8, Child C does not show differences of liking among the three behaviors as significantly as the other children. This instance of less-distinct preference could result in more inconsistent rewards/punishments and the robot switching behaviors more frequently. However, despite the above possible reasons for choosing less-preferred behaviors, Table V and Fig. 10 show that the robot is capable of identifying and selecting the preferred behavior automatically in most of the epochs for all participants and thus positively influencing the subjective liking level of the children with ASD (as shown in Fig. 11).

In Fig. 11, we present results to demonstrate that active monitoring of participants' liking and automatically selecting the preferred behavior allowed children with ASD to maintain high liking levels. The average labeled liking levels of the participants as reported by the therapist during the two sessions were compared. The agreement between the therapist and parent on the subjective liking level is substantial for both RBB sessions and has a larger kappa statistic value (0.71) than that of the other two possible reporter pairs (0.39 for the therapist and children and 0.43 for the parent and children). The lighter bars in Fig. 11 indicate the liking level during the RBB1 session (i.e., when the robot selected behaviors randomly), and the darker bars show the liking level during the RBB2 session (i.e., when robot learned the individual preference and chose the appropriate behavior accordingly). It can be seen that for all participants liking level was maintained, and for five of the six children liking level increased. There was no significant increase for Child C during the liking-based session as compared to the non-affect based session. As mentioned earlier, the impact of the different robot behaviors on the liking level of Child C is not as significant as that of the others, which may impede the robot in finding the preferred behavior and hence impede the robot in effectively influencing the subjective liking level positively. Note that RBB1 presents a typically balanced interaction with equal numbers of most-preferred, moderately-preferred, and least-preferred epochs and the comparisons in Fig. 11 are not between liking-based sessions and sessions of least-preferred epochs. In order to determine the effects of the session type and participant on the reported liking, a two-way ANOVA test was performed. The null hypothesis that there is no change in liking level between liking-based sessions and non-affect based sessions could be rejected at the 99.5% confidence level. Additionally, no significant impact due to different participants was observed. This was an important result as the robot continued learning and utilizing the information regarding the probable liking level of children with ASD to adjust its behaviors. This ability enables the robot to adapt its behavior selection policy in real time and hence keep the participant in a higher liking level.

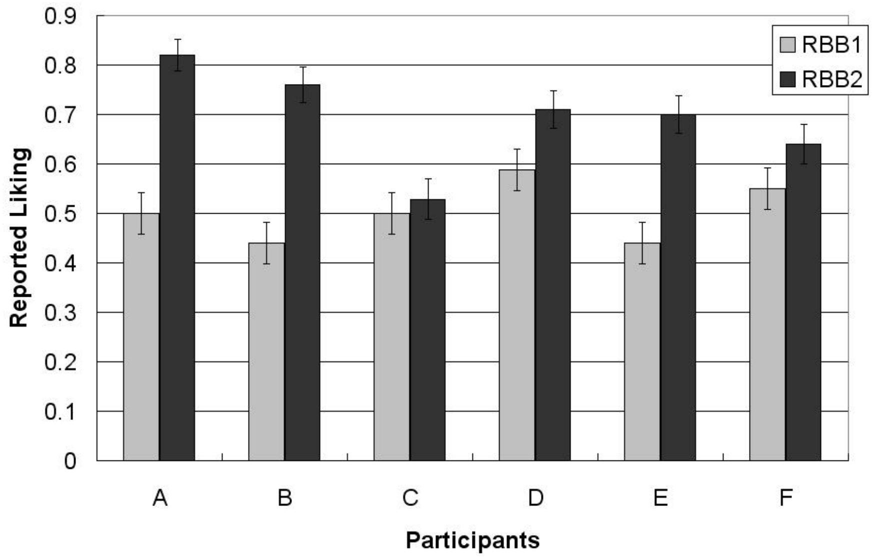


Fig. 11. Subjective liking as reported by therapist

V. Conclusions and Future Work

There is increasing consensus in the autism community that development of assistive tools that exploit advanced technology will likely make application of intensive intervention for children with ASD more readily accessible. In recent years, robotic technology has been investigated in order to facilitate and/or partially automate the existing behavioral intervention that addresses specific deficits associated with autism. However, the current robot-assisted intervention tools for children with ASD do not possess the ability to decipher affective cues from the children, which could be critical given that the affective factors of children with ASD have significant impacts on the intervention practice. In this work, we have proposed a novel framework for affect-sensitive human-robot interaction where the robot can implicitly detect the affective states of the children with ASD as discerned by the therapist and respond to it accordingly.

The presented affective modeling methodology could allow the recognition of affective states of children with ASD from physiological signals in real time and provide the basis for future robot-assisted affect-sensitive interactive autism intervention. In Phase I, two cognitive tasks – solving anagrams and playing Pong – have been designed to elicit the affective states of liking, anxiety, and engagement for children with ASD. To have reliable reference points to link the physiological data to the affective states, the reports from the child, the therapist, and the parent were collected and analyzed. A large set of physiological indices have been investigated to determine their correlation with the affective states of the children with ASD. We have experimentally demonstrated that it is viable to detect the affective states of children with ASD via a physiology-based affect recognition mechanism. A SVM-based affective model yielded reliable prediction with a success rate of 82.9% when using the therapist's reports.

In order to investigate the affect-sensitive closed-loop interaction between the children with ASD and the robot, we designed a proof-of-concept task, robot-based basketball, and developed an experimental system for its real-time implementation and verification. The real-time prediction of liking level of the children with ASD was accomplished with an average accuracy of 81.1%. The robot learned individual preferences of the children with ASD over time based on the interaction experience and the predicted liking level and hence automatically selected the most-preferred behavior, on average, 72.5% of the time. We have observed that such affect-sensitive robot behavior adaptation has led to an increase in reported liking level of the children with ASD. This is the first time, to our knowledge, that the affective states of children with ASD have been detected via a physiology-based affect recognition technique in real time. This is also the first time that the impact of affect-sensitive closed-loop interaction between a robot and children with ASD has been demonstrated experimentally.

