Report on NIAC Phase I Grant entitled
A Novel Interface System for Seamlessly Integrating Human-Robot Cooperative Activities in Space

Principal Investigator
Professor Nilanjan Sarkar
Department of Mechanical Engineering (Primary)
Department of Electrical Engineering and Computer Science (Secondary)
Vanderbilt University
Tel: (615) 343-7219
Fax: (615) 343-6687
Email: nilanjan.sarkar@vanderbilt.edu

Co-Principal Investigator
Professor Craig A. Smith
Department of Psychology and Human Development
Vanderbilt University
(615) 322-8298 (Office)
615) 343-9494 (Fax)
E-mail: Craig.a.Smith@vanderbilt.edu

Date of submission
December 2, 2002
Phase I Report

1. Introduction

To quote Weinberg, “The effort to understand the universe is one of the very few things that lifts human life a little above the farce and gives it some of the grace of tragedy.” NASA’s effort to understand the universe is to answer the most fundamental questions: How did we get here? Where are we going? Are we alone in the Universe?

To address these questions one of NASA’s strategic enterprises – Human Exploration and Development of Space (HEDS) – has as its first major objective to explore the space frontier. HEDS has proposed to undertake human-robotic expeditions throughout the inner solar system to meet this objective. It is believed that space exploration will be most effective if human capabilities are synergistically combined with those of robots. Such a human-robot system, developed effectively, will reduce exploration risks, improve efficiency, and achieve the overall mission goals more quickly and in a better manner. In the future human-robot teams will be expected to help: expand human presence beyond the vicinity of the earth, understand the future habitability and sustainability of the earth, and search for extraterrestrial life. Realization of these goals will require dramatically improved control and communication capabilities between human and robot. With these considerations in mind, the Office of Space Flight is interested in understanding the optimized use of humans and robots in space. NASA Johnson Space Center is working on advanced extravehicular activity (EVA) capabilities. NASA Jet Propulsion Laboratory is focusing on human-robot teams beyond low-earth orbit. It is expected that within 40 years a robot will be able to perform autonomous EDL (explore, drill and discover), and that human and robot will communicate using multiple modalities.

However, there are two major issues that must be adequately addressed before human and robot can be seamlessly integrated into a versatile team suitable for space applications. One, human and robot should be able to interact more naturally. And two, the robot, should be able to change its own autonomy level or task priority based on the interaction, which is a hallmark of intelligence. The existing human-robot paradigms, however, do not adequately address these requirements because of the robot’s inability to understand the human’s needs. Additionally, limitations in current sensor and communications technologies are further impediments. Nonetheless, consideration of human factors is very important for any successful human-robot cooperative system. A lesson learned from process automation, called “ironies of automation” argues that without human factor considerations any such effort is likely to be more problematic than beneficial.

The futuristic goals of NIAC has provided us with an ideal platform to propose a revolutionary concept for the seamless integration of human and robot where the robot will implicitly understand and act on human needs by following implicit commands and reconfiguring its autonomy and/or task priority. The proposed concept draws its strength from the recent advances in affective computing, signal processing, experimental psychology and control theory. It projects the current state-of-the art and trends of wearable sensor technology and real time systems into the future, and develops the proposed concept, with accompanying systems and architecture, to potentiate the development of a truly versatile human-robot system for space exploration within the next 10-40 years. Allowing more natural interaction between the human and robot will undoubtedly improve human-robot cooperation, increase productivity and reduce mission risk. It will have the potential to be instrumental in achieving NASA’s goals in
space exploration and development. Our Phase I work investigated the basic feasibility of this concept.

2. Overall Goals and Phase I Objectives

In order to establish natural human-robot interaction, we proposed to examine the role and potential of implicit communication. The current trend in human-robot interaction requires explicit communication from the human. Typically a person communicates with a robot through either typed or (more rarely) spoken explicit instructions. Such modes of communication may not be ideal for space applications. Additionally, a human-robot interaction that relies solely on explicit communication ignores the potential significant gains of implicit communication that have been demonstrated by studies in experimental psychology.

There are two eventual goals of the present research. One, we want to investigate whether and how a human can implicitly command a robot for specific task requirements. For example, can a human implicitly communicate with a robot to ask it to fetch a tool? This will, among other benefits, enable hands-free and nonverbal communication between an astronaut and her/his cooperative robot during mission tasks such as space station repair and maintenance. If such communication capabilities can be achieved, they will revolutionize state-of-the-art human-robot space exploration by improving command and control, and productivity. Two, we would like to further investigate whether a robot can sense the psychological states of the human and can make a decision regarding whether to change its autonomy level or task priorities. For example, consider a futuristic scenario where a human will explore a planet with a robot. Both human and robot will have their own exploration tasks to carry out. However, if the human is in trouble or in danger, can the robot sense his/her panicky/frightened psychological state and come to help the person, suspending its originally assigned duties without being explicitly commanded to do so by the person? Such a capability would reduce the mission risk by enhancing crew safety.

There is scientific evidence in experimental psychology that supports the feasibility of affect recognition in off-line, controlled experimental situations. Our objective in this research is to investigate how we can sense affect in real-time and in realistic task conditions, as well as how a robot can use this sensed state to make decisions for itself. Such work has not been done before. Once we successfully achieve this goal, our next objective will be to analyze the affective signal to extract any meaningful task-specific intention from it. This intention will then be translated into executable commands for the robot, achieving the goal of implicit communication. Such human-robot cooperation will herald a new era for human-robot teaming.

We have proposed to achieve these goals in the following steps. First, we sought to incorporate implicit communication through the understanding of affective states. In particular, we sought to give the robot the ability to sense and interpret the human’s affective state. As affective states, we include such things as frustration, anxiety, engagement, and fatigue as well as more traditional emotions such as happiness, sadness, and anger. It is important to mention that other researchers have argued the need for a computer to understand human emotion \[3\] as well as demonstrated that even a rudimentary implicit communication ability significantly improved the performance of human-computer interactions \[5\]. Second, we propose to design a control architecture for the robot that can be sensitive to the affective state of the human. That is, when the robot senses the affective state it will make a decision regarding the human’s needs, and if required will change its autonomy level or task priority to address the human need. Such a capability will be a hallmark of a successful human-centered robotic system that is robust and fault-tolerant. Third, we will investigate how to extract task-specific intention from the sensed affective states. We are not proposing thought recognition since it is an unsolved problem.
Because one can have a virtually limitless number of thoughts without any physical manifestation outside of the brain, the problem of understanding thoughts from brainwaves is likely an intractable problem. Nevertheless, it may be possible to infer intention in a specific task-related context. We propose to develop learning, probabilistic and heuristic methodologies to combine affective states and brainwaves to determine task-specific intention that will be reasonably robust. The intention will be finally translated into a set of executable commands for the robot to carry out.

From the outset it was obvious that this grand objective could not be achieved in Phase I alone. Therefore, in Phase I, our primary objective was to investigate how to best sense affect robotically, including an examination of the advantages and disadvantages of various measurement modalities. In addition, we proposed to investigate the current noninvasive methods of brainwave monitoring such as electroencephalographs (EEG) and magnetoencephalographs (MEG) to compare their respective advantages and disadvantages. More specifically we proposed to:

1. Use physiological sensing and real-time signal processing for affect recognition in realistic experimental conditions in order to demonstrate the feasibility of developing a real-time affect-detector that might be useful for human-robot implicit communication, and
2. Survey and make a cost-benefit analysis of the existing brainwave monitoring technologies, such that we could make a decision as to which of these technologies, if any, we could best integrate into our subsequent work on affect detection and recognition.

In our Phase I activities, we not only were able to accomplish both of these objectives, but for each objective we were able to advance our research somewhat further than we had proposed to do. Specifically:

1. We developed a preliminary robotic control architecture by which a robot could utilize the information to be generated by the affect detector we are trying to develop;
2. We were able to integrate this control architecture with a rudimentary prototype of a functioning affect detector to produce a feasibility demonstration of a simple human-robot implicit communication system in action; and
3. Because the outcome of our survey of existing brainwave monitoring technologies clearly implicated EEG as the most promising such technology for our purposes, we conducted a small pilot test to begin to examine the feasibility of integrating EEG technology into the affect detector we are trying to develop.

In conducting these Phase I activities, our hope was that we would be able to demonstrate the basic feasibility of the proposed human-robot cooperative framework. In the process, we hoped we would generate enough scientific understanding of what would be involved in the further development of such a framework that we would be able propose a detailed Phase II research plan to move us significantly toward the actual development of a functional human-robot implicit communication system. We believe that we have fully met these goals, and in the next two sections of this document we provide a detailed report of our Phase I activities and accomplishments. Then in the final section we outline our proposed Phase II activities.

3. Affect Detection and Recognition

An important initial step in developing a real-time implicit communication system in which a robot is able to diagnose and respond to a human’s affective state is to develop an “affect recognizer” that the robot can use to detect when the human is in a noteworthy affective state,
and to diagnose what that state is, so that the robot can respond to the state appropriately. Much of our Phase I research activities were directed toward, first, beginning the development of such an affect recognizer, and then examining the feasibility of using such a recognizer in real-time to modulate a robot’s behavior. In the next sections of this report we first review the theoretical and methodological considerations and decisions that guided our empirical research on this problem; then we report on the research we have conducted to date toward developing an affect recognizer; and finally we describe a human-robot experiment we conducted to examine the feasibility of using such a recognizer to modulate a robot’s behavior in real-time.

3.1 Developing an Affect Recognizer: Theoretical and Methodological Considerations

Three major decisions, made to ensure the tractability of our Phase I investigation, guided our initial investigations: First, we decided to limit ourselves to the use of non-invasive physiological sensing to recognize affect; second, we proposed to examine the physiological markers of affect in a highly person- and context-specific manner; and third, rather than trying to classify the person’s discrete emotional states in terms of such categories as anger, fear or sadness, etc., we initially attempted to locate the person’s state along a dimensional continuum corresponding to task-engagement/arousal, with some attempt to differentiate a state of high task-engagement from one of high anxiety. The rationales for these three decisions are briefly described below.

Decision 1: Focus on non-invasive physiological sensing. Affective states are embodied, and they have potentially observable effects on a wide range of response systems, including facial expressions, vocal intonation, gestures, and physiological responses such as cardiovascular activity, electrodermal responses, muscle tension, respiratory rate and amplitude, and others [6]. These diverse response modalities represent a wealth of potentially observable information that can be used to infer an individual’s affective state. However, an attempt to examine all the types of observable information available to convey affective state would be immensely complex, both theoretically and computationally. Moreover, some physical expressions are culture, gender, and age dependent, which creates additional difficulties in analyzing them. Physiological responses, on the other hand, are generally involuntary and less dependent on those factors. They offer an avenue for recognizing affect that may be less obvious for humans but more suitable for computers, which can quickly implement signal processing and pattern recognition tools to infer underlying affective states. Recent advances in wearable computers and affective computing [4] have ushered in the era of small and lightweight biofeedback sensors that can sense and process physiological signals in a non-invasive manner which are comfortable for the user to wear, unobtrusive, and fast enough for real-time applications [7]. Such capabilities have inspired us to use physiological sensing as an initial means to recognize human affect for our proposed controller. Specifically, we decided to focus initially on physiological indicators encompassing both cardiovascular and electrodermal activity, along with a limited number of facial activities. Then, as our initial efforts met with success, our plan was to gradually expand the range of indicators we might use in our affect recognizer.

Decision 2: Adopt a highly person- and context-specific approach. Despite numerous attempts over the past century, efforts to identify the distinctive physiological signatures associated with various discrete emotions, such as anger, fear, or sadness, have proven to be largely unsuccessful [8], due, in part, to the related phenomena of person-stereotypy and situation-stereotypy [9]. That is, within a given context, different individuals will often express the same emotion rather differently, with different individuals demonstrating somewhat different characteristic patterns of response (person stereotypy). For example, some individuals may be
very expressive facially, but show relatively little autonomic activity in response to affective stimulation, whereas others may show little facial activity, but strong autonomic responses (e.g., [10]). In a similar manner, across contexts the same individual might express the same emotion differentially, with different contexts pulling for different characteristic responses (situation stereotypy). Part of the reason for situation stereotypy is that none of the physiological indicators of affective state are solely devoted to affect expression, and thus there is no one-to-one mapping between a physiological indicator and an affective state [11]. Instead the physiological activities associated with affect each serve a variety of homeostatic, metabolic, and motivational functions that are often unrelated to the person’s emotional state. Thus the pattern of activity associated with anxiety when a person is at rest will invariably look rather different from the pattern associated with anxiety when the person is physically active. Moreover, because the primary functions associated with various physiological indices are in service of metabolic demands (e.g., a primary purpose of cardiovascular activity is to keep muscles sufficiently oxygenated -- i.e., fueled -- such that they can remain active), the physiological patterns associated with two different emotions, such as anger and fear, at high activity levels may be more similar to one another than the physiological patterns associated with the same emotion at two different activity levels.

In order to circumvent the difficulties represented by the above considerations, we decided to develop our affect-recognition system to be both individual- and context-specific. Within a given context, we will intensively examine the affective responses of a small number of individuals, and will develop an affect recognizer for each. By comparing the similarities and differences in the recognizers across individuals, we will be able to determine, for that context, which recognition cues generalize across individuals, and which are person-specific. As functioning recognizers are developed across individuals within a given context we will also examine responses within individuals across contexts, and thus be able to identify the aspects of the recognizer that are cross-situationally general versus context specific.

As noted above, a key factor to control in selecting the contexts to examine is the activity level called for by that context. Traditionally, most of the research examining the patterns of physiological activity associated with affect has involved participants at rest passively being exposed to such stimuli as affectively evocative slides. The rationale has been to examine the patterns of activity unique to emotion, as uncontaminated as possible by homeostatic and metabolic activity [8]. However, such observations are undesirable for the present purposes. First, because the physiological activity in emotion is believed to be in service of the person’s attempts to contend with the emotion-eliciting situation, by restricting such efforts one gets a very biased and restricted view of the physiological activity associated with the emotion or affect. Second, such a context is of very limited utility for present purposes since few, if any, of the envisioned space applications of person-robot interaction involve such a passive human participant. Thus, the recognition algorithms developed under such conditions are unlikely to have broad practical utility. On the other hand, starting with an individual vigorously engaged in physical activity is likely to make the search for recognition algorithms very difficult, as the affect-related contributions to the physiological activities are likely to be obscured by the large contributions mandated by metabolic demands. Ultimately we plan to develop recognition algorithms under such conditions. However, such difficult conditions did not seem to be the optimal ones to begin our development efforts.

