# The Utility of Affect Expression in Natural Language Interactions in Joint Human-Robot Tasks

Matthias Scheutz
A.I. & Robotics Lab
University of Notre Dame
Notre Dame, IN 46556, USA
mscheutz@cse.nd.edu

Paul Schermerhorn
A.I. & Robotics Lab
University of Notre Dame
Notre Dame, IN 46556, USA
pscherm1@cse.nd.edu

James Kramer
A.I. & Robotics Lab
University of Notre Dame
Notre Dame, IN 46556, USA
jkramer3@cse.nd.edu

## ABSTRACT

Recognizing and responding to human affect is important in collaborative tasks in joint human-robot teams. In this paper we present an integrated architecture for HRI and report results from an experiment with this architecture that shows that expressing affect and responding to human affect with affect expressions improves performance in a joint human-robot task.

## 1. INTRODUCTION

Social robots that interact with humans have become an important focus of research in robotics and human-computer interaction (e.g., see [13] for a comprehensive overview). As "human-robot interaction" (HRI) is being recognized as an independent, interdisciplinary field of its own, a variety of technological challenges need to be addressed by the HRI community, from general communication issues (including direct or mediated human-robot communication or HRI interface), to modeling (e.g., cognitive modeling of human reasoning), to teamwork (e.g., architectures for joint human-robot teams), and more (see the final report of the DARPA/NSF Interdisciplinary Study on Human–Robot Interaction [7]). Questions such as how to interpret commands given by humans, how to derive human intentions, how to recognize non-verbal cues including affect expressions or gestures, and others will be critical to joint human-robot teams that have to achieve a task together (e.g., for human-robot teams as envisioned by NASA for future space missions [14], but also elsewhere). We believe that understanding *human affect* and reacting to it appropriately might not only be essential for robots in some situations (e.g., in order to avoid misunderstandings or to allow for more natural interactions between robots and humans), but could potentially also improve the task performance of a joint human-robot team.

In this paper we report results from a study that is intended to measure and quantify the role of affect in human-robot interactions and its impact on the task performance in joint human-robot teams. Specifically, we investigate the question whether expressing affect and responding to human affect with affect expressions in natural language can facilitate task performance in mixed human-robot teams.

*Human Robot Interaction* '06 Salt Lake City, Utah USA

## 2. BACKGROUND

Affect is deeply intertwined with cognitive processing in humans and is, consequently, an integral part of human communicative situations. *Negative affect*, for example, can bias problem solving strategies in humans towards local, bottom-up processing, whereas *positive affect* leads in many cases to global, top-down approaches [2]. Affect is also crucially involved in *social control* ranging from signaling emotional states (e.g., pain) through facial expressions and gestures [12] to perceptions of affective states that cause approval or disapproval of one's own or another agents' actions (relative to given norms). Many aspects of natural language communication cannot properly be understood without taking the accompanying affect expressions into account.

While affect has been investigated to varying degrees since the beginning of AI [25], *affective computing* has become more prominent only since the publication of Picard's seminal work on the topic [26]. Since then various architectures for affective robots have been proposed [37, 24, 20, 6, 22, 31, 29, 21]. These architectures differ in several respects and can be categorized along several dimensions, for example, in terms of the architecture schema within which they are defined (e.g., a behavior-based approach like subsumption or motor schemas vs. other approaches), the employed deliberative components (if present), or whether natural language processing is integrated.

More importantly in the present context, they also differ with respect to the notion of affect and how affect is used in the architecture: (1) how affect is (functionally) defined and implemented, (2) how it can influence the robot's behavior, (3) where and how affect mechanisms are integrated into the architecture, (4) whether affect in others (e.g., in humans) can be perceived, (5) whether affect can be expressed (e.g., in the voice of the robot), and (6) whether affect can be internally generated without perceptions.[1]

[1] In some cases, for example, affective states like emotions are taken to be discrete and are architecturally represented by a corresponding number of components (e.g., neural network-like units with activations as in [37, 31, 21]), whereas others construe them as continuous subspaces of an $n$-dimensional space determined by some basic variables

# 3. THE DIARC ARCHITECTURE

We believe that affect can play an important role in HRI on both the interaction side (i.e., via affect recognition and expression) as well as the architecture-internal side (e.g., see [32] for different roles of emotions in agent architectures). Hence, we have developed a robotic architecture called "DIARC" (for "distributed integrated affect, reflection, and cognition") for HRI over the last several years that integrates cognitive and affective mechanisms.[2] Figure 1 depicts a partial view of the functional organization of the architecture, showing only the components relevant to the experiment described (see [35] for a more detailed overview). [3]

For space reasons, we will only describe the three components of the architecture that are relevant to our experiment, because they are involved in affect processing: the *affective action interpreter*, *affect recognition in spoken language*, and *affective speech production*.