In order to account for the phenomenon of person-stereotypy and the diverse affective characteristics of the children with ASD, we employed an individual-specific approach for affective modeling. An intensive study was preformed based on a large sample size of observations (approximately 85 epochs over 6 hours) for each of the six children with ASD. The time spent collecting the training data for affective modeling can be justified by the current ASD intervention practices [5]. However, note that the methodology for inducing, gathering, and modeling the experimental data is not dependent on the participants. The consistently reliable prediction accuracy for each participant demonstrated that it was feasible to model the affective states of children with ASD via psychophysiological analysis.

The presented work requires physiological sensing that has its own limitations. For example, one needs to wear physiological sensors, and use of such sensors could be restrictive under certain circumstances. But given the rapid progress in wearable computing with small, non-invasive sensors and wireless communication, e.g., physiological sensing clothing and accessories [51], we believe that physiology-based affect recognition can be appropriate and useful for the application of interactive autism intervention. None of the participants in this study had any objection to wearing the physiological sensors.

Future work will involve designing socially-directed interaction experiments with robots interacting with children with ASD. Specifically, we plan to integrate the real-time affect recognition and response system described here with a life-like android face developed by Hanson Robotics (www.hansonrobotics.com), which can produce accurate examples of common facial expressions that convey affective states. This affective information could be used as feedback for empathy exercises to help children recognize their own emotions. Enhancements on the intervention process could also be envisioned. For instance, the robot could exhibit interesting behaviors to retain the child's attention when it detects his/her liking level is low. Additionally, besides liking, anxiety and engagement are also considered important in autism intervention practice (as described in Sections I and II). For example, anxiety is considered "as both a possible consequence of, and a possible cause of, aspects of the behavior of children with autism [33]." While the affective model developed in this work is capable of predicting the intensity of liking, anxiety, and engagement simultaneously, more sophisticated behavior adaptation mechanisms would be demanded to incorporate multiple inferred affective cues and account for other intervention information of interests, such as the intervention goals, historical records, and contextual inputs. We will investigate fast and robust learning mechanisms that would permit a robot's adaptive response in the more complex human-robot interaction tasks and allow the affect-sensitive robot to be adopted in the future autism intervention.

Appendix

A. Pattern Recognition using Support Vector Machines

SVM, pioneered by Vapnik [44], is an excellent tool for classification [45]. Its appeal lies in its strong association with statistical learning theory as it approximates the structural risk minimization principle. Good generalization performance can be achieved by maximizing the margin, where margin is defined as the sum of the distances of the hyperplane from the nearest data points of each of the two classes. SVM is a linear machine working in a high k-dimensional feature space formed by an implicit embedding of n-dimensional input data X into a k-dimensional feature space ($k > n$) through the use of a nonlinear mapping $\phi(X)$. This allows for the use of linear algebra and geometry to separate the data, which is normally only separable with nonlinear rules in the input space. The problem of finding a linear classifier for given data points with known class labels can be described as finding a separating hyperplane $W^T\varphi\left(X\right)$ that satisfies:

$$y_i\left(W^T\varphi\left(X_i\right)\right) = y_i\left(\sum_{j=1}^{k} w_j\phi_j(X_i) + w_0\right) \geq 1 - \xi_i \tag{1}$$

where N represents the number of training data pairs (X_i, y_i) indexed by $i = 1,2,\ldots, N$; $y_i \in \{+1, -1\}$ represents the class label; $\varphi(X) = [\phi_0(X), \phi_1(X),\ldots, \phi_k(X)]^{\mathrm{T}}$ is the mapped feature vector ($\phi_0(X) = 1$); and $W = [w_0, w_1,\ldots, w_k]$ is the weight vector of the network. The nonnegative slack variable ξ_i generalizes the linear classifier with soft margin to deal with nonlinearly separable problems.

All operations in learning and testing modes are done in SVM using a so-called kernel function defined as $K(X_i, X) = \varphi^T(X_i)\varphi(X)$ [44]. The kernel function allows for efficient computation of inner products directly in the feature space and circumvents the difficulty of specifying the non-linear mapping explicitly. One distinctive fact about SVM is that the learning task is reduced to a dual quadratic programming problem by introducing the Lagrange multipliers α_I [44]:

$$\textit{Maximize} \qquad Q(\alpha)=\sum\nolimits_{i=1}^{N}\alpha_i-\frac{1}{2}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N}\alpha_i\alpha_j y_i y_j K\left(X_i,X_j\right) \qquad (2)$$

$$\textit{Subject to} \qquad \sum\nolimits_{i=1}^{N}\alpha_i y_i=0 \quad \text{and} \quad 0\le\alpha_i\le C$$

where C is a user-defined regularization parameter that determines the balance between the complexity of the network characterized by the weight vector W and the error of classification of data. The corresponding αi multipliers are only non-zero for the support vectors (i.e., the training points nearest to the hyperplane).

The SVM approach is able to deal with noisy data and over-fitting by allowing for some misclassifications on the training set [45]. This characteristic makes it particularly suitable for affect recognition because the physiology data is noisy and the training set size is often small. Another important feature of SVM is that the quadratic programming leads in all cases to the global minimum of the cost function. With the kernel representation, SVM provides an efficient technique that can tackle the difficult, high dimensional affect recognition problem.

B. Behavior Adaptation using QV Learning

QV-learning [46], a variant of the standard reinforcement learning algorithm Q-learning [52], was applied to achieve the affect-sensitive behavior adaptation. QV-learning keeps track of both a Q-function and a V-function. The Q-function represents the utility value *Q(s, a)* for every possible pair of state *s* and action *a*. The *V*-function indicates the utility value *V(s)* for each state *s*. The state value $V(s_t)$ and Q-value $Q(s_t, a_t)$

at step t are updated after each experience (s_t, a_t, r_t, s_{t+1}) by:

$$V(s_t) := V(s_t) + \alpha\left(r_t + \gamma V(s_{t+1}) - V(s_t)\right) \tag{3}$$

$$Q(s_t, a_t) := Q(s_t, a_t) + \alpha\left(r_t + \gamma V(s_{t+1}) - Q(s_t, a_t)\right) \tag{4}$$

where r_t is the received reward that measures the desirability of the action a_t when it is applied on state s_t and causes the system to evolve to state s_{t+1}. The difference between (4) and the conventional Q-learning rule is that QV-learning uses V-values learned in (3) and is not defined solely in terms of Q-values. Since $V(s)$ is updated more often than $Q(s, a)$, QV-learning may permit a fast learning process [46] and enable the robot to efficiently find a behavior selection policy during human-robot interaction.