Therefore we opted to begin our efforts in the context of cognitive problem-solving tasks. In such a context the individual will be actively involved in contending with the task, such that the
affect system can be fully engaged in a way that is not possible with a more passive task. However, the amount of physical activity involved is low enough that the affective signals need not be overly obscured by the influences of metabolic demands.

Decision 3: Initially examine a restricted range of affective states. Finally, in developing an affect recognizer, one can attempt to identify the discrete categorical state of the participant, or one can adopt a more dimensional approach in which one attempts to locate the affective state along some number of continuous affective dimensions, such as arousal or task-engagement. In our initial efforts we have adopted the latter approach. As mentioned above, previous efforts to reliably identify emotion-specific patterns of physiological activity have had very limited success, and such patterns, when found, have proven to be extremely context-dependent. In contrast, efforts to link parameters of physiological activity to such broader dimensions as arousal and task engagement have been much more successful. Thus, although we ultimately plan to explore the feasibility of an affect-recognizer inferring discrete emotional states from physiological signals, we believe that focusing our early efforts on a more dimensional approach has a much higher probability of success.

Therefore in our initial efforts, we have been attempting to identify a more restricted range of affective states that largely vary along a dimension of task engagement. Thus, our research strategy was to present our research participants with a series of tasks in which task difficulty was systematically varied in order to produce within-person variability in task engagement. At times the tasks were trivially easy in a way that was intended to produce states of boredom and low task engagement; at other times they were moderately difficult, in a way that was designed to produce relatively high, optimal levels of task engagement; and at other times the tasks were extremely difficult, in a way that was designed to produce very high levels of task engagement, accompanied by such states as anxiety and frustration. Our initial data-analytic strategy was to try to identify physiological correlates of task engagement that could be used in the affect recognizer to be developed. In addition, though, with an eye toward beginning to identify differences between discrete affective states, we also examined the correlates of anxiety, such that we could begin to examine the feasibility of differentiating states like anxiety from task engagement, per se.

3.2 Experimental Tasks, Procedures and Participant Information

Participants and Design  Participants were 6 (4 female, 2 male, age range 18-54) members of the Nashville, TN community. Participants were recruited via word of mouth and announcements in relevant courses. All participants received payment of $75 in return for involvement in the study. In line with the person-specific approach we have adopted, the design was fully within-subjects, such that all participants completed a series of 6 experimental sessions. They performed one task during each experimental session, which lasted approximately one and a half hours. Sessions were scheduled on separate days.

Tasks  Three different problem-solving tasks were employed in this study – an anagram task, a math task, and a sound discrimination task. These tasks were chosen to assess a variety of problem-solving skills, but all were considered to be within the restricted cognitive problem-solving context we decided to focus on initially. Two versions of each task were created – an easy sequence and a difficult sequence. In the anagram and sound tasks the easy sequence started out with moderately difficult trials which over the first half of the sequence became trivially easy, and then stayed that way for the remainder of the task. In responding to these sequences, participants were expected to start out at fairly high levels of task engagement, but the long series of trivially easy problems in these sequences were expected to produce low levels
of engagement corresponding to affective states of fatigue and boredom. In the math task, the problems started out easy, and remained moderately easy throughout the task. This was expected to produce fairly high levels of engagement throughout the task. In the difficult sequences for all three tasks, the trials started out moderately difficult, as in the easy sequence, but then increased in difficulty until by the end of the task they were virtually impossible to solve or perform correctly. This difficult sequence was expected to produce increasing levels of task engagement across the task for most participants, although as the tasks became virtually impossible, it was expected that anxiety and frustration would also increase. It was also anticipated, that in the face of the virtually impossible trials in the difficult sequences, some participants would become resigned and become less engaged as they gave up on the task and reduced their efforts. All participants completed both versions of all three tasks. Details for each of the three types of tasks are listed below.

**Anagram Task.** Both sequences of the anagram task were developed from a master list of over 800 5-letter anagrams that had been derived from a variety of sources. In a pilot study, five members of the research team attempted to solve each of the anagrams and then classified it on a five-point scale ranging from “extremely easy” to “extremely difficult.” An example of an extremely easy anagram is: AWADR (award), and a sample difficult anagram is OCHSA (chaos). The easy sequence consisted of a series of 600 anagrams. In this sequence all anagrams were comprised of easy words that were scrambled so that the anagram could be solved by switching just two of the letters (see very easy example above). The difficult sequence consisted of a series of 200 anagrams. The first 50 anagrams in the sequence were a mix of difficulty levels except that no extremely easy anagrams were included. As the sequence progressed only difficult and extremely difficult anagrams were included, with the extremely difficult ones dominating the latter portions of the list.

For each trial of the task, the anagram was presented to the participant on a computer screen for up to 45 seconds, and once the participant had solved the anagram s/he was to enter the correct answer via the computer’s keyboard. If an answer was not entered within the allotted time, then the participant’s (non)response was coded as wrong, and the program advanced to the next trial. Participants received feedback on their performance following every trial.

**Math Task.** Both sequences of the math task consisted of a series of 12 math-word problems. In the easy sequence all problems were judged, based on previous research [12] and pilot work to be very easy or moderately easy. In the difficult sequence, the first five problems had parallel structure, and were of similar difficulty as the corresponding problems in the easy sequence, but the remaining seven problems were very difficult, and based on pilot work, were judged to have a low probability of being solved by any given participant. An example of an easy problem is:

An astronaut requires 2 pounds of oxygen per day while in space. How many pounds of oxygen are needed for a team of 3 astronauts who are going to spend 5 days in space? (Answer: 30)

An example of a difficult one is:

Tammy has $9.70 in nickels, dimes, and quarters. The number of nickels is 4 more than 3 times the number of dimes, and the number of quarters is 5 fewer than 2 times the number of nickels. How many nickels does Tammy have? (Answer: 19)

In solving each problem, the participant was able to use a scratch pad and calculator, and s/he had up to four tries to solve each problem correctly. Feedback on the participant’s performance was provided after each solution attempt for every problem.
Sound Discrimination Task. Both the easy and difficult versions of the sound discrimination task consisted of a series of 380 trials in each of which the participant heard a sequence of three tones and had to decide whether the first and third tones were the same or different. Thus, the first and third tones were “comparison” tones, and the second tone served as a “distractor” tone. The function of the distractor tone was to serve as a “mask” designed to increase the base difficulty of the trials by disrupting the participants’ iconic memory of the first tone. The distractor tone was always presented 25 ms after the offset of the first comparison tone, and the second comparison tone was always presented 250 ms after the offset of the distractor tone.

In order to control for guessing, on half of the trials the two comparison tones were identical, whereas on the remaining trials they were different. The difficulty of a trial was manipulated by varying both the duration of each tone in the sequence and the frequency difference between the first and third tones. Tones could be presented for one of five durations (50, 100, 250, 500, or 1000 ms), and within a trial, all three tones were presented for the same duration. On “different” trials, the two comparison tones could differ by one of four frequency discrepancies (25, 50, 100, or 250 Hz). Across trials, half of the comparison tones were centered around 2000 Hz and half were centered around 3000 Hz. Thus on half of the different trials one of the comparison tones had a frequency of 2000 Hz and the other was the assigned discrepancy higher or lower than this frequency. In the remaining “different” trials, one comparison tone had a frequency of 3000 Hz and the other was the assigned discrepancy higher or lower than this frequency. The order in which the two discrepant tones were presented was counterbalanced across trials. In addition, the “same” trials were conceptually matched to the “different” trials selected for the sequence, such that half of the time both tones would be either 2000 or 3000 Hz, and for each the remaining trials, both tones would differ from this base frequency by the discrepancy used in the different trial to which it was conceptually matched. Thus, the task was constructed such that the frequency of the first comparison tone did not provide any information regarding whether the trial would be a same or different trial. Distracter tones were selected to vary between 2300 and 2700 Hz, and thus were always outside of the range of the two comparison tones.

Based on extensive pilot work, it was determined that all trials in which the frequency difference was 250 Hz and the tones were presented for at least 100 ms were quite easy. In contrast all trials in which the tones were presented for only 50 ms (any frequency discrepancy) were very difficult, and all trials in which the discrepancy was only 25 Hz (any duration) were slightly more difficult still. The remaining possible trials were all found to be of moderate difficulty. In constructing the two sequences, both consisted entirely of a mix of moderately difficult trials for the first 80 trials. Over the next 120 trials the proportion of easy or difficult trials increased steadily for the easy and difficult sequences, respectively. The final 180 trials of the easy sequence was comprised totally of very easy trials, and those of the difficult sequence were composed totally of very difficult trials, with the final 60 difficult trials involving only 25 Hz discrepancies.

The tones were presented binaurally to the participants via headphones, and after listing to the entire sequence of three tones, the participant indicated whether the comparison tones were the same or different by entering “y” or “n” on the keyboard, respectively. Participants received performance feedback following every trial.

Procedure At the initial session, participants were met by a female experimenter and seated in front of a table holding a computer. The experimenter provided a study overview, then left the room so participants could privately read a consent form. After participants had given their
consent to proceed with the study, the experimenter gave them a background information form to fill out. Upon completion of this form, the experimenter returned, collected the forms, attached the physiological sensors (see below), and began a 10-min baseline period, during which time the participant read through some magazines. On subsequent experimental sessions, the experimenter greeted the participant, hooked up the sensors, and proceeded directly to baseline.

Following baseline, the experimenter returned to the subject room and instructed the participants to read through the instructions for the task for that day, which were presented on the computer. She also informed them that after reading through the instructions, they would be presented with a practice task to help them get used to the procedures before the main task began. The participants then performed the practice task (which consisted of 10 anagrams of varying difficulty, three easy math problems, or 20 sound discrimination trials of varying difficulty for both versions of the anagram, math, and sound tasks, respectively), and had an opportunity to ask the experimenter any procedural questions about the upcoming main task. After any such questions were addressed, the main task was begun. Upon completion of the main task, the physiological sensors were removed, and after the participant was given an opportunity to clean him or herself up, the appointment for the next session (if any) was confirmed, and the participant was dismissed. At the conclusion of the sixth and final session the participant was thoroughly debriefed regarding the purposes of the study, and any questions the participant had at that point were addressed.

Apparatus. The experimental task and all associated instructions and feedback were presented to participants via a Pentium-based Compaq computer located in the participants’ room, while the physiological data were collected via a second Pentium-based Compaq computer in a nearby control room. The trial order for each task, participants’ responses and reaction times for each trial, and their self-report data (see below) were stored on the hard disk of the participant’s computer. In addition, unique signals were generated via this computer’s parallel port to mark key events during the task performance (onset of problems, time of participant’s response to the problem, whether the problem was answered correctly, etc.) to be linked with the physiological data record.

These event-marker signals, and all physiological measures were recorded using Coulbourn Instruments Lablinc equipment. All channels were sampled continuously at 1000 Hz and stored
on the data-acquisition computer’s hard drive through a Coulbourn Instruments Wingraph port (V19-02). The event signals from the participant’s computer were sampled through the 8-bit digital IO port of this unit, and the physiological signals were digitized with 12-bit accuracy through the unit’s analog ports. Information on the modules used to condition the physiological signals is included in the description of measures, below.

3.3 Physiological Measures and their rationale

A variety of physiological measures were collected continuously during the pre-task baseline and over the entire performance of each task – a measure of electrodermal activity (skin conductance), several indicators of cardiovascular activity (the electrocardiogram [ECG], relative pulse volume, and digit skin temperature), and two channels of facial electromyography (EMG, in the brow [corrugator] and the jaw [masseter] regions). These signals have been selected because they can be measured non-invasively and are relatively resistant to movement artifact. A number of them – electrodermal activity, the various parameters of cardiovascular activity, and jaw EMG – are all indices of tension, arousal, and/or task engagement. However, it is important to note that arousal is not a unidimensional construct, and since the seminal work of the Lacey’s it has been known that these measures reflect different aspects of it. Thus, for example, estimates of vagal tone/parasympathetic activity, derived from the ECG, can be interpreted as an indicator of the person’s stress level, jaw EMG can reflect a general level of somatic tension, whereas the electrodermal and other cardiovascular indicators can reflect task engagement, although the electrodermal activity appears to be more closely related to the attentional aspects of task engagement, whereas the cardiovascular activities appear to be more closely related to preparation for, and actual, physical or mental exertion. In general, we expect these indicators of arousal to be lowest when the person is under-engaged, at moderate levels, when s/he is optimally engaged, and highest when the person is over-engaged. However, differences in the functions and systematic activities of the different parameters should allow us to further differentiate between various states associated with task engagement. For instance, one might expect especially high levels of electrodermal activity to be associated with anxiety relative to non-anxious task-engagement. Brow EMG has been added to these other measures, because it may provide additional information to allow us to further differentiate among more specific affective states. Specifically, EMG activity in the eyebrow (corrugator) region has been shown to be related to the perception of goal obstacles, and thus should be useful in discriminating frustration (an especially high-obstacle state) from anxiety.

In addition to the continuous monitoring of the physiological measures, participants also reported periodically on their subjective emotional states. These reports were used as reference points with which to link the observed physiological activities to the participants’ affective states. In addition, objective indices of the participants’ performance on the tasks were also collected, and these were used to help validate the subjective affective reports. For instance, all else being equal, how engaged a person is on a task should be related (albeit imperfectly) to how well that person does on the task. Thus, to the extent to which the participants’ reports of task-engagement are valid, we would expect them to be reliably correlated with task performance. Each of the measures collected is described below.

Skin conductance (SC) was recorded using Med Associates silver finger strip electrodes (TDE-50A). The sensors were placed on the distal phalanges of the second and fourth fingers of the non-dominant hand. A Coulbourn Isolated Skin Conductance Coupler (Model V71-23) was
used to apply a constant 0.5 volt current across the sensors and to condition the skin conductance signal.

**Digit skin temperature** was recorded using a thermistor placed on the palmar surface of the distal phalanx of the fifth finger on the non-dominant hand. A Coulbourn Temperature Module (Model S71-30) was used to convert temperature into a voltage.

The **Electrocardiogram** (ECG) was measured using a three-lead placement, with leads at the top and bottom of the breastbone and a ground on the back of the non-dominant hand. This placement was chosen to minimize movement artifact. A Coulbourn Bioamp with BandPass Filter (Model S75-01) was used to amplify and filter the ECG wave.

Relative **Finger pulse volume** was measured using a Coulbourn Pulse Monitor Densitometer (Model S71-40), with the optical pulse transducer placed on the middle finger of the nondominant hand. The pulse transducer sends a beam of light through the finger, and measures how much light is reflected back vs how much is absorbed by the blood in the finger tip, which itself continuously changes as a function of the relative dilation/constriction of the capillaries in the finger and the ebb and flow of the pulse wave of blood in the finger.