Acknowledgements

The authors gratefully acknowledge the MARI grant, the Vanderbilt Treatment and Research Institute for Autism Spectrum Disorders for guidance during the development of experiments involving children with ASD, and the parents and children who participated in the presented research.

References

[1] American Psychiatric Association., *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*, 4th ed. Washington, DC: American Psychiatric Association, 2000.

[2] CDC, "Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, 14 sites, United States, 2002," *MMWR Surveill Summ,* vol. 56, pp. 12-28, Feb 9 2007.

[3] S. J. Rogers, "Empirically supported comprehensive treatments for young children with autism," *J Clin Child Psychol,* vol. 27, pp. 168-79, Jun 1998.

[4] M. Rutter, "Autism: its recognition, early diagnosis, and service implications," *J Dev Behav Pediatr,* vol. 27, pp. S54-8, Apr 2006.

[5] L. Tarkan, "Autism therapy is called effective, but rare," in *New York Times*, 2002.

[6] V. Bernard-Opitz, N. Sriram, and S. Nakhoda-Sapuan, "Enhancing social problem solving in children with autism and normal children through computer-assisted instruction," *J Autism Dev Disord,* vol. 31, pp. 377-84, Aug 2001.

[7] S. Parsons and P. Mitchell, "The potential of virtual reality in social skills training for people with autistic spectrum disorders," *J Intellect Disabil Res,* vol. 46, pp. 430-43, Jun 2002.

[8] K. Dautenhahn and I. Werry, "Towards interactive robots in autism therapy: background, motivation and challenges," *Pragmatics & Cognition,* vol. 12, pp. 1-35, 2004.

[9] F. Michaud and C. Theberge-Turmel, "Mobile robotic toys and autism," in Socially Intelligent Agents: Creating Relationships with Computers and Robots, K. Dautenhahn, A. H. Bond, L. Canamero, and B. Edmonds, Eds.: Kluwer Academic Publishers, 2002, pp. 125-132.

[10] G. Pioggia, R. Igliozzi, M. Ferro, A. Ahluwalia, F. Muratori, and D. De Rossi, "An android for enhancing social skills and emotion recognition in people with autism," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 13, pp. 507-515, Dec 2005.

[11] B. Robins, P. Dickerson, P. Stribling, and K. Dautenhahn, "Robot-mediated joint attention in children with autism: A case study in robot-human interaction," *Interaction Studies*, vol. 5(2), pp. 161–198, 2004.

[12] H. Kozima, C. Nakagawa, and Y. Yasuda, "Interactive robots for communication-care: A case-study in autism therapy," in *IEEE International Workshop on Robot and Human Interactive Communication*, Nashville, TN., 2005, pp. 341-346.

[13] B. Scassellati, "Quantitative metrics of social response for autism diagnosis," in *Proc. IEEE International Workshop on Robot and Human Interactive Communication*, Nashville, 2005, pp. 585- 590.

[14] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42(3-4), pp. 143-166, 2003.

[15] R. W. Picard, *Affective Computing*. Cambridge: The MIT Press, 1997.

[16] J. Seip, *Teaching the Autistic and Developmentally Delayed: A Guide for Staff Training and Development*. Delta, British, Columbia, 1996.

[17] L. Ernsperger, Keys to success for teaching students with autism. Future Horizons, 2003.

[18] S. Wieder and S. Greenspan, "Can children with autism master the core deficits and become empathetic, creative, and reflective?" The Journal of Developmental and Learning Disorders, vol. 9, 2005.

[19] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real time face detection and facial expression recognition: development and applications to human computer interaction," in *Computer Vision and Pattern Recognition Workshop*, Madison, Wisconsin, 2003.

[20] C. M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Trans. on Speech and Audio Processing*, vol. 13, pp. 293-303, Mar 2005.

[21] P. R. D. Silva, M. Osano, A. Marasinghe, and A. P. Madurapperuma, "A computational model for recognizing emotion with intensity for machine vision applications," *IEICE Transactions*, vol. 89-D(7), pp. 2171-2179, 2006.

[22] F. Nasoz, K. Alvarez, C. L. Lisetti, & N. Finkelstein, "Emotion recognition from physiological signals for presence technologies." *International Journal of Cognition, Technology, and Work – Special Issue on Presence*, vol. 6, no. 1, 2003.

[23] D. Kulic and E. Croft, "Physiological and subjective responses to articulated robot motion," *Robotica,* vol. 25, pp. 13-27, Jan-Feb 2007.

[24] R. L. Mandryk and M. S. Atkins, "A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies," *Int J of Hum-Com Studies,* vol. 65, pp. 329-347, 2007.

[25] J. Groden, M. S. Goodwin, M. G. Baron, G. Groden, W. F. Velicer, L. P. Lipsitt, S. G. Hofmann, and B. Plummer, "Assessing Cardiovascular Responses to Stressors in Individuals with Autism Spectrum Disorders," Focus on Autism and Other Developmental Disabilities, vol. 20, pp. 244-252, 2005.

[26] M. M. Bradley, "Emotion and motivation," in *Handbook of Psychophysiology*, J. T. Cacioppo, L. G. Tassinary, and G. Berntson, Eds. New York: Cambridge University Press, 2000, pp. 602-642.

[27] P. Rani, C. Liu, N. Sarkar, and E. Vanman, "An empirical study of machine learning techniques for affect recognition in human-robot interaction," *Pattern Analysis Appl.,* vol. 9, pp. 58-69, May 2006.

[28] P. Rani, N. Sarkar, C. A. Smith, and L. D. Kirby, "Anxiety detecting robotic system-towards implicit human-robot collaboration," *Robotica,* vol. 22, pp. 85-95, Jan-Feb 2004.

[29] P. Rani, C. Liu, and N. Sarkar, "Interaction between human and robot – an affect-inspired Approach," *Interaction Studies*, vol. 9, no. 2, pp. 230-257, 2008.

[30] D. Ben Shalom, S. H. Mostofsky, R. L. Hazlett, M. C. Goldberg, R. J. Landa, Y. Faran, D. R. McLeod, and R. Hoehn-Saric, "Normal physiological emotions but differences in expression of conscious feelings in children with high-functioning autism," *J Autism Dev Disord,* vol. 36, pp. 395-400, Apr 2006.