**Facial electromyography** (EMG) was measured at the Corrugator (brow) and Masseter (jaw) sites, using the skin preparation procedures and placement sites recommended by Fridlund and Cacioppo [16]. At each site, 2 4 mm electrodes were placed side by side over the target muscles. The signal was filtered and amplified using Coulbourn Bioamps with Bandpass Filters (Model S75-01). The filters were set to retain activity between 8 and 1000 hz. The waveforms were then rectified and integrated with a time constant of 20 ms using a Coulbourn Countour Following Integrator (Model S76-01).

**Self-reported affective state** was measured periodically throughout each task. At each assessment, the participant responded to a series of 14 items using 9-point scales (1 = not at all, 9 = extremely much). First participants were asked how much they cared about how well they did on the task, how well they thought they were doing on the task, and how difficult they thought the task was. Then they rated the extent to which, during the previous interval, they had been feeling each of the following 11 affective states: resignation, hope, anger, fatigue, challenge/determination, anxiety, boredom, overload, calmness, interest, and frustration. For each affective state, the participant was presented with a cluster of two or three adjectives (e.g., “nervous, anxious, tense” to define anxiety), and the participant provided a single rating to indicate the degree to which s/he had been feeling the state represented by that adjective cluster. This technique has been shown to be a relatively quick and efficient way to obtain reasonably valid emotion reports [17].

**Objective performance** was measured by recording for each problem or trial in the task whether the participant responded correctly or incorrectly.

### 3.4 Data Reduction and Parameterization

For purposes of data reduction, each task sequence was subdivided into a series of discrete epochs that were bounded by the self-reports of affective state that were taken periodically during each of the tasks. Self-reports were collected immediately after performing the practice task (i.e., just before beginning the main task), and then were periodically assessed throughout the task. For the anagram task, the first main assessment occurred 2 minutes into the task, and then assessments were repeated every seven minutes thereafter. For the math task, assessments were collected after the first three problems, after the next two problems, and subsequently upon completion of every problem thereafter. For the sound task, the first main assessment was taken
after the first 20 trials of the main task, and every 60 trials thereafter. For analysis purposes the practice task data, and the post-practice self-report assessment were not retained for analysis. For the main tasks the physiological record was divided into discrete problem-solving epochs bounded by the self-report assessments. Each physiological measure was parameterized separately for each epoch, and the parameters corresponding to a given epoch were referenced to the self-report data derived from the assessment immediately following the epoch.

**Skin Conductance.** Several indicators were derived from the measure of skin conductance. A typical skin conductance recording is presented in Figure 2. The skin conductance record can be separated into tonic and phasic components [13]. The tonic component of skin conductance is a relatively slow moving wave that often reflects the person’s overall state of activation or arousal at any given time. In many experimental settings it often shows a distinct downward trend (as in the above figure), as the participant adapts to the research context. Phasic responses reflect event-related reactions that occur in an individual. Occurrences of various types of stimuli produce discrete, relatively brief upward inflections in the skin conductance record. The type of stimuli that evoke skin conductance responses can range from internal stimuli, such as the person’s own thoughts to experimentally presented external events. A phasic skin conductance response (SCR) has a characteristic form, the major components of which are depicted in Figure 3.
In the present case, the data record for this measure was separated into tonic and phasic components, by identifying and extracting from the record all skin conductance responses (defined as upward deflections of skin conductance level greater than 0.05 uSiemens and having a rise-time greater than 250 ms; the duration of the response was defined as spanning from the point of upward inflection to the point of 3/4 recovery, or the instigation of the next response, whichever came first). From this extracted phasic component, we computed the rate of skin conductance responses (in resp/min), the average amplitude of the responses (in uSiemens), and the amplitude of the largest response observed during the epoch (maximum amplitude). From the remaining tonic array, the mean skin conductance level (SCL) and the slope of change in SCL (in uSiemens/Min) across the epoch were computed and retained.

Digit skin temperature was summarized as the mean temperature during the epoch, and the slope of temperature change across the epoch.

**Figure 4. Major components of the cardiac cycle as reflected in the ECG.**

The major components of a cardiac cycle as reflected in the ECG are depicted in Figure 4. Several indicators were derived from the ECG record. First, the interbeat interval (IBI) was computed for successive heart beats by computing the time interval (in ms) between the peak of successive R-waves (representing the point of systole in the cardiac cycle, which is the point at where the ventricles of the heart of contracted maximally to eject the blood in the heart into the arteries, [18]), and the mean IBI, and its standard deviation (IBI SD) were retained.

Second, spectral analysis of heart-rate variability, as reflected in the IBI record, can be used to estimate the influence of both the sympathetic (activation of which serves to increase heart-rate in emergency, or otherwise very demanding situations) and parasympathetic (activation of which slows down heart-rate, and activity of which primarily controls heart-rate variation under most normal circumstances) branches of the autonomic nervous system on heart rate. Parasympathetic activation of the Vagus nerve slows down heart-rate and also increases the variability of heart-rate due to respiration, with heart-rate accelerating during inspiration and decelerating during expiration (this respiration-related variation in heart rate is commonly
referred to as “vagal tone” or “respiratory sinus arrhythmia”). Therefore, by examining the spectral power of IBI variability in the frequency band (known as the “high frequency” [HF] band, 0.15-0.4 Hz) that corresponds to normal respiration, can be used to estimate the influence of parasympathetic influence on heart rate. Spectral power in what is known as the “low frequency (LF) band (0.04-0.15Hz) can also be used to estimate the sympathetic influence on the heart [9,11], although this estimate is not as reliable or as pure as the parasympathetic estimate because slow respiration rate can sometimes contaminate this estimate with parasympathetic activity as well [18].

In the present case, the computed time series of IBIs from the last three minutes of each epoch (for only those epochs 3 min or longer in duration) were subjected to a power spectral analysis to provide estimates of sympathetic and parasympathetic activity, with the observed spectral power in the HF band being used as an estimate of parasympathetic activity, and that in the LF band being as an estimate of sympathetic activity.

Mean relative pulse volume (RPV) and its standard deviation were derived from the finger pulse record by extracting the peak pulse volume associated with each cardiac cycle and averaging and assessing the variability of these peak amplitudes across the epoch. As noted above, by capturing the ebb and flow of blood into the finger tip, this signal can track the cardiac cycle from a distance. In addition, the relative level of blood in the finger at the peak of any given cycle can be used to track both the level (mean) and variability (standard deviation) of the relative constriction versus dilation of the capillaries in the finger tip. Figure 5. depicts a sample data record containing both the ECG and RPV signals.

Figure 5. Sample ECG (upper panel) and RPV (lower panel) Data Records

Exploiting the fact that RPV can be used to track the cardiac cycle from afar, data from this record were combined with the ECG to compute pulse transmission time. This parameter was computed by deriving the time difference between the peak of the R-wave in a given cardiac
cycle and the peak of the subsequent pulse-wave in the finger tip, which reflects the amount of
time it takes the pulse wave to travel from the heart to the finger tip. This measure primarily
reflects the level of constriction versus dilation of the major blood vessels between the heart and
the finger tip. Both the mean and standard deviation of pulse transmission time were computed
and retained for each epoch.

Figure 6 depicts a sample data record containing sample rectified and integrated EMG data
from both the brow (upper panel) and jaw (lower panel) regions. These signals reflect the
electrical activity produced by the contraction of the muscles underlying the EMG sensors, and
thus reflect the degree of tension in these muscles at any given moment. Thus upward
excursions in these signals reflect a momentary tensing of the relevant muscles [19]. For both
the brow and jaw EMG records, both the mean and standard deviation of activity levels observed
in each channel were computed and retained for each epoch.

![Figure 6. Sample data from both the Brow (Corrugator, upper panel)
and Jaw (Masseter, lower panel) regions.](image)

The self-report data were used to derive two affective indices that were of primary interest in
the present analyses: an engagement index and an anxiety index. The engagement index was
computed for each assessment by averaging the item designed to measure how much the
participant cared about his/her performance with the affective items assessing hope, challenge/determination, interest, resignation (reverse coded), and boredom (reverse coded). The
anxiety index was computed by averaging the items assessing anxiety, overload, and calmness
(reverse coded). The reliabilities of these indexes were assessed by computing estimates of
alpha reliability across the epochs associated with all three tasks for each participant separately,
and then averaging these estimates. For the engagement index, the mean reliability estimate was
.71 (range: .60-.84), and for the anxiety index it was .71 (range: .60-.90), indicating that both
indexes were of adequate reliability for research purposes. In addition, to the two affect indexes,
the single item assessing tasks difficulty was also retained for analysis in order to assess the
effectiveness of the intended experimental manipulation of difficulty across the tasks.

Finally, an index of objective performance was computed for each epoch of the three tasks,
and in each case this index could vary between 0 and 1. For the anagram task, this index was
simply the proportion of anagrams solved correctly during the epoch. For the math task, a
correct problem solution was coded as “1,” whereas failing to solve the problem was coded as
“0”. For the first two epochs of the task, which involved attempting to solve multiple problems, the scores were averaged across the problems for that epoch, whereas for the subsequent, single-problem epochs, the performance score was simply the 1 or 0 for that problem. For the sound task, the performance score was proportion of sound trials within the epoch that the participant answered correctly, after controlling for guessing.

3.5 Results of the Preliminary Human Experiments
3.5.1 Manipulation and Validity Checks

The three tasks were designed to systematically manipulate task difficulty across the epochs of each task in order to produce variation in the participants’ subjective states of engagement and anxiety during the performance of these tasks. The three panels of Figure 7 depict the mean subjective ratings of task difficulty, and those of Figure 8 depict the mean objective performance scores, for all six participants across the epochs of the easy vs. difficult sequences of the anagram (panel A), math (panel B), and sound (panel C) tasks, respectively.

![Difficulty Ratings](image)

**Anagram Task (A)**

**Math task (B)**
As can be seen in the figures, both in terms of subjective difficulty, and even more strongly for objective performance, the tasks manipulated difficulty largely as intended. Specifically, for the anagram task, the anagrams in the difficult sequence were generally rated as more difficult than those in the easy sequence, $F(1, 5) = 7.37, p < .05$, and although the interaction of task epoch and difficulty was not reliable for the difficulty ratings of this task, $F(1.23, 6.17) = 2.70, p = .15$ (note, for all effects involving epoch, degrees of freedom have been adjusted using the Greenhouse-Geisser correction), a tendency for the difficulty ratings to become more disparate across between the tasks becomes is somewhat evident from the graphed means. In terms of objective difficulty, not only were the difficult anagrams less likely to be solved overall, $F(1, 5) = 13.50, p < .05$, the spreading of the performance differences across the task was also evident, as indicated by a difficulty by epoch interaction, $F(2.82, 14.11) = 10.60, p < .001$.

Rather similar results were observed for both the math and sound tasks. Although, overall, neither task was rated as being reliably more difficult in the difficult sequence than in the easy one, Math $F(1, 5) < 1, ns$, Sound $F(1, 5) = 2.01, ns$, the difficult sequences were more difficult in terms of objective performance, Math $F(1, 5) = 17.16, p < .01$, Sound $F(1, 5) = 56.64, p < .001$. Moreover, for both tasks, the difficulty by epoch interaction was reliable for both the subjective difficulty ratings, Math $F(1.54, 7.68) = 5.09, p < .05$, Sound $F(2.67, 13.35) = 5.03, p < .05$, and the objective performance index, Math $F(3.21, 16.04) = 5.06, p < .05$, Sound $F(3.27, 16.39) = 27.03, p < .001$. In each case, inspection of the figures verifies that these interactions were due to difficulty tending to increase, and performance tending to decrease across the difficult versions of the task, with difficulty remaining relatively low or decreasing, and performance remaining relatively high or increasing across the easy versions of the tasks, as intended.
Figure 8. Objective performance scores for the easy vs. difficult versions of the Anagram Task (Panel A), Math Task (Panel B), and Sound Task (Panel C)

Given that the manipulation of task difficulty was largely successful for all three tasks, an important further preliminary question is whether this difficulty manipulation succeeded in producing variable levels of task engagement and anxiety in the participants across the various
epochs of the task. Table 1 depicts the means, standard deviations and ranges for both the engagement index and the anxiety index for each participant considered individually.

Inspection of the table indicates that although reports of task engagement were generally higher than those of anxiety, both indexes showed variability across the tasks, with in most cases, scores on the index spanning at least a third of the 9-point response scale (Note, however, that participant #6 showed notably lower and less variable levels of anxiety than most of the other participants). Thus, given that the tasks produced variability in both engagement and anxiety within each participant, there is good basis for examining the extent to which the various physiological parameters assessed in this study are systematically associated with the two affective states.

Table 1. Descriptive summaries for the Engagement Index and Anxiety Index broken down by participant engagement and anxiety indexes for each participant.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Engagement Index</th>
<th>Anxiety Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St. Dev.</td>
</tr>
<tr>
<td>1</td>
<td>5.91</td>
<td>1.53</td>
</tr>
<tr>
<td>2</td>
<td>7.90</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>6.37</td>
<td>0.53</td>
</tr>
<tr>
<td>4</td>
<td>5.50</td>
<td>1.33</td>
</tr>
<tr>
<td>5</td>
<td>7.17</td>
<td>0.66</td>
</tr>
<tr>
<td>6</td>
<td>6.73</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Before doing so, however, the data can also be used to assess whether the self-report indices appear to be valid. As mentioned above, one indication of the validity of the self-reports would be if the self-reported engagement index was found to be associated with objective performance on the tasks. In a similar manner, anxiety might be expected to be negatively correlated with performance, as one might expect participants to become anxious about their performance when the tasks became difficult, and they started performing less well. These correlations were computed for each participant across all of the epochs comprising all three tasks. On average, across all six participants, engagement was found to be positively correlated with performance, ave $r = .40$. In contrast, anxiety was found to be negatively correlated with objective performance, ave $r = -.34$. Moreover, in accord with the reasoning above, anxiety was also found to be positively correlated with the subjective difficulty of the task, ave $r = .43$, a relation that was not observed for engagement and task difficulty, ave $r = -.02$. Thus, the self-reports of both task-engagement and anxiety demonstrated relations with objective task performance that gives reason to believe that these reports have considerable validity.

3.5.2 Identification of physiological correlates of task engagement and anxiety

Given their apparent validity, the participants reports were correlated with the parameters of physiological activity to identify physiological correlates, or markers, of these affective states. Because we anticipated that, due to the phenomenon of person stereotypy discussed above [9], the correlates would vary somewhat from person to person, these analyses were conducted entirely within-persons. That is, for a given participant, the data from all useable epochs across all six experimental sessions were combined, and the reports of anxiety and of task engagement were correlated across the epochs. These analyses were repeated separately for all six
participants. Table 2 reports the correlations of the physiological parameters at or above .23 that were observed with task engagement, and Table 3 reports the corresponding correlations that were observed for anxiety. Across the approximately 50 epochs over which the correlations were computed, this magnitude correlation has a two-tailed alpha probability associated with it of approximately .10. This relatively lenient level of statistical significance was selected due to the preliminary and exploratory nature of these analyses.

**Table 2. Correlations of physiological parameters with Task Engagement**

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symp Power</td>
<td>-.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.34</td>
</tr>
<tr>
<td>Parasymp. Power</td>
<td></td>
<td>.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBI</td>
<td></td>
<td>-.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPV</td>
<td>.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPV-St. Dev</td>
<td>.27</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transit Time</td>
<td></td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCL</td>
<td></td>
<td>.59</td>
<td>.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCR-Rate</td>
<td>.38</td>
<td>.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCR-Ave</td>
<td>.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCR-Max</td>
<td>.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp.</td>
<td>.59</td>
<td>.61</td>
<td>-.37</td>
<td>.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp. Slope</td>
<td></td>
<td></td>
<td></td>
<td>.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrugator</td>
<td></td>
<td></td>
<td></td>
<td>.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corr-St. Dev.</td>
<td>.33</td>
<td>.37</td>
<td></td>
<td>.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masseter</td>
<td>.44</td>
<td>.53</td>
<td>.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mass-St. Dev</td>
<td>.34</td>
<td>.33</td>
<td>-.24</td>
<td>.26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Examination of the tables reveals several notable points. First, a number of correlates, and hence, potential markers, for both task-engagement and anxiety are evident. With the exception of anxiety for participant #6 (who, it will be recalled, demonstrated very little anxiety overall), at least two physiological correlates were observed for each emotion, and in most cases the individual demonstrates substantially more correlates than this.

However, there is considerable evidence of person stereotypy: The specific parameters that are correlated with a particular emotion vary considerably from person to person. Thus, parasympathetic power is positively associated with engagement only for participant #4; skin conductance level is associated with anxiety only for this same individual, and so on. Notably, several of the parameters show similar relations across multiple individuals, which increases one’s confidence that such relations are not spurious. Thus, the rate of skin conductance responses increased with task engagement for both participants #2 and #4; these same two
participants demonstrated increased variability in relative pulse volume for both task-engagement and anxiety; participants #’s 2, 3, and 6 each demonstrated increased variability in brow (corrugator) activity with increasing task engagement, and so on.

Table 3. Correlations of physiological parameters with Anxiety

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symp Power</td>
<td>.40</td>
<td>.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parasymp. Power</td>
<td>.23</td>
<td>.23</td>
<td>.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBI-St. Dev</td>
<td>.51</td>
<td></td>
<td></td>
<td>.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPV-St. Dev</td>
<td>.27</td>
<td></td>
<td></td>
<td>.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transit Time</td>
<td>-.24</td>
<td>.40</td>
<td>-.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCL</td>
<td></td>
<td></td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCR-Rate</td>
<td></td>
<td>.41</td>
<td>-.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCR-Ave</td>
<td>.27</td>
<td></td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCR-Max</td>
<td>.43</td>
<td></td>
<td>.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp.</td>
<td>-.84</td>
<td></td>
<td></td>
<td>.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp. Slope</td>
<td></td>
<td></td>
<td>.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrugator</td>
<td></td>
<td></td>
<td>-.47</td>
<td>.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corr-St. Dev.</td>
<td>.23</td>
<td>.27</td>
<td></td>
<td>-.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masseter</td>
<td></td>
<td>.29</td>
<td>-.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mass-St. Dev.</td>
<td></td>
<td>.26</td>
<td>-.39</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

At the same time, however, no single parameter was associated with either task engagement or anxiety in all six participants, and in several cases observed relations involving a given parameter were actually in opposite directions across individuals. Thus, for example participants #4 and #5 showed rather different patterns of facial activity while anxious. Whereas participant #4 tended to show reduced brow tension and increased and variable jaw tension with increasing anxiety, the pattern was reversed for participant #5. Similarly, whereas digit skin temperature increased with increased task engagement for participants #1, #2 and #5, temperature decreased systematically with increasing engagement for participant #4, and so on. It is beyond the scope of the present investigation to determine why these individuals are showing such different patterns of activity for similar affective states. However, these findings clearly validate our supposition that individual differences in response patterning (i.e., due to person steroptyp) would be a major issue that would need to be confronted in the development of an affect recognizer, and they clearly support our methodological decision to adopt a person-specific approach in our efforts to do so.

It is also notable, that within individuals, the observed patterns for task-engagement and anxiety are somewhat different. This is notable because both anxiety and task-engagement are
“high arousal” states in which the individual is mobilizing for action, albeit for different purposes – self-protection and the avoidance of harm in the case of anxiety versus task performance and mastery in the case of task-engagement. Thus it is to be expected the arousal patterns associated with these two states would show some similarities, which, in fact, they do. For example, there is a clear tendency across participants for SC activity to be positively associated with both task engagement and anxiety. However, with the exception for participant #4, for whom the patterns of associated physiological activity are quite similar for the two states, it is notable that for each individual somewhat different parameters are associated with the two states. This result indicates considerable potential for us to be able to develop an affect detector that will be able to reliably distinguish between different affective states for a given individual.

Finally, it should be noted that virtually all correlations depicted in the table are modest in magnitude. One likely reason for this is that, in retrospect, our assessment procedures in this pilot work had poor temporal resolution. In assessing subjective state we asked participants to describe their states for the previous task epoch, which could have lasted up to 7 minutes, and in referencing the physiological data to the self-reports, we summarized the data over these same intervals. It is possible for one’s affective state to change appreciably over such intervals, and to the extent to which it did, our “snapshots” of both subjective state and physiological activity would have been blurred, and this blurring would have attenuated the observed correlations between subjective state and physiological activity. Thus, using techniques to yield better temporal resolution of our parameters should reveal stronger, and potentially more numerous, relations between subjective affect and physiological activity than we have obtained. This is one of the issues to be addressed in our Phase II research that we outline at the end of this report.

To summarize the results of these initial studies, we have documented the existence of reliable associations between subjective emotion and physiological activity that can be potentially used by an affect recognizer to infer an individual’s affective state. However, as suspected at the outset, the specific patterns associated with a given state appear to be quite person-specific in a way that necessitates the adoption of a person-specific approach to the development of such a recognizer.

3.6 Affect Recognition
Given that we had successfully identified and provided preliminary validation for potential physiological to be used in affect detection, we decided to try to demonstrate the feasibility of an implicit human-robot communication system by developing a real-time experiment in which a robot would be able to detect a human’s affective state based on incoming physiological data, and then would modify its own behavior accordingly. As an initial step toward building this experiment, we needed to develop an affect recognizer that could infer the human’s emotional state from the physiological data. To do this, we needed to decide upon a suitable signal processing approach. That is, we needed to decide upon a signal processing approach by which a robot, in real time, could parameterize the incoming physiological signals from the human, and then combine these parameters to infer the human’s affective state. We have not yet made a final commitment to a particular information processing approach, and as we outline below, we intend to examine and compare several such approaches in our Phase 2 research. However, for the purposes of demonstrating the feasibility of a human-robot communication system, we adopted a fuzzy logic approach to interpreting the physiological signals. The next few sections review the workings of a fuzzy logic inferential system.
3.6.1 Fuzzy Logic –Theoretical Background

We chose to work with fuzzy logic because of the nature in which humans shift from one affective state to another in real life. Such transitions are neither abrupt nor crisp. Instead the transition of one state, such as relaxation, to another, such as anxiety, can often be gradual, and in many transitions, the person may be simultaneously experiencing multiple states, such as anger and anxiety \[20\]. Such considerations make classical set theory a poor candidate for modeling such transitions, because according to such theory, if we consider the transition from relaxation to anxiety, a person can either be relaxed or anxious at a given instance, s/he cannot be both. Also the transition from one set to another is rather abrupt which does not happen in real life. In such a case, fuzzy reasoning can prove invaluable in making the decision-making process resemble human reasoning. Figure 9 captures the advantage of using fuzzy logic over boolean logic in determining the affective state of a person.

![Figure 9. Comparison of Boolean and Fuzzy logic](image)

Fuzzy logic is based on the theory of fuzzy sets. Fuzzy set theory implements classes or groupings of data with boundaries that are not sharply defined (i.e., fuzzy)\[21\]. A fuzzy set can

![Figure 10. Overview of a fuzzy logic system](image)
contain elements with only a partial degree of membership. This enables fuzzy models to exercise flexibility in capturing various aspects of vagueness in the data available to us [22-23]. Fuzzy set theoretic methods have been used extensively for pattern recognition in the past.

The design and implementation of a fuzzy model involves the following steps [24-26]: (1) Specifying the input and output variable membership functions, (2) Fuzzification of the input variables, (3) Defining the rule statements that relate the input variables to the output, (4) Aggregating all outputs (Fuzzy Inference), and (5) Defuzzification of the output variable. Figure 10 depicts the structure of a fuzzy logic system. Each of the steps that are involved in the implementation of a fuzzy logic system are discussed below.

3.6.2 Definition of membership functions

A Fuzzy set \( F \) in a space of points \( S = \{s\} \) is a set of elements with a varying grade of membership and is characterized by a membership function \( M_F(s) \) that maps each element of \( S \) to a real number in the interval \([0, 1]\). The value of \( M_F(s) \) for any given \( s \) indicates the degree of \( s \) in \( F \) or the degree an \( s \) belongs to \( F \) [26]. The value 1.0 indicates complete inclusion of \( s \) in \( F \), the value of 0.0 indicates complete exclusion of \( s \) in \( F \), and the intermediate values indicate partial inclusion. There is a wide variety of membership functions built from piecewise linear functions, the Gaussian distribution function, the sigmoid curve, or the quadratic and cubic polynomial curves as these standardized functions have adjustable parameters. The only constraint for a membership function is that it must vary between 0 and 1. The function itself can be an arbitrary curve whose shape we can define as a function that suits us from the point of view of simplicity, convenience, speed, and efficiency.

3.6.3 Fuzzification of inputs

Fuzzification of inputs is necessarily determining the degree to which they belong to each of the appropriate fuzzy sets via membership functions [23]. We resolve the inputs into a number of different fuzzy linguistic sets. For the particular fuzzy system that we are interested in, the inputs are various physiological responses. Before the rules can be evaluated, the inputs must be fuzzified according to each of the linguistic sets. For example, to show the extent to which a given parameter indicates anxiety, we can fuzzify it into the following three sets: least anxiety, medium anxiety and high anxiety. Figure 11 shows the fuzzification of the input tonic level of skin conductance.

![Figure 11. Fuzzification of Skin Conductance Level](image-url)
3.6.4 Definition of rules

The next step is to define rules that relate the inputs to an output. The parallel nature of the rules is one of the most important aspects of fuzzy logic systems. The transition from a region where the system’s behavior is dominated by one rule to a region where another rule dominates is smooth, avoiding sharp switching between modes based on breakpoints. A single fuzzy if-then rule assumes the form:

\[
\text{if } x \text{ is } A \text{ then } y \text{ is } B
\]

Where A and B are linguistic values defined by fuzzy sets on the ranges (universes of discourse) X and Y, respectively. The if-part of the rule "x is A" is called the antecedent or premise, while the then-part of the rule "y is B" is called the consequent or conclusion. The antecedent of a rule can have multiple parts connected by the logic “AND” or “OR”, for example:

if \( x_1 \) is A and \( x_2 \) is B then \( y \) is C

Every rule has a weight (a number between 0 and 1), which is applied to the number given by the antecedent. Generally this weight is 1 (as it is for this example) and so it has no effect at all on the implication process. From time to time we may want to weight one rule relative to the others by changing its weight value to something other than 1.

3.5.5 Aggregating Output (Fuzzy Inference)

Because decisions are based on the testing of all of the rules, the rules must be combined in some manner in order to make a decision. Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set [25]. The output of the aggregation process is one fuzzy set for each output variable. All the rules are evaluated together and the output of each rule is combined, or aggregated, into a single fuzzy set whose membership function assigns a weighting for every output value.

3.6.6 Defuzzification of the output

The defuzzification process transforms the fuzzy set (the aggregate output fuzzy set) into a single number. The aggregate of a fuzzy set encompasses a range of output values, and so must be defuzzified in order to resolve a single output value from the set. This defuzzification process could employ methods like centroid, bisector, middle of maximum (the average of the maximum values of the output set), largest of maximum, smallest of maximum and other such criteria.

Figure 12 shows how the output of each rule is determined and outputs from all the rules are defuzzified to get the anxiety index of 0.336.

In the figure, for each parameter, the red line running through the distributions reflects the amount of activity observed for the parameter at a particular time for which we are trying to estimate the person’s anxiety level. This activity level intersects with two membership distributions for the parameter corresponding to “not anxious” (upper distribution) and “anxious” (lower distribution). The probability that the observation is in one or the other of these classes is determined by where the activity level intersects the membership distribution. These probabilities are then translated to output estimates for each channel of the degree to which the person is anxious (the rightmost column of the figure) via the output rules (not depicted) for each parameter. These individual output estimates are then aggregated according to a predefined aggregation scheme (not depicted) to yield an overall estimate of the person’s anxiety level (the bottommost box of the output column).
3.6.7 Designing an affect (anxiety) detector using fuzzy logic

Having settled upon fuzzy logic as our affect-detection inferential tool, our next step was to use fuzzy logic to implement an actual affect recognizer. For this implementation we selected six physiological parameters to focus on, based on their ability to differentiate between states of high and low anxiety for a particular individual (Participant #2, see below). In the implementation, these six parameters are computed in real time, based on incoming physiological signals, and then the parameters are combined in a fuzzy logic analysis as described above, and a prediction is generated regarding the person’s affective state (in this case, how anxious the person currently is). In the human-robot experiment, this prediction would then be available for use by the robot’s controller.

Selected physiological data from Participant #2 in the studies described above were used in the development and implementation of this affect detector. Specifically, six parameters were selected for analysis based on the anxiety-related differences they demonstrated in a comparison of the epochs in which this participant reported the highest and lowest levels of anxiety, respectively. These parameters were sympathetic and parasympathetic power, derived from a spectral analysis of the ECG, the rate and average amplitude of phasic skin conductance responses, and both the variability in the brow region EMG and mean tension level in the jaw region EMG. The data that provided the basis for selecting these parameters are depicted in Figures 13-15.