[31] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *IEEE Signal Process Mag,* vol. 18, pp. 32-80, Jan 2001.

[32] R. M. Brown, L. R. Hall, R. Holtzer, S. L. Brown, and N. L. Brown, "Gender and video game performance," *Sex Roles,* vol. 36, pp. 793 – 812, 1997.

[33] A. Gillott, F. Furniss, and A. Walter, "Anxiety in high-functioning children with autism," *Autism,* vol. 5, pp. 277-86, Sep 2001.

[34] L. A. Ruble and D. M. Robson, "Individual and environmental determinants of engagement in autism," *J Autism Dev Disord,* 2006.

[35] K. Vansteelandt, I. Van Mechelen, and J. B. Nezlek, “The co-occurrence of emotions in daily life: A multilevel approach,” Journal of Research in Personality, vol. 39(3), pp. 325-335, 2005.

[36] J. I. Lacey and B. C. Lacey, "Verification and extension of the principle of autonomic response-stereotypy," *Am J Psychol,* vol. 71, pp. 50-73, Mar 1958.

[37] K. Dautenhahn, I. Werry, T. Salter, R. T. Boekhorst, "Towards adaptive autonomous robots in autism therapy: varieties of Interactions", *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pp. 577-582, Kobe, 2003.

[38] M. D. Lloyd and M. D. Leota, *Peabody Picture Vocabulary Test--Third Edition (PPVT-III)*, American Guidance Service, 1997.

[39] A. Pecchinenda, and CA. Smith, “The affective significance of skin conductance activity during a difficult problem-solving task,” *Cognition and Emotion*, vol. 10(5), pp. 481-504, 1996.

[40] J.T. Cacioppo, G.G. Berntson, J.T. Larsen, K.M. Poehlmann, and T.A. Ito, "The psychophysiology of emotion" in *Handbook of Emotions*, M. Lewis and J.M. Haviland-Jones, Eds.: New York: The Guilford Press, 2000, pp. 173–191.

[41] Papillo, J.F. and D. Shapiro, "The cardiovascular system" in *Principles of Psychophysiology: Physical, Social, and Inferential Elements*, J.T. Cacioppo and L.G. Tassinary, Eds.: New York: Cambridge University Press, 1990, pp. 456–512.

[42] C. Iani, D. Gopher, and P. Lavie, "Effects of task difficulty and invested mental effort on peripheral vasoconstriction," *Psychophysiology*, vol. 41(5), pp. 789-798, 2004.

[43] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system," in *Principles of Psychophysiology: Physical, Social, and Inferential Elements*, J.T. Cacioppo and L.G. Tassinary, Eds.: New York: Cambridge University Press, 1990, pp. 295-324.

[44] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley-Interscience, 1998.

[45] C. J. C. Burges, "A tutorial on Support Vector Machines for pattern recognition," *Data Min. Knowl. Discov.* vol. 2, pp. 121-167, Jun 1998.

[46] M. A., Wiering, "QV (λ)-learning: A New On-policy Reinforcement Learning Algorithm," *The 7th European Workshop on Reinforcement Learning,* Napoli, Italy, October, 2005.

[47] M. Pantic and L. J. M. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," *Proceedings of the IEEE,* vol. 91(9), pp. 1370-1390, 2003.

[48] R. A. Barkley, Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment 2ed. New York: Guilford Press, 1998.

[49] E. Hill, S. Berthoz, and U. Frith, "Brief report: cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives," *J Autism Dev Disord*, vol. 34 (2), pp. 229-235, 2004.

[50] S. Siegel and J. N. J. Castellan, *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill, 1988.

[51] R. Jafari, F. Dabiri, P. Brisk, and M. Sarrafzadeh, "Adaptive and fault tolerant medical vest for life critical medical monitoring," in *ACM Symposium on Applied Computing* Santa Fe, NM, 2005, pp. 272-279.

[52] C. J. C. H. Watkins and P. Dayan, "Q-Learning," *Machine Learning,* vol. 8, pp. 279-292, May 1992.

CHAPTER VII

ACTIVE LEARNING USING SUPPORT VECTOR MACHINE FOR PHYSIOLOGY-BASED AFFECTIVE MODELING FOR CHILDREN WITH AUTISM

In any classification procedure, samples must be collected, processed and labeled with a known class membership in order to determine a model (e.g., ± 1 for two classes). This procedure can be demanding in terms of the efforts required for (i) sample collection/processing and (ii) sample labeling. These two issues have particular impacts on developing human machine interaction applications when the participation of human subjects and the classification are demanded (e.g., in affective computing, Picard, 1997). For example, in order to build physiology-based affective-models, a human-machine interaction task usually has to be performed for 2-5 minutes to get the physiological signals and subjective reports have to be collected afterward (which may interrupt the interaction) in order to get one labeled training data point (Liu, et al., 2008; Picard, Vyzas, & Healey, 2001; Rani, et al., 2006). The experimental data measurement and labeling can be both time consuming and expensive. For developing an affective model for a real-world human-machine application, it would be important to investigate the efficient machine learning paradigms that depends relatively less on the availability of a large set of labeled training examples (Picard, et al., 2004).

In this work, we addressed this challenge in the context of building physiology-based affective model for children with Autism Spectrum Disorders (ASD). Specifically, we focused on the second issue and investigated the feasibility of alleviating the efforts of sample labeling by using Support Vector Machine active learning (SVM-AL).

As discussed in (Liu, et al., 2008, Chapter V), due to the unreliability of self-reports from the children with ASD, an autism therapist and a parent of the participant were also involved in the study. They were seated at the back of the experiment room, watching the experiment from the view of the video camera, observing how the task progressed on a monitor, and providing the subjective reports. For details about the experimental setup, please refer to Section 4 of Chapter V. This requirement posed additional challenges in participant recruitment and coordination of schedules among the therapist, participating child, and his/her parent. An alternative approach could be video-recording experiment (including the task progress), sending the video tapes to the therapist/parent, and allowing them to label the video segments at their own time. However, it can be expected costly and time consuming to label such video records. For example, in a study like (Liu, et al., 2008), for each child with ASD, there would be 6 hours of video needed to be reviewed and around 86 segments needed to be labeled. In such circumstances a method that allows the construction of reliable affect recognizers while only needs the labeling of a small fraction of samples can be of advantage, speeding up the procedure and possibly reducing costs due to extra analysis.