Cardiac activity was a strong indicator of anxiety. The power in the sympathetic activity frequency range increased and the power in the parasympathetic activity frequency range decreased significantly as the human-subject showed anxiety (Fig. 13a and 13b, respectively).
Fig. 13 Power of sympathetic activity (a) and parasympathetic activity (b) of heart in relaxed and stressed states (msec²)

There were marked increases in both the mean amplitude (Fig. 14a) and rate (Fig. 14b) of phasic skin conductance responses with increases in anxiety.

Fig. 14 (a) Mean level of phasic activity and (b) Rate of response of phasic activity in relaxed and stressed states

Finally, Muscle tension was generally higher in the brow region (Fig 15a), and more variable in the jaw region (Fig 15b) as anxiety increased.

Fig. 15 (a) Mean activity of the corrugator muscle and (b) Variability of the masseter muscle
Using these parameters, five blocks of Participant #2’s data were selected to use to fine-tune the weight of each rule and the shape of the membership functions that comprised the fuzzy logic analyzer until very accurate results were obtained. Table 4 shows the results of testing the fuzzy logic system on the five selected blocks of this participant’s data. The self-reports of the participant are also indicated in the table to illustrate the close correspondence between the anxiety index generated by the fuzzy system against the participant’s own self-reported anxiety.

<table>
<thead>
<tr>
<th>Physiological Parameter</th>
<th>Epoch 1</th>
<th>Epoch 2</th>
<th>Epoch 3</th>
<th>Epoch 4</th>
<th>Epoch 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power (sympathetic activity)</td>
<td>84023.02</td>
<td>54180.64</td>
<td>278221.78</td>
<td>41628.53</td>
<td>319226.20</td>
</tr>
<tr>
<td>Power (parasympathetic activity)</td>
<td>28142.35</td>
<td>48625.93</td>
<td>40523.58</td>
<td>4762.10</td>
<td>29481.15</td>
</tr>
<tr>
<td>Mean phasic SCR Amplitude</td>
<td>0.8742</td>
<td>0</td>
<td>0.66586</td>
<td>1.3795</td>
<td>0.7285</td>
</tr>
<tr>
<td>Phasic SCR rate</td>
<td>6.86</td>
<td>0</td>
<td>8.38</td>
<td>7.31</td>
<td>6.98</td>
</tr>
<tr>
<td>Mean of the corrugator activity (EMG)</td>
<td>1.22</td>
<td>11.02</td>
<td>9.02</td>
<td>7.94</td>
<td>6.80</td>
</tr>
<tr>
<td>Variability of the masseter activity (EMG)</td>
<td>2.4153</td>
<td>1.3856</td>
<td>2.9474</td>
<td>1.9658</td>
<td>2.3959</td>
</tr>
<tr>
<td>Self report of the participant (scale of 7)</td>
<td>2.00</td>
<td>2.66</td>
<td>4.33</td>
<td>4.67</td>
<td>5.67</td>
</tr>
<tr>
<td>Anxiety index (scale of 1)</td>
<td>0.31</td>
<td>.41</td>
<td>0.62</td>
<td>0.63</td>
<td>0.65</td>
</tr>
</tbody>
</table>

4 Robot Control

The final consideration we needed to address in our efforts to implement a feasibility demonstration of implicit communication between a human and robot was to decide upon a control architecture to use in the robot.

In the demonstration we wanted to experiment under laboratory conditions a human-robot system working in close coordination on an exploration task. An real life situation analogous to this may be an astronaut and a mobile robot exploring a planet, each carrying out separate tasks in a coordinated fashion, with the robot being responsive to the astronaut’s affective states (for instance stress, panic or fatigue). In this scenario a mobile robot is navigating a workspace, while its human companion, whose physiological state is being continuously monitored, is engaged in some other task.

With this scenario in mind, and following a growing consensus among robotics researchers, we decided to develop a hybrid architecture that would combine both deliberative and reactive control paradigms. Deliberative control strategies require a relatively known world model. The world model is used to predict the outcome of the control actions and optimize the performance. This type of controller is highly effective in a structured environment. Reactive control techniques, on the other hand, are suitable for dynamic and unstructured worlds. Perceptions and actions are tightly coupled in this paradigm. There are several reactive control architectures that have been shown to work well in real life situations.

We reasoned that affect-sensitive robot control, where the robot is expected to perform its own task in a relatively unknown and unstructured environment, as well as to monitor the affective states of the co-working human, would naturally lend itself to a hybrid control architecture. Consider a human-robot planetary exploration task or a human-robot space station repair and maintenance task as examples. It is unlikely that we will have a precise world model in such cases. The environment will be dynamic, unstructured and to some extent, uncertain. As
a result, a detailed planner, which is the backbone of a deliberative controller, will be largely infeasible. A well-designed behavior-based reactive controller will be better suited for such tasks. However, when the robot is also expected to “look after” the human based on its inference of the human’s affective states, the robot will need a reasoning mechanism that will analyze the physiological signals to infer affect and determine a possible course of action. Given a good model of the human’s characteristic physiological affective patterning, this monitoring task is well suited for deliberative control. Consequently deliberation must be integrated into the controller along with the reactive components.

Motivated by this reasoning, we embarked on the design of a hybrid control architecture for the proposed human-robot cooperation activity. The reactive part of our controller was designed based on Subsumption architecture that was originally proposed by Brooks [27-29]. In a subsumption architecture task achieving behaviors are represented as separate layers. Individual layers work on individual goals concurrently and asynchronously. Complex actions subsume simpler behaviors. A priority hierarchy determines the topology. The lower levels in the architecture have no awareness of higher levels. This is conducive to incremental design where higher-level behaviors can be added without disturbing the lower-level behaviors. These advantages of subsumption architecture are very useful for our proposed human-robot coordination. The Phase I work was designed to demonstrate the feasibility of designing an affect-sensitive controller. We started with a simple exploration task for the robot. A set of simple behaviors were developed and programmed for this purpose. However, more complex task behaviors can be added to this controller without destabilizing this working controller in Phase II work.

The deliberative part of our hybrid controller was designed to infer the affective states of the human. All the necessary physiological signals are processed and used to infer the affect. The robot controller senses the degree of the affect and determines a course of action. This whole process was embodied in a layer called the affect layer and was integrated within the subsumption architecture as a middle layer.

The top most layer was again a reactive layer called the survival layer, which completed the current control architecture. The idea behind this layer was to ensure the survival of the robot before it would attempt to either sense the affect of the human or perform its own tasks.

The specific behaviors we programmed into each layer were determined based on a consideration of the capabilities of the robot we would be using for our demonstration. The robot we used was the mobile robot-Trilobot [30]. Given this robot’s sensing capabilities, we gave the robot the following tasks:

- Wandering in a random manner exploring the workspace
- Avoiding obstacles in the workspace
- Responding to the affect signals of the human companion

At any given time the robot’s sensors map its environment onto its memory. The sonar range finder and the touch sensors give information regarding the obstacles in the workspace, the compass indicates the orientation of the mobile robot, the optical encoders indicate the motor speed and distance traveled and the physiological sensors give an indication of the affect state of the human. The priorities with which the robot performs its various tasks based on the information it receives is determined by the subsumption control architecture we developed. The complete control architecture is presented in Fig. 16.
In Figure 16, the lowest priority task for the robot is wandering. It can be subsumed by the wall following behavior, which in turn can be subsumed by the obstacle avoidance behavior. All three behaviors are reactive behaviors and form the basis of an exploration task. The letter “S” represents the signals or response suppressing ability of a higher layer so that the lower layer may not be activated if required. The middle layer, called the affect layer, is a signal processing and fuzzy logic reasoning block that is purely deliberative. It determines whether the human needs help and how to respond to this need. This layer has a higher priority than the lower exploration task layers simply because we put a higher weight on the robot’s capability to respond to human’s affective need. The top most layer is again a set of reactive behaviors designed to help the robot survive in an uncertain and dangerous environment. The whole system will work as follows: the robot will perform exploratory task if there is no reason to attend to the human affective need and there is no danger to its own survival. If it senses that human requires its help, the exploratory tasks will be temporarily suspended and the robot will attend to the human. If there is any danger to its own survival, the robot will disregard any other activity and perform its survival actions. However, once the danger is taken care of, the robot will come back to its original exploratory tasks.

This architecture was implemented in the Trilobot through the use of three triggers that were continuously generated using the feedback from the robot’s environment: a survival trigger (which when triggered indicated any immediate danger to the robot’s own safety at any given instant), an affect trigger (which when triggered indicated that its human companion was getting anxious) and a wander trigger (which when triggered indicated that the mobile robot was expected to be in its wander mode, i.e., carrying on the exploration task unimpeded). By default the mobile robot was programmed to be wander triggered at all times.

At any given instant the robot received one or all of the three triggers. The decision regarding the trigger to be processed first was determined by the subsumption architecture. A trigger that was being processed caused the robot to go into that particular mode. For instance if the robot

![Figure 16. A hybrid subsumption control architecture for human-robot cooperation](image-url)
was processing the affect trigger then the robot would go into the affect mode in which it would do any particular task or sequence of tasks that enabled it to respond to the human’s state in the best possible manner. The affect mode included tasks like assisting the companion in navigation, querying its companion about his/her well being, locating its companion in the workspace and returning to him/her or simply raising an alarm.

5. A Working Demonstration of Human-robot Affect-sensitive Cooperation

In order to provide a demonstration that attempted to mimic the real-time functioning of the envisioned human-robot communication system, the fuzzy-logic-based affect recognizer was combined with the robotic control architecture. For this experiment the fuzzy logic analyzer was implemented in MATLAB and it was programmed to read in the selected data from Participant #2 as if they were being obtained in real time. The analyzer then parameterized the incoming data stream, and these parameters were subjected to a fuzzy logic analysis that updated its prediction of the participants’ anxiety level every 30 sec. This output was then fed to the robot’s controller. A diagrammatic representation of the above-described experimental setup is shown in Figure 17.

For the purposes of demonstration, the data being fed into the affect detector was known to contain three instances where the physiological patterning had originally been associated with the participants’ reports of elevated anxiety. The two critical questions of interest in this demonstration were: first, whether the fuzzy logic analyzer would correctly identify these instances as reflecting elevated anxiety, and if so, whether the robot would successfully modify its behavior accordingly.

Figure 18 depicts an actual timing diagram obtained from one of our experiments using the data from Participant #2. As seen in the figure, in absence of any affect or survival trigger the only trigger that was active was the wander trigger, as a result of which the robot stayed in the wander mode. On three occasions (circled in red), the survival trigger became active (in these cases because the robot sensed that it was about to run into a wall or other obstacle). In each
case, as soon as the survival trigger was received, it suppressed the wander trigger and the robot went into survival mode: It suspended the wandering behavior it had been performing, and engaged in survival behavior. That is, the robot diagnosed the nature of threat it was facing and took steps to get out of that threatening situation. In these cases, the robot backed away from the obstacles it was sensing and turned to change its direction of wandering, and in two of the three cases, having attended to its survival needs it returned to wandering mode.

As can be seen from the timing diagram, through the fuzzy logic analysis the robot was able to detect the anxious responses of Participant #2, which activated the affect trigger. On the first and last occasions, the affect trigger fired while the robot was in wander mode. Thus, in a fashion parallel to the firing of the survival trigger, in these cases, the robot suspended its wandering and went into affect mode to respond to the needs of the operator. For this demonstration, the robot was programmed to query the operator as to whether she needed assistance. When informed that she did not, the robot returned to wondering mode. In the second instance, the affect trigger fired virtually simultaneously with the survival trigger. In this

Figure 18. State Diagram of the Trilobot during an implicit human-robot communication experiment.
instance, the survival trigger took precedence, and the robot went immediately into survival mode to deal with the obstacle it was sensing. Once it had done so, however, because the affect trigger was still active, the robot went into affect mode rather than returning to wander mode, and queried the operator as to whether she needed assistance. As indicated in the diagram, it was only after the robot received appropriate feedback from the operator that it once again retu

6 Brainwave Monitoring

6.1. Survey and Cost Benefit Analysis of Existing Technologies

Four different modalities for assessing brain function and activity were investigated, they are Functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), Magnetoencephalography (MEG), and Electroencephalography (EEG). The goal was to determine which method would be the most appropriate to measure brain function on an astronaut in real time and to make determinations about the individual’s state and needs in a near instantaneous time frame.

Functional Magnetic Resonance Imaging (fMRI)

Functional Magnetic Resonance Imaging is the newest modality that was examined. MRI technology is approximately 30 years old and functional imaging was not discovered until 1990.

Brain activation results in an increase in cerebral blood flow (CBF) and oxygen consumption, it is this last effect that is most often exploited in fMRI studies. As oxygen is withdrawn from blood the magnetic properties of the blood are altered. Deoxyhemoglobin (blood lacking oxygen) is paramagnetic, however oxyhemoglobin (oxygen-rich blood) is not. Therefore, higher levels of deoxyhemoglobin results in less signal being detected by the MRI pickup coil. The normal human brain consumes about 40% of the oxygen sent to the brain [31]. During brain activation the quantitative amount of oxygen consumed increases, however the percent of usage actually decreases. Increased activation and oxygen consumption triggers a corresponding increase in CBF. This increase in CBF more than offsets the oxygen consumption and the blood in the activated region actually contains more oxyhemoglobin then while at rest.

The greatest strength of fMRI is the excellent spatial resolution, which is currently at 1-mm [31]. This allows for very precise localization of activation areas within the brain while performing specific tasks. The high level of spatial resolution comes at the expense of temporal resolution. Temporal resolution is limited by the speed of metabolic processes to about 1 second at the absolute best. However this level of temporal resolution would never be utilized because the spatial resolution would decrease to unacceptable levels. As the temporal resolution increases the spatial resolution decreases and vice-versa.

MRI images are created by dividing the brain into separate slices; with each slice being imaged separately. This means getting a scan of the entire brain must be done one slice at a time. This process is time consuming, for example to obtain a complete scan of the brain with spatial resolution at 1 mm requires between 5 and 10 minutes. High levels of spatial resolution require thinner slices and longer imaging times for each slice, as the magnetic field applied is increased in length as well as the resting time between images. A compromise is therefore reached between the two domains and presently whole brain volumes can be imaged with acceptable spatial resolutions in 4 minutes [31]. This delay in full brain imaging means that sudden activity and changes can be missed by this method. For this reason current experiments place subjects in a particular state or physical activity for several seconds to minutes. In a real
time analysis situation missing a sudden or momentary activation may be unacceptable and result in misclassification of a subject’s needs.

A drawback to fMRI is that it does not directly measure the neural activity but instead it measures an effect of neural activity, oxygen consumption and CBF. Therefore there is an inherent delay between neural activity and changes in blood behavior. This delay is believed to be between 1 and 2 seconds from neural firing until changes can be seen in oxyhemoglobin levels \[32\].

MRI machines are in no way wearable and probably never will be. Powerful superconducting magnets must be used to create large uniform magnetic fields around the subject. These magnets must be kept at superconducting temperatures through the use of liquid Helium. The size of the dewar to store the Helium and the magnet require far too much space to imagine that a wearable system will be developed, furthermore there is no push in the industry to try to create such a system.

Functional MRI provides high levels of spatial resolution. Low temporal resolution, non-wearability, as well as a high cost however offset this advantage.

**Positron Emission Tomography (PET)**

PET uses positron emitters to measure regional cerebral blood flow (rCBF). This technology takes advantage of the fact that many of the elements within the human body have positron-emitting radioisotopes. These emitters are injected into the body and as they decay the energy given off can be detected by detectors surrounding the subject’s head. Images are then created from the collected data showing the distribution of the emitters within the brain, higher accumulation indicates more activity. The specific emitter selected determines what is to be monitored, PET can detect blood flow, and oxygen or glucose metabolism. These three parameters can all be used to assess brain function.

PET provides excellent spatial resolution on the order of millimeters however the temporal resolution is quite low. The poor temporal resolution is caused by the delay in changes of blood flow within the body that is associated with increased neural activity. This lag in blood flow change is unavoidable and is limited by metabolic rates, which cannot be changed and are in the range of 1 to 2 seconds \[31,33\]. The shortcomings of fMRI’s temporal resolution are the same as those for the PET. Sudden and short-term activations may not cause changes in rCBF and therefore will not be detected by PET.

In a real time monitoring situation a subject’s brain activity would have to be monitored constantly and over an extended period of time. To accomplish this using PET, the subject would have to be supplied a continuous or nearly continuous supply of emitters. The use of constant injections or a drip IV does not seem to be a very viable delivery method. Another approach to deliver positron emitters continuously would be to provide the emitter in the air supply with labeled CO\(_2\) \[33\]. The body would then reach a steady state of emitter concentration and changes in emitter concentration to regions can be tracked. However over extend periods of time of measurements (hours) there will be an ever increasing risk to the subject caused by exposure to the radiation given off by the positron emitters.

Another issue with PET is the production of the positron emitters that are used. These radioisotopes have short half-lives (as low as 120 seconds to 110 minutes). In a space environment either large amounts of these materials would have to be transported on the mission or a Cyclotron, which produces these positron emitters, would have to be on the voyage as well.
Current PET systems by their construction will not allow for measurements during dynamic movements and do not offer wearability. The subjects must lay flat and still while they are placed inside the ring of detectors. At this time it does not seem reasonable to imagine that a system in the future will be developed that will be able to be used in an ambulatory environment.

**Magnetoencephalogram (MEG)**

Cohen recorded the first recordings of magnetic signals from the brain in 1972 using a SQUID (Superconducting QUantum Interference Device) magnetometer inside a magnetically shielded room. The first full head scanning system was developed in 1992 and present systems may have up to 306 channels.

Magnetic fields produced within the brain are generated by intracranial currents. These currents are the same currents that produce the voltages recorded on the scalp surface using the EEG. Therefore there is strong correlation between EEG and MEG recordings. The strongest magnetic component is perpendicular to the direction of the current as demonstrated by the Biot-Savart Law and illustrated by the right hand rule. Figure 19 shows a diagram of a current and the associated magnetic field.

In general the recent trend in MEG advancement has been working toward full head coverage and the addition of more detection channels. Advancements of the past 5 years have brought about complete helmet systems with an ever-increasing number of channels. The present systems have channel numbers on par with EEG, this should allow for other avenues of development to be pursued, such as noise cancellation and cryogenic improvements. The main disadvantages of the MEG are the need for near noiseless or at least noise reduced recording environments, the near 0 K operating temperature for the electronics, and the extremely high cost of the systems. These areas will be explored further and current work in these areas will be discussed.

Magnetic fields from the brain are extremely small ($10^{-13}$ Tesla or 100 million times smaller than the Earth’s magnetic field) and great care must be taken in order to measure them. Shielded rooms are constructed from μ metal and Aluminum to attenuate the environmental noise from the earth’s magnetic field as well as electronic noise. Shielded rooms are extremely expensive; depending on the size of the room the cost will vary. However a reasonable estimate is $300-400k$ [34]. As well as being costly these rooms are also heavy and can weight as much as 7 tons [34]. Therefore selection of a suitable location becomes more difficult because the structure must be able to support such large weights.

![Figure 19: Current and associated magnetic field](Picture taken from CTF Inc.)
Presently noise cancellation systems are being used in addition to the shielded rooms. Noise subtraction can be accomplished using noise detector arrays [35]. These arrays are setup to detect noise in the environment around the source. The noise array is placed far away (relative distance) from the source, the theory is that these noise channels will detect the environmental noise and not the signal of interest. The detectors near the source should then detect the same noise as the noise array as well as the signal of interest. Signal subtraction can then be performed using the noise array data and the source data to collect a noise reduced or noise absent signal. This analysis is presently performed off-line, post data acquisition. This method is beneficial in a shielded environment however in an unshielded environment the noise is so great that it completely masks the signal of interest and the signal cannot be recovered from the source detectors.

Another method of noise reduction is known as active shielding. Active shielding is presently being used to reduce noise in shielded environments. This method uses a coil system in addition to a noise array to reduce noise. In this system the noise array detects the environmental magnetic noise present and the coil system is used to create an identical magnetic field around the source. The source detectors use a common mode rejection scheme to then eliminate the induced magnetic field from the source signal, and thus the environmental noise. A system such as this is presently being used by Nowak in Berlin. Using a 4 layer shielded room and an embedded coil system they have achieved attenuation factors in the 50-dB range for near DC fields and 20 dB for power line frequencies [36].

These shielding systems are encouraging advances in technology however for MEG to become truly useful it must be functional outside of a shielded room. This will require advances in active shielding and noise subtraction techniques. Reasonable estimates of such a system may be 20 years in the future [37].

Low Temperature SQUIDs (LTS) require liquid Helium cooling in order to operate. LTS are employed in nearly all MEG systems with few exceptions, a few systems use High Temperature SQUIDs that only require liquid Nitrogen for cooling. Liquid Helium is expensive at around $5.00 per liter. Liquid Helium also has a relatively short shelf life with most storage dewars allowing 1-2 liters of boil off per day. MEG systems allow much greater rates of boil-off in the range of 5-12 liters per day depending on the particular system. The required use of liquid Helium makes the operating costs of an MEG quite high. Assuming $5.00 cost per liter and 5 liters boil off per day the cost of keeping an MEG system cool everyday of the year would be ~$10,000. This estimate however is very unreasonable, this estimate assumes 100% transfer between the storage dewar and the MEG dewar, in reality this transfer rate is much closer to 75 or 50%.

Presently HTS are not used because they do not have the same high level of sensitivity that LTS have. If HTS could be improved to the levels of performance of their counter-parts this would greatly reduce the operating costs. Liquid Nitrogen is about 10 times less expensive then liquid Helium. HTS will probably catch up to the present levels of sensitivity found in LTS however when this will happen is difficult to estimate.

The most advanced MEG systems providing whole head coverage and 300 channels cost approximately $2 million [37]. A minimalist system that would still provide whole head coverage but only contain 37 channels could bring the cost down to about $500,000 [37]. These cost are at least 20 times that of EEG systems as will be shown below. This price range would put these systems in about the same price range as MRI and somewhat cheaper than PET.
**MEG Advantages:** Magnetic fields are not attenuated and altered by the skull and tissues of the brain in the same way as electric currents. Electric signals are subject to the factor of volume conduction. When a current is generated within tissue this primary current also creates secondary currents within the surrounding tissue. These secondary currents make their way to the surface of the scalp and result in a voltage being recorded by an EEG electrode. Tissue surrounding a source could be composed of glial cells, cerebrospinal fluid or many other possible tissue types each of which has a unique conductivity and geometry. The skull itself is also quite variable in terms of shape and thickness. These irregularities combine to result in a smearing of the signal across the surface of the scalp. The EEG is indirectly related to the neuronal source because it is not a measurement of the actual primary current but rather a measurement of the volume currents that arise from the initial current.

The MEG does not offer a complete fix to this spatial resolution problem but it does provide an improvement. This improvement is possible because of the differing properties of electric currents and magnetic fields. The magnetic permeability of living tissue is approximately equal to that of free space and little to no distortion of the magnetic fields occur as they travel from their source to the scalp. The spherical shape of the skull allows for magnetic fields perpendicular to its surface to be easily recorded without the distortion that is associated with volume currents. This allows for a direct measurement of the primary currents rather than an indirect measurement of volume currents with the EEG. The end result is better source localization with the MEG over the EEG, MEG can achieve spatial resolutions of 3-mm [38].

**Electroencephalogram (EEG)**

The study of the human EEG over 80 years old. It began with single channel recordings and has progressed to today’s system’s which use as many as 256 different locations on the scalp to record activity. Of the four modalities being investigated EEG is by far the best known and has the most information available on its use. Since its discovery in 1929 predictions and questions were already being asked about whether the EEG had the power to predict one’s thoughts, intelligence, or emotions.

The simple interpretation of the EEG is to imagine it as specialized voltmeter. Metal electrodes on the scalp detect voltages produced by the currents of neurons within the cortex. These electrodes record voltages when large numbers of neurons are activate at the same time; a measurable source is approximately $10^6$ neurons activating at once [39]. This large area of activation is also needed by the other modalities for measurements to be made. Some of the disadvantages of EEG will now be discussed.

EEG is more invasive then MEG, EEG requires direct contact with the subject for measurements to be taken. With MEG recordings the subject must be very close to the detector, approximately 1-cm, however the detector does not need to touch the person. EEG electrodes must be positioned correctly on the scalp and then attached. This can result in long preparation time before data can be taken. The sites on the scalp usually must be prepped using abrasive gels and cleaners to lower the impedance of the scalp as much as possible before the electrodes can be applied. Furthermore after skin preparation the electrode must be firmly secured to the scalp using conductive pastes. Depending on the number of channels being used this prep time can range from just a couple of minutes up to an hour.

EEG recordings are always electrically referenced to a location on the body; common places may be the forehead, ear, or the center of the head. Because this reference location has not been
standardized, comparison between data using different reference locations is not possible [40]. This can result in confusion when comparing results from other researchers as well as the need to repeat previous procedures using different referencing schemes.

EEG however has many advantages over the other modalities. Perhaps the greatest advantage is the cost. The most advanced EEG systems using 256 channels and advanced signal processing software will cost near $200,000. However a simpler wearable system with 20 to 32 channels for data collection is between $15,000 and $25,000 [41]. This is at least 20 times less expensive then MEG systems, as well as MRI and PET systems, which cost several million dollars.

Furthermore the fact that wearable or ambulatory EEG systems presently exist sets it apart from all other modalities. EEG ambulatory systems have been designed for day-long studies on epilepsy patients. The systems are battery powered and can run for up to 48 hours without loss of power and the data can be saved onto memory cards in the system. Presently no system has a radio transmitter to send data wirelessly to a workstation, however it would seem reasonable that this sort of technology would not be difficult to implement on a wearable systems. Xltek Inc manufactures one such wearable product. Their system has the option of 24 or 32 channels of data collection and the ability to store up to 48 hours of this data onboard the device. A picture of the device is shown in Figure 20. The whole system is no larger than the length of pencil and only weighs 2 pounds. A present MEG system would weight in the range of 250 to 400 pounds and be 3 or 4 feet in length.

Another strength of the EEG is the excellent temporal resolution. EEG like MEG has basically no time delay in recordings. The time from the production of an electric current and the recording of this activity is within milliseconds. The spatial resolution is also comparable to the other modalities at 7-11 mm. This is perhaps an order of magnitude worse then PET and MRI, but this range is very similar to MEG. When the issue of cost is considered, the drastic difference does not
justify the small gain in spatial resolution.

EEG has another advantage over MEG. MEG detectors are only able to detect sources that are located tangential to the head. This limitation is caused by the orientation of the detectors used in MEG systems. The EEG on the other hand is able to detect all three vector components of a source with the brain. Figure 21 shows a diagram with an example of a source and the components that are detectable. An EEG contains more information than the MEG signal because of this limitation.

Analysis Techniques of EEG: Frequency Analysis of EEG

EEG signals have traditionally been divided into four main frequency components. The lowest frequency component is known as the Delta rhythm and is only present in deep sleep and ranges from 1-4 Hz. The theta band is generally classified between 4 and 7 Hz. This rhythm is very difficult to detect in humans but is thought to be associated with lighter sleep and perhaps meditative state. The alpha band has been studied extensively over the history of the EEG and is in the frequency range of 8-13 Hz. This wave pattern can be seen in patients while their eyes are closed and during relaxed states. Beta waves have generally been considered any frequency above 13 Hz. However as more research is being done into these waves higher levels of classification are starting to be developed.

Frequency analysis began with Hans Berger in the early stages of EEG research when he found that alpha waves were blocked or decreased during mathematics exercises. The idea of alpha blocking first demonstrated by Berger is now known as event-related desynchronization (ERD) and it’s activation is now associated with not just mathematical tasks but rather general arousal of the individual. Studies were undertaken in the 60’s and 70’s involving ERD and it’s association across the two hemispheres of the brain [42]. The results from these studies however were very few and often were not repeatable by other researchers. For these reason’s frequency analysis fell out of favor with many researchers until the 90’s when interest was again sparked as advances in technology made analysis of large amounts of data possible and much easier.

The interesting and exciting research involving cognitive task and EEG presently is centering on the higher frequency band, or the so called gamma band, anything above 20 Hz. Local gamma band activity has been found to increase while preparing for and during motor movement, during language processing, and during visual tasks [42]. Activity in the 40 Hz range has been shown to increase during simple and repetitive motor movements [42]. Hypothesizes are now being put forward to divide the gamma band into a lower (20-40 Hz) and upper (40-60 Hz) gamma band each with it’s own properties. The lower band seems to show activity during complex processing of tasks such as language processing and the higher band provides activity during sensory processing tasks. It is believed that these different frequencies are created by different groups of cells needing to be activated for these different tasks to be performed. In the lower frequency bands the activated cells are further apart from one another so the traveling loop of activity is longer and only repeats every 20-40 ms. Higher frequencies would be caused by cells in closer proximity to each other with shorter loops of travel and a faster rate of repetition.