A strategy to tackle the problem is to use active learning, where the algorithm interacts with the samples prior to their labeling with the purpose of indentifying informative samples for which to request labels (Cohn, Atlas, & Ladner, 1994).

The most established method for training physiology-based affective models is passive learning (Liu, et al., 2008; Mandryk & Atkins, 2007; Nasoz, et al., 2003; Picard, Vyzas, & Healey, 2001; Rani, et al., 2006). In this case the set of training examples are the physiological features and the corresponding subjective reports, drawn at random from the training set. There is no relation between the expected error rate and a training example.

It has been shown in machine learning research that only a small portion of a large unlabeled data set may need to be labeled to train an active learning in order to achieve a strong classification performance (Lewis & Catlett, 1994; Cohn, Atlas, & Ladner, 1994; Dagan & Engelson, 1995). Previous work in active learning has concentrated on two approaches: certainty-based methods and committee-based methods. In the certainty-based methods (Lewis & Catlett, 1994), an initial system is trained using a small set of annotated examples. Then, the system examines the unannotated examples and determines the certainties of its predictions of them. The samples with the lowest certainties are then presented to the labelers for annotation. In the committee-based methods (Dagan & Engelson, 1995), a set of distinct classifiers is also created using the small set of annotated examples. The unannotated instances, whose predicted annotations differ most when presented to different classifiers, are presented to the labelers for annotation. In both paradigms, a new system is trained using the new set of annotated examples, and this process is repeated until the system performance converges.

In this work, we applied SVM-AL to reduce the label efforts in affective modeling for children with ASD. The seminal theoretical and practical discovery about SVM-AL was made by (Tong & Chang, 2001). It is based on version space reduction in Support Vector Machine (SVM) classifiers and was used in image retrieval. SVM-AL for active sample selection was also applied in drug discovery process (Warmuth, et al., 2002) and text classification with success (Tong & Koller, 2001). While active learning is an appealing tool for sample selecting/labeling, till date no published study has been found that specifically explored its application in building reliable affect recognizers while reducing the labeling efforts.

The rest of this chapter is organized as follows: A description of the machine learning algorithm (i.e., SVM-AL) is presented in Section 1. Section 2 presents experiment design for applying active sample selection and learning in physiology-based affective modeling for children with ASD. This is followed by a results and discussion section (Section 3). Finally, Section 4 summarizes the conclusions of this work.

1. Support Vector Machine Active Learning

SVM-AL is a learning system that is consisted of the SVM and a query function. A detailed description of SVM can be found in Appendix of Chapter V. SVM is one kind of passive learning. It is fed with a large pool of randomly selected samples and requires that the whole training set be labeled. In SVM-AL, the oracle (e.g., a user or labeler) becomes an integral part of the learning process. By using the query function, the system selects the most informative samples and asks the oracle for labels. The goal is to achieve strong classification performance with relatively few labeling requests.

One issue that needs to be addressed is: given an example, how should the query function be defined so that system can select the "informative" samples and request labels accordingly? In this work, we used margin-based query (Tong & Chang, 2001). Assume that samples are drawn from $X \subseteq R^n$, and that each sample $x_i \in X$ has an associated label y_i drawn from {−1, 1}, representing the affective states. Assume that L is a classifier with weight vector w, that a value p_i is defined for each x_i by $p_i = \langle w, x_i \rangle$, and the prediction by L for example x_i is given by $sign(p_i)$. The value p_i measures how close a sample is to the separating hyperplane. The closer a sample is to the hyperplane, the less confidence we have in L's prediction. One sampling/query function can be defined to request labels for such samples x_i that have relatively small distance to the decision boundary (i.e., choosing top k samples that have relatively small $|p|$). Tong and Chang (2001) proved this intuitive result: those unlabeled points closest to the separating hyperplane are the optimal choice to most reduce the size of the version space and thus improve the classification accuracy efficiently at each round.

2. Experimental Investigation

A simulation study was performed to examine SVM-AL and SVM passive learning (SVM-PL) both in terms of relative prediction performance and number of labels required to develop an affective model with satisfactory performance.

2.1. Experiment Design

We ran the experiment on the dataset obtained in our previous work (Liu, et al., 2008; Chapter V), where the physiological features of the children with ASD and the labels for the target affective state level from the therapist are given. For details of data collection and data set derivation, please refer to Section 3 and Section 4 of Chapter V.

In the experiment, SVM-AL was used to learn a classifier for each query in the simulation. A search is then a mapping from physiological features to the level of the target affective states (e.g., low/high level of anxiety) of children with ASD.

In a real application, an oracle (e.g., a therapist) would have physiological feature vectors and the corresponding video segments that are linked by the experiment epoch ID.

A label of a physiological feature vector can be obtained from the therapist after he/she reviews the corresponding video segment. A physiological feature vector will have the same ID as a video segment if that feature vector is derived from the physiological signals that were collected at the same epoch as the one when the video segment is recorded. To begin the search (i.e., active learning), the system asks the therapist to label a small number of video segments, and uses the corresponding labeled physiological feature vectors as the "seeds" for the first round training. After that, the system iterates between training a new classifier on the labeled physiological features and soliciting new labels from the therapist for informative samples. The sample selection/query is determined by the distance of the samples to the decision boundary of SVM as described in Section 1.

In the simulation, the labels in the dataset obtained in Chapter V will be fed to the system only when they are requested. This emulates the process of video review and labeling by the therapist.

<u>2.2. Procedure</u>

In summary, SVM-AL affective modeling in simulation is consisted of the following steps:

1) Randomly select s "seed" physiological feature vectors and get their corresponding labels
2) Train a SVM classifier
3) Record prediction performance on the testing dataset of size t
4) Calculate the distance of unlabeled physiological feature vectors from the current separating hyperplane, select top k feature vectors that has the least distances

(i.e., informative samples), and retrieve their labels

5) Train the SVM classifier on all the labeled training instances (i.e., update the separating hyperplane)

6) Record prediction performance on testing dataset of size t

7) Repeat steps 4), 5), and 6) iteratively until stop criterion is reached

The simulation procedure for SVM-PL is similar, except in step 4. SVM-PL selectes k feature vectors randomly and retrieves the labels for training, with no regard to the expected impacts of the selected training samples on the prediction performance.