Analysis involving coherence between electrode sites also shows promise. Coherence is the correlation between two signals with as a function of frequency. Patterns of change can be studied in the coherence of differing electrode sites during mental tasks and complex patterns have been found to exist in inter and intra hemispherical coherence.
Event-Related Potentials (ERP)

Event-related potentials are another analysis technique used for detection of an event or activity. ERP’s occur after an event or stimuli with a given time delay and polarity. ERP’s are difficult to detect with the EEG because of the small amplitude and are typically found by increasing the SNR of the signal through averaging of the signal to draw out the ERP. The most well known ERP is known as the P-300, the name means that the shift in voltage is positive and it occurs approximately 300 ms after the stimulus.

DC potentials

DC potential shifts occur during long periods of activation in a specific region of the cortex. This occurs when a subject has to think for an extended period of time about a certain task.

It was found in the 1960’s that a negative DC shift could be recorded in anticipation of movements or in the expectance of stimuli, this shift started to develop 800-1000 ms before movement. This negative DC shift in anticipation of voluntary motor movement was given the name “Bereitschaftspotential” or readiness-potential.

EEG in Space

EEG has a limited and fairly uninteresting history in space. However, some important information did come from these few experiments performed on astronauts in space environments. The most important observation that has been made is that there appears to be no significant difference in the EEG of humans on Earth and humans in space. The only documented difference was a slight increase in theta activity. Therefore the conclusion and argument can be made that any EEG paradigms on Earth will hold true in space. EEG has been used previously as a tool to try to diagnosis physical problems expressed by astronauts while in space.

The most extensive studies were conducted during the Skylab missions and involved sleep studies. These studies were motivated by complaints of fatigue by many astronauts while on missions. These studies had the purpose of determining if sleep cycles were being compromised during space travel leading to poor sleep and tired astronauts. It was found that the astronauts were in fact receiving adequate amounts of sleep and only minor variations were seen in the astronauts sleep cycle from preflight sleep patterns. Some other minor variations were also observed from preflight data during post flight analysis of the data. Two of the subjects showed an increase in alpha-rhythm frequency and beta frequencies and all three subjects showed an increase in delta frequency. The cause of these variations has not been identified because too many potential factors could be labeled as a cause.

In a more recent study (1993) sleep was once again examined using EEG. Here sleep patterns were shown to be shifted from pre-flight sleep patterns with shorter periods of sleep and more disruption of sleep in-flight.

EEG was also used to investigate space motion sickness syndrome in 1985 using auditory and visual evoked potentials. These tests provided normal results as would be expected if they had be carried out on earth rather than in space.

EEG and Prediction

Some promising research has been done using EEG as a prediction tool. Pfurtscheller et. al used EEG to classify hand movements of human subjects in an online environment. One of the very impressive qualities of this work is that it was done using only three electrodes, two signal
channels and a reference channel. Six subjects were studied at three different times, in each of
these sessions subjects were instructed by a video monitor to either move their right or left arm. The first two sessions were used as training data for a neural network classifier and the third session was online classified by the algorithm as either right or left arm movement. Four of the subjects showed very high results of 89-100% correct classification, the other two subjects had classification percentages near 50%, which could be attributed to chance [43].

This particular study is very encouraging for the possibility of being able to online classify a person’s needs, movements and emotions. There is room for vast improvement in this study, which could lead to higher levels of classification. Increasing the amount of training data would undoubtedly improve accuracy. The use of only two monitoring sites could be increased, while the use of two sites keeps the data analysis fairly simple, more information would improve the power of this classifier.

Conclusions of the review of Brainwave Monitoring Technologies

Table 5 provides a visual comparison of the four modalities and several important parameters. Functional MRI and PET each offer very high levels of spatial resolution, much better than both MEG and EEG. However this is their only strong attribute for an application such as this. They provide poor levels of temporal resolution, which in the best case could be no better than 1 seconds but is closer to several seconds in practice. These systems are expensive and bulky, the idea of a wearable or even slightly portable version of these systems is very hard to imagine anytime in the future.

<table>
<thead>
<tr>
<th>Modality Comparison</th>
<th>EEG</th>
<th>MEG</th>
<th>fMRI</th>
<th>PET</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial Resolution</strong></td>
<td>7-11 mm (6)</td>
<td>3-11 mm (7)</td>
<td>1 mm(2)</td>
<td>5 mm(8)</td>
</tr>
<tr>
<td><strong>Temporal Resolution</strong></td>
<td>Milliseconds(1)</td>
<td>Milliseconds(1)</td>
<td>1 second to minutes(limit due to metabolic processes)(2,5)</td>
<td>45 seconds to minutes (2,3,5)</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>$15-200k</td>
<td>$500k (37 channel) $300-400k Shielding</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Wearability</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

MEG and EEG are very similar systems in what they are capable of. However EEG has the advantage of being about 50 years the senior of MEG. This has resulted in cheaper instrumentation and more compact size. At this point the cost of MEG systems does not seem to justify the advantages gained. Furthermore MEG is bulky, and while there probably will be a day when there are smaller, possibly wearable MEG systems, that day is most likely very far in the future. For these reasons EEG would be the better system for this application. The spatial...
6.2 EEG Pilot Test

Given the results of our analysis of brainwave monitoring technologies clearly indicated that EEG was the best potential measure for our affect detection system, we conducted a very preliminary pilot test to get a sense of the feasibility of using EEG to assess affect-related information in the problem-solving context with which we are working. In this pilot test, EEG from the standard central parietal site (Pz) was monitored while a 44-year old male participant engaged in eyes open and eyes closed baseline periods, and then worked through the difficult version of the math task we used in the human experiments described above. The signal was collected using a Beta version of a portable Flexcomp Infinity model from Thought Technology Inc. that we had been loaned to test. The EEG signal was sampled at 2000 Hz. The Pz signal was referenced to a neutral site on the participant’s ear lobe.

![Figure 22. EEG Pilot Test – Comparison of Eyes Closed and Eyes Open Baselines](image)

Thirty seconds of EEG data were extracted from the eyes open and eyes closed baseline periods, as well as from 30 sec each from three of the math problems of varying difficulty. For each 30 sec epoch, spectral power in the alpha band (8-13 Hz, [44]), was estimated, and compared across the various epochs. In general, alpha band EEG activity has been associated with mental alertness versus fatigue, with higher levels of alertness being associated with lower levels of alpha power [44]. In addition, in the parietal region, alpha power is known to markedly increase when the eyes are closed. Figure 22 depicts the comparison of the eyes open and eyes closed. As can be seen the alpha power was much higher during the eyes closed baseline than in
the eyes open one, which gives us confidence that we were collecting and processing the EEG data correctly.

![Figure 23. EEG Pilot Test – Comparison of Eyes Open Baseline with Solving Math Problems of Varying Difficulty](image)

Figure 23 depicts a comparison of the eyes open baseline with the three epochs from the math task, with problem difficulty increasing from left to right. As can be seen, there is a steady decline in alpha power as the problems become more difficult, and presumably demanded more of the participants attentional resources. Although these results must be viewed as extremely preliminary, they suggest that EEG data, may indeed, be able to be used to supplement the physiological data we are already obtaining to assist in the task of affect detection and recognition.

7 Conclusions and Future Directions

7.1 Summary of the contributions
For both of the major aims we initially set out to address, we have accomplished what we have proposed to do, and have gone a step beyond. Thus, in working toward the development of an affect recognizer to be used in human-robot communication, we not only demonstrated the feasibility of using affect related physiological information to identify a person’s emotional state, we also were able to use this information in a working demonstration of how a functioning human-robot communicative system might actually work.

Similarly, in surveying existing brainwave monitoring technologies for potential use in such a communicative system, we were able to conclude that EEG was the most suitable technology for our purposes. In addition, after making this determination, we were able to collect some preliminary data that indicated that, not only is it feasible to integrate EEG technology into our
physiological affect assessment battery, but also that doing so is likely to have considerable payoff.

We believe that the progress we have achieved on both fronts not only demonstrates the feasibility and utility of the implicit communication system concept, but it also puts us on a strong footing to make substantial progress toward the development of an actual functioning system with our Phase II activities.

7.2 Overview of Phase-II

The research to be pursued in Phase II is designed both to consolidate the progress that we achieved in Phase I and to significantly advance the development of a human-robot system capable of implicit communication between human and robot. The advances we will pursue in Phase II will be developed along two distinct lines. First we will continue and enrich our Phase I efforts to develop a robust affect detector that can be used by the robot to modify its behavior in response to information about the human operators’ spontaneous affective state, without any communicative intent being necessary on the part of the human operator. Second, we will expand our examination of implicit communication to include different paradigms in which the communication on the part of the human is more intentional, albeit not involving explicit spoken or written commands. Below we briefly outline our plans for both of these lines of development.

7.2.1 Continued development of a robust affect detector

In the work we have described above, we believe that we have clearly demonstrated the feasibility of such a system. Through human psychophysiological experimentation we have verified the existence of physiological markers, indicative of both autonomic and facial activity that have the potential to be used in real-time to diagnose the person’s emotional state. In addition, we have demonstrated the importance of adopting a person-specific approach in developing an affect recognizer. Within a particular context, such as the cognitive problem-solving domain we have examined, physiological activities are associated with the person’s affective state in systematic ways. However, the specific associations vary considerably from person to person. Thus, much as is the case with voice-recognition technology, it appears that an affect detector designed for a particular individual will need to be tuned to that person’s specific patterns of physiological responding. Therefore, in pursuing our Phase II research we will continue, and, in fact, intensify, our person-centered research approach. Accordingly, in Phase II we will focus on a relatively small number of individuals who will be brought into the laboratory repeatedly over the course of several weeks to engage in an extended series of tasks that will enable us to develop a robust affect detector for each individual. These research activities will include the following enhancements over what was done in Phase I.

7.2.1a Improved assessment of subjective affective states. One limitation of the Phase I research was that in assessing the research participants’ affective states we essentially asked them to summarize what they had experienced over intervals as long as 7 minutes, and we summarized the physiological data over these same extended intervals. Of course, one’s affective state can change considerably over the course of 7 minutes, and to the extent to which it did in these tasks, the “snapshots” we obtained of the person’s physiological activities and affective states were blurred due to poor temporal resolution. This blurriness likely attenuated the observed correlations between physiological activity and emotion that we obtained. Thus the results we reported are likely underestimates of the degree to which physiological activities are
reliably linked to the individual’s affective state. In the next phase of this research we will address this problem by improving the temporal resolution of our assessments of both subjective emotion and physiological activity. Rather than stopping participants at natural break-points in the task (e.g., at the end of a math problem in the math task), participants will be probed at predetermined intervals while the task is ongoing (e.g., while still working on a math problem, or while attempting to solve a particular anagram, etc.), and they will be asked to indicate their affective state at that particular moment. These reports will then be referenced to the physiological activity during the interval immediately preceding the affect assessment. As described in the section on “multi-rate signal processing” below, the exact time interval preceding the affect assessment to be referenced to the affective reports may depend on the measure being considered. Thus, relatively “fast” measures, such EMG and SC activity may be summarized over a very brief interval of 30 seconds to a minute, whereas spectrally derived parameters of interbeat interval may be summarized over a 3-minute period. In any event, the temporal resolution of each physiological parameter will be as fine-grained as the parameter itself will allow.

7.2.1b Examination of an increased range of experimental tasks. Another enhancement is that with the increased levels of participation for each individual, we will increase the range of problem-solving tasks to be examined. For instance, we are already working on adding a simple, “pong”-type computer/video game to the battery. Notably, each of the tasks we have developed so far, and are contemplating developing for Phase II are such that they can be extended indefinitely (e.g., through the addition of new math problems or anagrams), such that they can be meaningfully engaged in repeatedly by the same individual, as will be required by our person-specific approach.

7.2.1c Identification of an increased range of affective states. The increased range, and extended, repeated use of the experimental tasks will allow us to expand the range of affective states we can meaningfully elicit and attempt to characterize physiologically beyond task engagement and anxiety. For example, through extended sessions of trivially easy trials of the various tasks we can elicit boredom and fatigue; by combining one of the other tasks with the sound discrimination task (which was specifically designed with this purpose in mind), we can elicit high levels of confusion and cognitive overload; by suddenly changing the response properties of the mouse or joystick in a video-game task, we can elicit anger and frustration, and so on. As we become confident of our ability to reliably differentiate task engagement and anxiety, we will systematically set out to elicit a broader range of affective states through our tasks, such that the identification of these states can be incorporated into the development of the affect recognizer.

7.2.1d Utilization of an increased range of physiological measures. Our efforts to reliably discriminate between a broader range of affective states will be facilitated through the use of an increased range of physiological measures. We are planning of adding two major types of additional measures to our current battery. First, given the outcome of our review of brainwave monitoring technologies, we intend to incorporate the assessment of EEG into our battery. At the very least, use of EEG should provide us with an index of mental alertness [44], which is an important component of task-engagement, and the use of EEG should be especially useful in identifying when the human operator has become dangerously fatigued [45]. Second,
we will enhance our assessment of cardiovascular functioning through the addition of impedance cardiography, which, based on an analysis of blood flow through the chest, provides a number of useful indicators of cardiovascular functioning. For instance, the impedance-derived measure of pre-ejection period (PEP), which corresponds to the time difference between the onset of the electrical signal to contract the heart (derived from the ECG) and the onset of the physical ejection of blood from the left ventricle, provides a much purer estimate of sympathetic nervous system influences on the heart than can be derived from the analysis of IBI [46]. In addition, the index of total peripheral resistance (TPR, an index of the degree to which the peripheral vasculature is constricted or dilated) also yielded by impedance cardiography has been shown to be useful in differentiating between states of threat, associated with anxiety, and of challenge, associated with task engagement [47].

7.2.1e Use of a 4-stage validation procedure. The Phase I human experiment work we have reported represents only the first step of a multistep development and validation procedure that is necessary to produce a truly functional affect detector. In this step we have documented the correlates between affect and physiological activity in a given person. An important second step, that we will pursue in Phase II, is to cross-validate these relations by demonstrating through a second assessment that the observed relations are stable and can be replicated. In this step the participant will be put through a series of tasks similar to the first series, and the degree to which the same relations can be observed across both series will be assessed. As the third step, we will incorporate our improved signal processing and analysis techniques (see next section, below) into our assessment, as the participant is put through a series of tasks a third time. This time, instead of assessing subjective affect at experimenter-determined points in the task, the participants’ physiological signals will be analyzed in real time, and the computer administering the task will interrupt the participant for affect assessments when it detects that the person either is, or clearly is not, in a given target state. In this way we can verify that the analyzer is accurately detecting the person’s affective states. At this point, we will have effectively developed a functioning affect detector for the particular individual. In the final step, rather than being used to determine when to assess the participants’ affective state, the participant will perform a series of problem-solving tasks while interacting with a robot whose behaviors are being modulated based on information derived from the affect detector, utilizing the formalized affect-sensitive control architecture discussed below. At this point we will have a functioning prototype human-robot communication system, albeit still a primitive one.