In practice, the stop criterion can be defined when a system performance converges or a predetermined number of iterations have been reached. In this study, the SVM-AL affective modeling will not stop until all the training set is labelled, for the purpose of performance comparison.

We compared the relative prediction performance and number of labeling requests of SVM-AL and SVM-PL. In order to avoid the bias due to the impacts of different choice of "seed" and testing sets, the following conditions were imposed:

1) 30 trials for each target affective state of each participant were performed for both SVM-AL and SVM-PL, respectively.

2) In each trial, the same "seed" and testing dataset were used for both SVM-AL and SVM-PL.

3) The class distribution in "seed" and testing dataset is about the same as the whole data set.

In each trial, the experimental parameters were set as: $s = 5$, $t = 10$, and $k = 5$. The whole dataset for one target affective state of each participant has approximately 86 samples, which will result in 15 prediction performance records in each trial.

3. Results and Discussion

We investigated the feasibility of applying SVM-AL to develop affective models for children with ASD. SVM-AL and SVM-PL were compared in terms of both relative performance increase and number of requests for labeling to achieve a satisfactory performance. As shown in the results, SVM-AL is more advantageous by both evaluation criteria.

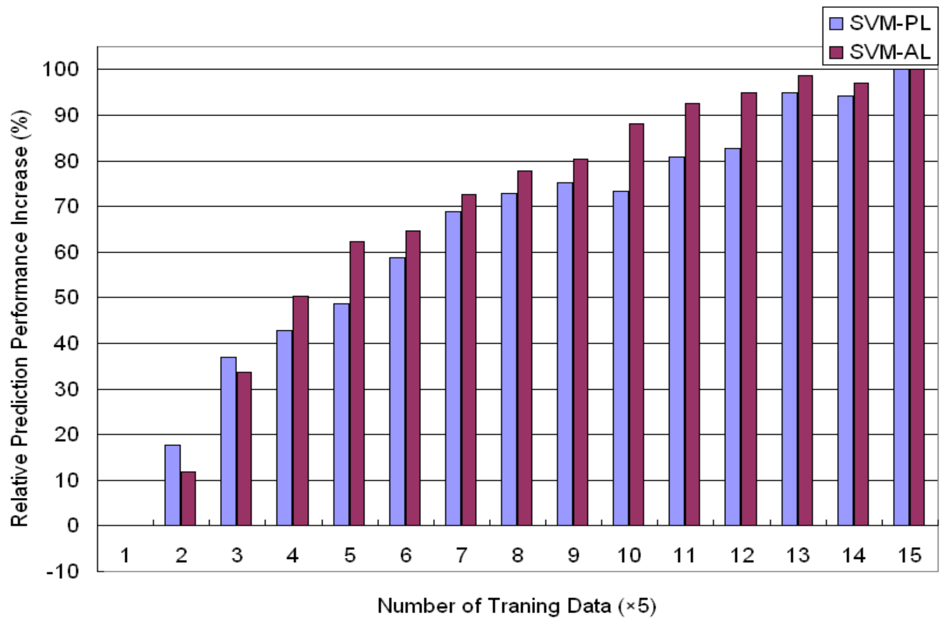


Figure 1. Relative Liking Prediction Performance Increase vs. Number of Labeling Requests for SVM-AL and SVM-PL

First, we present the result to demonstrate that active sampling allowed larger relative prediction performance (*RPP*) increase in the development of liking recognizers for children with ASD. Figure 1 shows that *RPP* for liking level increases as more labeled training data are available. It should be noted that, in this work we are focused on studying to what degree SVM-AL can enhance the prediction performance by choosing the informative samples for labeling, instead of examining the absolute prediction accuracy of SVM. In another word, we are more interested in investigating the effects of query function part of SVM-AL rather than the performance of SVM, which is another component of SVM-AL. The detailed results of SVM prediction performance in physiology-based affective modeling for children with ASD can be found in our previous work (Liu, et al., 2008; Chapter V). As a consequence, *RPP* was used for such measurement when the average is computed across the participants. *RPP* is calculated by linearly normalizing the absolute prediction accuracies to [0, 1]. The normalization scope is defined as the difference of two prediction accuracies: one is achieved after the system is trained with only the "seed" dataset and the other one is obtained after the system is trained with the whole training dataset.

In Figure 1, we can observe that SVM-AL has consistently larger *RPP* than SVM-PL after 15 sample requests. This showed that margin-based query of SVM-AL has more positive impacts on the prediction performance improvement of the liking model than the random query used in SVM-PL. By choosing more informative samples and asking the oracle (e.g., the therapist) for labels, SVM-AL enhanced the liking model's prediction performance in a more efficient manner. In Figure 1, SVM-AL and SVM-PL have the same *RPP* after trained with the "seed" dataset and the whole training dataset, with the value of 0% and 100%, respectively. This is because of the linear normalization as described before and the fact that the same "seed" and testing dataset were used for both approaches in each trial (as mentioned in Section 2). This leads to the fact that both SVM-AL and SVM-PL have the same *RPP* after those two trainings.

It can also be observed from Figure 1 that on average SVM-AL achieved 80% *RPP* by using 45 training samples and 90% *RPP* by using 55 training samples, whereas SVM-PL has to obtain 55 training samples and 65 training samples to get such performance improvement for liking prediction. By only taking 60%-70% sample labeling, SVM-AL provided the liking prediction models for children with ASD with acceptable performance. If 80% or 90% *RPP* is sufficient in certain applications, SVM-AL would significantly alleviate the costly and time consuming labeling process (e.g., hours of video review and annotations) for the therapist.

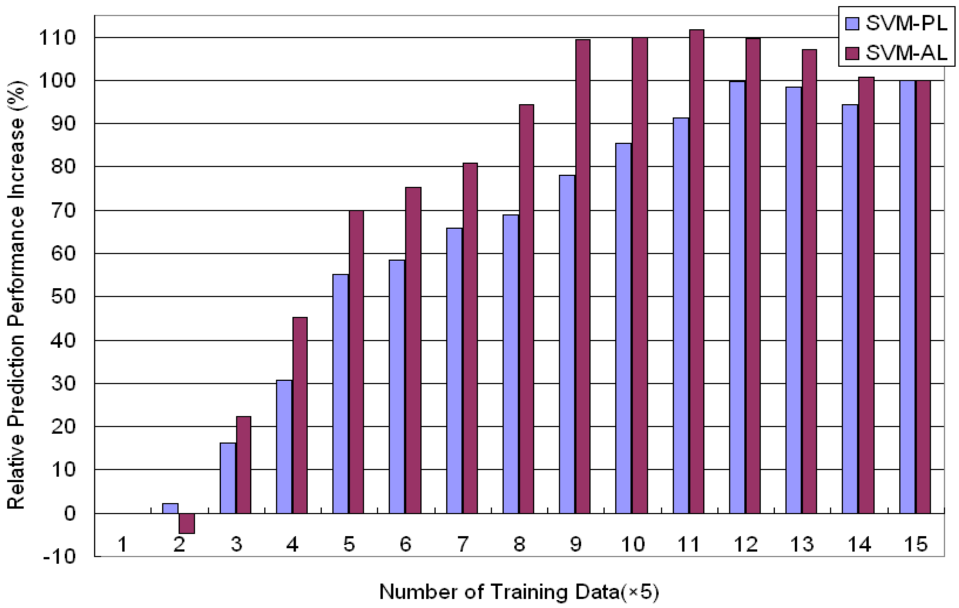


Figure 2. Relative Anxiety Prediction Performance Increase vs. Number of Labeling Requests for SVM-AL and SVM-PL

The *RPPs* in anxiety modeling for children with ASD was shown in Figure 2, with the lighter bars indicating *RPP* for SVM-PL and the darker bars showing *RPP* for SVM-AL. It was observed that *RPP* of SVM-AL is consistently larger than that of SVM-PL after they were trained by 10 selected samples. Furthermore, SVM-AL achieved 80% and

90% *RPP* for anxiety prediction after 35 sample requests and 40 sample requests, respectively. SVM-PL has to use 50 and 55 labeled samples to get the similar performance improvement. This result showed that the acceptable (e.g., with 80%-90% *RPP)* anxiety models for children with ASD could be developed by using only around 50% sample labeling with SVM-AL on average.

It should be noted in Figure 2, *RPP* of SVM-AL is larger than 100% after 9 rounds of training (i.e., 60% of the whole training dataset). This fact suggested that it is possible that an affective model trained by a fraction of samples, which were informatively selected for labeling, may have a better performance than the one trained by the whole labeled training dataset. This is generally due to the overtraining or to the presence of outliers as discussed in (Riccardi & Hakkani-Tur 2005). In the 2nd round of training, *RPP* of SVM-AL is of small negative value, which means the performance degenerated from the one obtained by being trained only with the "seed" dataset. As described in Section 2, the active query is based on the distance of unlabeled physiological feature vectors from the current separating hyperplane. At the first several training rounds, the model quality could be low and the choice of sample queries based on the decision boundary of such model could be biased and even results in performance decrease. This could also be the reason that in Figure 1 *RPP* of SVM-AL is smaller than *RPP* of SVM-PL for 2nd and 3rd round of training.

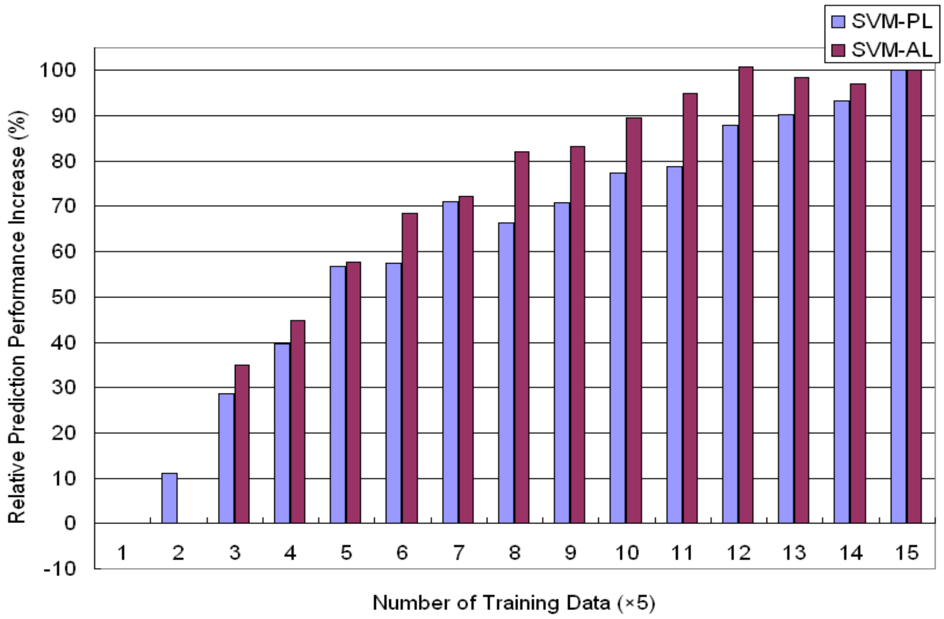


Figure 3. Relative Engagement Prediction Performance Increase vs. Number of Labeling Requests for SVM-AL and SVM-PL

Similar trend can also be observed in Figure 3 where results of *RPP* of engagement modeling for children with ASD are presented: larger *RPP* was obtained for SVM-AL after 10 training data and SVM-AL was capable of achieving acceptable affective models by the use of relatively less sample labels. For example, SVM-AL obtained 80% and 90% *RPP* for engagement prediction after 40 and 50 label requests (53% and 66% of whole training dataset), respectively; while SVM-PL demands more labeling efforts (e.g., 80% and 86% of whole training dataset) to get such a performance improvement.

4. Conclusions

In summary, we observed that: i) SVM-AL has larger *RPP* after the first several rounds of training (10-15 labeled training data on average) than SVM-PL; ii) acceptable model performance (e.g., with 80% or 90% *RPP*) can be achieved by asking the therapist to review and annotate only about 50%-60% of the dataset; iii) It could be possible to obtain a better performance by using a fraction of samples, which are informatively selected, than using the whole training dataset.