7.2.1f Use of advanced signal processing and analysis techniques to robustify affect detection and recognition. Phase I work involved basic off-line signal processing to identify important discriminating features of each physiological signal (e.g., ECG, SC, EMG etc.). These features were used in a fuzzy logic affect analyzer to infer the affective state of the human. These were the necessary steps to first understand the feasibility of our approach. Now that we have been able to demonstrate that several physiological signals can be analyzed to infer affective state, our next, Phase II objective is to further investigate several advanced signal modeling and recognition techniques to robustify the affect detection and recognition process. Space limitation does not permit us to describe each technique in any reasonable detail. As a result, we will describe only one technique for affect detection and one for affect recognition that will be the most logical extension of Phase I research. Other techniques of importance will be briefly mentioned.
Signal Analysis. The spectral analysis in Phase I was performed by Fourier transform (FT). Although the resulting frequency spectrum was useful in identifying the sympathetic and parasympathetic activities, and consequently affect detection, the temporal information was lost. Since all physiological signals of our interests are nonstationary, the natural extension in signal processing technique will be to analyze the signal using Wavelet transform (WT), which will provide both frequency and time localization of the signal. In other words, WT will allow us both to identify the frequency band of interest and the signal variation over time in that band. This will be very useful in real-time affect detection since we are interested in knowing exactly when in time a signal of particular frequency changes its nature. For example, when we monitor the EEG signal, we are interested in knowing when there is a sudden burst of alpha waves (i.e., a temporal variation in the 8-13 Hz band). The mathematical details of WT can be found in [48]. The basic idea is to represent a given function as a sum of time shifted (translated) and scaled (dilated) representations of some functions called mother wavelets. We have performed basic wavelet analysis of ECG signals in pilot studies [49], and will use it extensively in Phase II work.

Besides WT, we will also explore both parametric signal modeling and adaptive signal modeling techniques. We will develop person-specific parametric signal modeling techniques so that model-based matching and affect detection can be performed. The basic idea is to reduce a complicated process with many variables to a simpler one involving a small number of parameters. Adaptive signal modeling techniques, on the other hand, will allow us to adjust the weights in an adaptive algorithm to reduce the mean-square error towards its minimum value so that a good model can be achieved for the unknown signal. We will also incorporate multi-rate processing of the physiological signals. It is clear that not all physiological processes have the same dynamics. Thus it is computationally more efficient to sample a slow signal at a lower sampling rate while sampling a fast signal at a significantly higher rate.

Affect Recognition. There is very little systematic study to date on affect recognition techniques in the literature. As a result, we propose to investigate three competing techniques to assess their relative merits: learning theory based affect inference (e.g., neuro-fuzzy technique), purely statistical affect inference (e.g., regression analysis), and probability based affect inference (e.g., belief network). The natural extension of our Phase I affect recognition work is to investigate neuro-fuzzy technique, which we describe first. The fuzzy inference system captures input characteristics through the input membership functions and output characteristics through the output membership functions. In Phase I work, as a first step, we had only considered membership functions that were fixed, and were somewhat intuitively chosen according to the nature of the inputs and outputs. Neuro-adaptive learning, on the other hand, provides a method for the fuzzy modeling procedure to train itself from a given data set to compute the membership function parameters that best allow the associated fuzzy inference system to track the given input/output data. This learning method works similarly to that of artificial neural networks. The membership function parameters are tuned (adjusted) using either a back propagation algorithm alone, or in combination with a least squares method. The computation of these parameters (or their adjustment) is facilitated by a gradient vector, which provides a measure of how well the fuzzy inference system is modeling the input/output data for a given set of parameters. In Phase II, we will augment our fuzzy inference system by combining it with neuro-adaptive learning algorithms so that the membership function parameters can be automatically determined from a physiological signal.

Additionally, we will use regression analysis where we need to obtain the relationship
between a set of several physiological responses and the corresponding affective state that can be mapped to that set. Multiple regression can prove to be a useful tool to learn more about the relationship between several independent predictor variables and a dependent or criterion variable. Belief Net [50] will be another affect recognition tool for our Phase II work. The design of a reliable decision making system that takes as input various physiological parameters and determines the affective state of a human as an output must take into account the uncertainty and the noise associated with the input data. Probability theory can be an effective tool for dealing with this kind of uncertainty. A belief network, which is a computationally efficient probabilistic technique for determining a state based on conditions, will be explored in Phase II.

7.2.1g Formalization of the affect-sensitive control architecture. In Phase II work, we will formalize the control architecture developed in Phase I. The role of affective layer will be expanded. A multiple control action capability will be built into the layer and a mechanism to choose one depending on the affective feedback. Consequently, traditional control theory will not be of much help. A hybrid control architecture will be most appropriate because it will allow the transit from one controller to another based on certain events. A hybrid system involves both continuous and discrete event dynamics. Hybrid models operate in continuous modes but at points in time when signal values cross pre-defined thresholds or when explicit external events are imposed on the system, changes in model configurations cause discrete changes in system behavior. In a hybrid system, the temporal trajectory of system behavior becomes piecewise continuous, where simple discontinuities can occur at well-defined points in time. The discontinuities of these systems can occur from four main events: controlled jumps, autonomous jumps, controlled switching and autonomous switching. Controlled jumps occur when the state changes in response to a controlled command. When the state changes upon reaching a prescribed boundary, it is an autonomous jump. Controlled switching changes the continuous dynamics vector field in response to a control command and autonomous switching changes the vector field upon reaching a prescribed boundary. A detailed discussion on hybrid systems can be found in [51].

Next we will decide how to design the continuous controllers within a hybrid system. There could be both structured and unstructured scenarios in human-robot cooperative tasks that we are concerned with. For example, in an assembly line, the environment is very structured and the task of the robot is generally defined. On the other hand, during an exploration the environment is uncertain and the robot needs to work robustly. A well-designed traditional controller may be an optimal solution for structured, well-defined environment. But where the task requires negotiation with uncertainty and a model is hard to come by, it is prudent to design robot behaviors following the methods as outlined before.

Once we design each continuous controller, which could be a traditional controller or a behavior depending on the task at hand, we cast the overall control problem in terms of hybrid automata. We will extensively use and modify the theoretical framework developed in [52] for human-robot cooperative tasks. In this framework, each node will represent a behavior (or a traditional controller). The switching from one node to another will be dictated by both task related events and affective feedback. Since our main interest is to exploit the affective event we will not discuss the task-related events in the following discussion, although they will be considered during implementation. The affective event detection will be carried out by fuzzy analysis (e.g., is it a high level of engagement?) since crisp event detection will not be appropriate for affect dynamics. Switching from one node to another will be in response to an
affective event. However, switching may lead to chattering, which in some cases may exhibit Zeno properties (i.e., infinite number of discrete transitions in finite time). The reason for such behaviors is that the underlying system that the automaton seeks to model is essentially a switched system that has sliding properties in the sense of Filippov [53]. Fortunately, as shown in [52], the chattering effects can be avoided by regularizing the automata by introducing additional nodes to the automaton. These nodes will capture the sliding dynamics that are defined on the boundary between two behaviors. It is possible to identify the sliding solution without much difficulty in many cases (e.g., when the geometric description of the switching surface is available). The hybrid automata approach is also useful for ensuring safety and optimality [54]. It provides a theoretically elegant way to define safety functions in an optimal control framework. Although such optimal solutions are difficult to obtain and are generally not suitable for online tasks, sub-optimal solutions can be derived using heuristic methods.

In summary, the robot control architecture will be based on hybrid automata where each node will represent one continuous action. That continuous action could be a behavior or a predefined control action. The affect recognition algorithm will generate a fuzzy inference on the affective state of the person. An event will be defined for node transition based on the affective state. In order to avoid the potential chattering problem due to switching, a regularization scheme for the hybrid automata will be developed which will allow us to introduce additional nodes to capture the sliding solutions. The behaviors and predefined control actions could be combined for one robot based on task requirement. For example, in an exploration scenario, the robot may be ideally on so-called roam-and-explore mode. However, when it detects that the human is in danger, it may suspend its “roam-and-explore” behavior and switch to “reach the human fast” mode, which could be a PID control with a step input (step being the distance from the human). While running towards the human, the robot should incorporate “avoid obstacle” behavior so that it can safely reach the human. We will spend considerable effort to design the behaviors and predefined control actions for our experimental tasks. Understanding and analyzing the resulting hybrid automata will provide a basis for the proposed control architecture.

7.2.2 Examination of more intentional implicit communication paradigms

Beyond the enhancements designed to lead to the development of a robust affect recognizer, during Phase II we will also expand the range of paradigms examined to include more intentional implicit communication on the part of the human. We will develop this research line in two distinct ways, first, we will draw upon the human developmental psychology literature on social referencing [55] to develop a paradigm in which, under conditions of uncertainty, the robot looks to the human operator for guidance, and then modifies its behavior on the basis of a facial signal the human produces. Second, using a somewhat similar paradigm, we will explore the feasibility of using EEG-based evoked potentials to send similar signals to the robot through more purely cognitive means.

7.2.2a Intentional communication through facial expressive behavior.

In the typical social referencing paradigm, older infants and young toddlers are presented with ambiguous stimuli. Under such conditions, the infant or toddler will look to their parent (the act of social referencing) as if seeking advice. The information parents provide to the child in response to being referenced is often experimentally manipulated, and the impact this information has on the child’s behavior is assessed. In the paradigm we will develop, a human operator will be monitoring a robot’s performance of a given task. At major decision points in
the task, the robot will initiate one of two or more potential courses of action (e.g., to select from an array of several tools the one most appropriate for the task at hand), but will also monitor the physiological information coming in from the operator for feedback. The operator in turn will be trained to facially signal approval (e.g., by smiling and/or raising his/her eyebrows) if the robot selects the correct/desired course of action, and to facially signal disapproval (e.g., through an eyebrow frown and/or pursed lips) if the robot selects the incorrect/undesired course of action. The operator’s facial signals will be assessed using EMG, and these signals will be transmitted to the robot. The robot’s control architecture will be programmed such that upon receiving the signal from the operator, the robot will go back to the decision point and select a different course of action if the disapproving signal is received, but will continue on in with the selected course of action if the approving signal is received.

The technical challenges of developing such an intentional communication paradigm are considerably less than those associated with the detection and identification of spontaneously produced affective states that we have been pursuing from the outset of Phase I. Thus, we do not anticipate that it will be especially difficult to develop a functioning prototype of this paradigm. We anticipate that one of the main issues to be confronted in the development of this prototype will be the ease with which the human operator can be trained to produce the appropriate facial signals to guide the robot’s behavior. However, by modeling the target signals for “yes” and “no” after naturally occurring expressions for approval and disapproval, we anticipate that most human operators will be able to learn how to deliberately produce the appropriate facial signals on demand, and that with a reasonable amount of practice, they will find the operation of the communicative channel to be easy and to feel natural.

7.2.2b Exploring the signal value of evoked potentials.

As noted above, in Phase II we will be incorporating measures of EEG into our assessment battery to be used to identify a person’s spontaneously produced affective state. In that usage, we will be using spectral analytic techniques to estimate the power of EEG activity in specific frequency bands (e.g., estimates of alpha power) that should yield useful information regarding the person’s affective state. In addition, in the research line being proposed here, we will explore the feasibility of using evoked potentials (EPs), a second major way of analyzing the EEG signal, in an implicit communication paradigm much like that being developed for facial communication as described above. The evoked potential is measured in a tightly time-linked fashion in response to a discrete stimulus. Although the EP evoked in any given single trial is quite noisy, when the EPs from multiple trials of a particular type are averaged, the signal in the EP is amplified, and the noise is canceled out across trials to yield a waveform with characteristic peaks and valleys. One of the best known characteristics of the EP is the “P3” or “P300” which is a positive wave that is initiated approximately 300 ms following stimulus onset. Other heavily studied landmarks are the “N4” or “N400,” which is a negative wave that follows the P3, and the “N2” or “N200” that precedes it. The characteristics (e.g., onset latency, amplitude) of these and other characteristics of the EP have been found to have considerable psychological significance. Notably all three of the components mentioned above are often evoked by some sort of mismatch between a stimulus event and some standard or expected event. For instance, the N2 often corresponds to some sort of perceptual mismatch, the P3 tends to be elicited by a stimulus event that in some way deviates from expectations, and the N4 is often elicited by a semantic or linguistic incongruity [56].
Our proposed investigation of the EP will be modeled loosely after the social-referencing paradigm described above. In this investigation the human operator will monitor a robot as it repeatedly attempts to perform a discrete task (e.g., to choose between two tools). On some trials it will make the choice that has been predefined as “correct”, and on other trials it will make the choice that has been predefined as “incorrect,” while the human operator’s EPs are monitored.

Initially, and in line with traditional EP research, we will aggregate the EPs within correct and incorrect trials to derive the characteristic EP associated with each trial type. These aggregated EPs from the two conditions will then be analyzed to test for systematic differences between them. To the extent to which systematic differences between the two types of aggregated EPs can be reliably characterized, we will then seek to apply signal processing techniques, such as those described above, to the EPs associated with single trials, in an attempt to reliably differentiate between the two types of EP at the single trial level. To the extent to which these efforts meet with success, we will then seek to develop an implicit communication prototype, analogous to the one to be developed for facial activity as described above, but where the communicative signal is based on purely cognitive activity, as reflected in the EP, rather than on facial expression.

7.3 Summary of Phase II Plans

In Phase I of this project we demonstrated the feasibility of developing an implicit communication system between human and robot, such that the robot, by detecting and analyzing the human’s emotional state, could be responsive to the human’s needs with receiving explicit commands. In the Phase II activities we have outlined above we will be taking several concrete steps to make such a communication system a reality. Moreover, by also investigating ways in which human operators can more intentionally communicate with the robot through nonverbal modalities (e.g., facial activity), we are further advancing toward a point where human robot communications can be truly seamless and natural, even under conditions in which spoken or typed verbal commands are impractical or impossible.
References

1. ICASE/NASA/LaRC Workshop on Revolutionary Aerospace Systems Concepts for Human Robotic Exploration of the Solar System, Nov. 6-8, 2001 at NASA LaRC.
32. www.fil.ion.ucl.ac.uk/spm/course/notes97/Ch8slides.pdf
34. 4-D Neuroimaging Inc. http://www.4dneuroimaging.com
35. Tristan Technologies http://www.tristantech.com/
37. CTF Systems Inc. www.ctf.com