SVM-AL is capable of improving the performance of affect recognizers efficiently by using relatively less labeled samples. This work experimentally demonstrated that it is feasible to use this technique to alleviate the costly and time consuming video review and labeling efforts in physiology-based affective modeling for children with ASD while still maintaining sufficient model performance.

References

Cohn, D., Atlas, L., & Ladner, R. (1994). Improving generalization with active learning. *Machine Learning*, 15, 201–221.

Dagan, I. & Engelson, S. P. (1995). Committee-based sampling for training probabilistic classifiers. *Proc. 12th Int. Conf. Machine Learning*, 150–157.

Lewis D. D. & Catlett, J. (1994). Heterogeneous uncertainty sampling for supervised learning. *Proc. 11th Int. Conf. Machine Learning*, 148–156.

Liu, C., Conn, K., Sarkar, N., & Stone, W. (2008). Physiology-based affect recognition for computer assisted intervention of children with autism spectrum disorder. *International Journal of Human-Computer Studies*, 66(9), 662–677.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies, 65*(4), 329-347

Nasoz, F., Alvarez, K., Lisetti, C. L., & Finkelstein, N. (2003). Emotion recognition from physiological signals for presence technologies. *International Journal of Cognition, Technology, and Work – Special Issue on Presence, 6*(1).

Picard, R. W. (1997). *Affective Computing*. Cambridge: The MIT Press.

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

Prendinger, H., Mori, J., & 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.

Rani, P., Liu, C. C., Sarkar, N., & Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis and Applications, 9*(1), 58-69.

Riccardi, G. & Hakkani-Tur, D. (2005). Active learning: theory and applications to automatic speech recognition. *IEEE Transactions on Speech and Audio Processing*, 13(4), 504- 511.

Tong, S. & Chang, E. (2001). Support vector machine active learning for image retrieval. *Proc. IEEE Int. Conf. Computer Vision Patter Recognition*, 150–157.

Tong, S. & Koller, D. (2001). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 45-66.

Warmuth, M. K., Ratsch, G., Mathieson, M., Liao, J. & Lemmen, C. (2002) Active learning in the drug discovery process. In: *Advances in Neural information processings systems*, ed. by Dietterich T.G., Becker S., & Ghahramani, Z., vol. 14, 1449-1456.

CHAPTER VIII

CONTRIBUTIONS AND FUTURE WORK

Contributions

The contributions of this dissertation are in the area of physiology-based affect-sensitive human-machine interaction (HMI). The main contributions of this dissertation are:

i) Performed a systematic comparison of the strengths and weaknesses of four machine learning methods (K-Nearest Neighbor, Regression Tree, Bayesian Network, and Support Vector Machines) when they were employed for the physiology-based affect recognition. The proposed individual-specific modeling approach accounted for the phenomenon of person stereotypy and was capable of delivering competitive prediction accuracy on the intensity of affective states.

ii) Designed and implemented an affect-based dynamic difficulty adjustment (DDA) mechanism for computer games. It was experimentally demonstrated that the gaming experience can be augmented by using the affect-based DDA though a systematic user study.

iii) Proposed a closed-loop human-robot interaction framework, which was capable of performing accurate real-time affect recognition and modifying robot's behaviors accordingly. A robot-based basketball game was designed where a robotic "coach" monitored the human participant's anxiety level and dynamically changed its behavior parameters based on a state-flow model. It allowed users' skill improvement while maintaining desired anxiety levels.

iv) Designed and implemented computer-based cognitive tasks that successfully elicited target affective states (i.e., liking, anxiety, and engagement) in the children with ASD. Multiple subjective reports from an autism therapist, a parent, and the participant were analyzed to account for the suspected unreliability of the subjective self-reports from children with ASD. Support Vector Machines (SVM) was employed to develop a therapist-like affective model that yielded reliable prediction performance. Furthermore, (i) the effects of reducing the number of physiological signals to achieve more economical modeling, and (ii) the correlation between the affective model's prediction performance and the agreement between the therapist and parent on the subjective reports were also investigated.

v) Proposed an online affect detection and robot behavior adaptation framework for intervention of children with ASD. A robot used a Support Vector Machines based affective model to implicitly detect the affective cues in real-time. A reinforcement learning based behavior adaptation mechanism was employed to enable the robot to adapt its behaviors autonomously as a function of the predicted child's affective state. The robot learned the individual liking level of each child with regard to the game configuration and selected appropriate behaviors to present the task at his/her preferred liking level. This work was the first step to develop robot-assisted intervention tools to help children with ASD to explore social interaction dynamics in an affect-sensitive and adaptive manner.

vi) Investigated Support Vector Machine active learning (SVM-AL) to alleviate the effort required for sample labeling. By using the margin-based query to select the informative samples for the label requests, SVM-AL was capable of improving the relative prediction performance of affective models efficiently with the use of relatively less labeled samples.

Future Work

One of the future developments of this research is to incorporate multiple inferred affective cues (e.g., anxiety and frustration at the same time) and to account for other interaction information of interests, such as the human's performance and the context and complexity of the interaction task. With more sophisticated adaptation mechanisms, it would permit the adaptive and affect-sensitive interactions in the more complex HMI applications. Another area of emphasis in future could be integrating the affect-detection and adaptation system with various other modalities, such as dialogue, facial-recognition, and gestures. It would allow richer and more meaningful human-machine interaction. Furthermore, it would also be useful to conduct a large-scale study of physiology-affective state correlations to determine age, gender and culture-related patterns. This will assist in making generic affect-recognizers that can predict affective responses of a class of people.

While the concept of providing affect-based feedback in computer/robot assisted autism intervention for children with ASD was demonstrated in this work, there are several future areas of work remained to be explored. One is to investigate the design of socially-directed interaction experiments with robots interacting with children with ASD, e.g., integration of the real-time affect recognition and response system described here with a life-like android face, which can produce accurate examples of common facial expressions that convey affective states. This affective information could be used as feedback for empathy exercises to help children recognize their own emotions.

Enhancements on the intervention process could also be envisioned.

It would also be of great importance to i) reduce the verbal components in the cognitive tasks which would allow application to a broader part of the ASD population and ii) develop computer-based intervention tools that address the social communication deficits of children with ASD. My colleagues in our laboratory are currently working on these two challenges.