## 3.1 The Affective Action Interpreter

The "affective action interpreter" is a novel interpreter for scripts that is used for natural language understanding as well as action selection, action sequencing and action execution. For this purpose, scripts can be augmented by action primitives that are grounded in basic *skills* of the robot (the bottom layer control structures are implemented as motor schemas as in [1]). Scripts can be combined in hierarchical and recursive ways, yielding complex behaviors from basic behavioral primitives.

Action selection is accomplished via a prioritized goal stack. The robot has high-level *permanent goals* that are always present (e.g., "be polite"). In addition, *transient goals* can be put on the goal stack as they are generated by pre- and post-conditions in scripts. Each transient goal has an expected *time-to-completion* and a *utility* associated with it, which reflects the benefit of completing the goal in time and the cost of performing the required actions.

Each script goal can consist of multiple subgoals. A subgoal may be another script goal or an atomic action. In general, a script's subgoals are pushed onto the stack in order; when one subgoal is accomplished, it is popped and the next is pushed. Subgoals can also be *conditional* (e.g., the outcome of an action can lead to one sequence of subgoals on success and to another on failure). Unlike a normal stack, the top of the prioritized goal stack is not always the most recently pushed goal. Rather, the order of the goal stack depends on the *priority* of each goal. A goal's priority ($P$) is essentially a measure (or function) of the *importance* ($I$) of the goal to the robot and of the goal's *urgency* ($U$).

*Urgency* is related to time (similar to [22]). Each goal is allotted a fixed amount of time ($T_A$) when it is pushed onto the stack, within which it has to complete.[4] The closer a goal is to timing out (i.e., the smaller its remaining time $T_R > 0$), the greater its urgency. Specifically, $U = \frac{T_A - T_R}{T_A}$. If reliable estimates of remaining time to completion can be made for subgoals, $T_R$ can be computed as the difference between the time remaining to complete the task and the time remaining before it times out. Otherwise, $T_R$ is just the time remaining before timeout.

The *importance* of the goal is based on the benefit of achieving the goal ($B$) and the cost of performing the actions required ($C$), along with the current *positive* and *negative affective* mood states of the robot ($A_P$ and $A_N$, respectively).[5] Specifically, $I = (B \cdot A_P) - (C \cdot A_N)$, i.e., the importance reflects some measure of expected utility if the intensity of the positive and negative affective mood states are taken to be self-generated "estimators" of future outlooks (e.g., positive moods in humans can lead to positive outlooks, top-down problem solving, welcoming of changes, etc., whereas negative mood leads to negative outlook, problem-focused search, avoidance of changes, etc.).[6] The *mood states* $A_P$ and $A_N$ themselves are computed based on the failure or success of computations in various submodules. $A_N$ is increased based on failures to recognize words, interruptions in motor actions, failure to complete goals, etc., while $A_P$ is increased based on successful completion of some computations (such as successful parses of sentences), completion of entire action sequences, or achievement of complex goals. In addition, $A_N$ can be increased by successful detection of certain negative properties (e.g., detection of stress in people's voices or detection of threatening stimuli such as rapidly approaching objects). Conversely, $A_P$ can be erroneously increased due to failures in detection of negative properties (e.g., the completion of a complex delivery action will result in an increase in $A_P$ if the robot does not notice that the object to be delivered was lost)–for a detailed exposition of the (complex) relationships between positive and negative affective states see [36].

Both affective states are updated according to the following equation (based on [30]): $\Delta act/\Delta t = trig - act \cdot (trig + dec)$, where $trig \in 0, 1$ reflects the infusion of affect (i.e., 1 for success for $A_P$ or failure for $A_N$, 0 otherwise)[7], $act \in [0, 1]$ is the level of activation, and $dec \in (0, 1)$ is a decay value that

such as Mehrabian's PAD model: "pleasure", "arousal", and "dominance" (e.g., [34, 3]).

[2] We will not be able to describe the distributed and reflective aspects of the architecture here; for the distribution components see [35].

[3] The implementation builds on the ADE system available at `http://ade.sourceforge.net/`. DIARC also makes heavy use of pre-defined components developed by other research teams (e.g., the OpenCV vision library for face detection and various image processing functions, the SONIC speech recognizer for spoken word recognition, the link parser for natural language parsing, VerbNet mappings, and an enhanced version of "Thought Treasure" for natural language understanding and production).

[4] $T_A$ is typically the "time-to-completion" associated with the goal in the script, but can be modified by the action interpreter based on context.

[5] The decision to model positive and negative affective states repeatedly was based on psychological (e.g., [10]) and neuropsychological (e.g., [9]) evidence indicating the representational independence of positive and negative affect.

[6] We are in the process of demonstrating the different effects of positive and negative mood influence on action selection, and thus behavior, in an independent study.

[7] Note that $A_P$ and $A_N$ are not complements. There are actions that can be accomplished without positive affect being triggered (e.g., recognizing words). Similarly, there may be action failures that do not trigger negative affect right away (e.g., when the robot interrupts itself while speaking to produce another more urgent sentence).

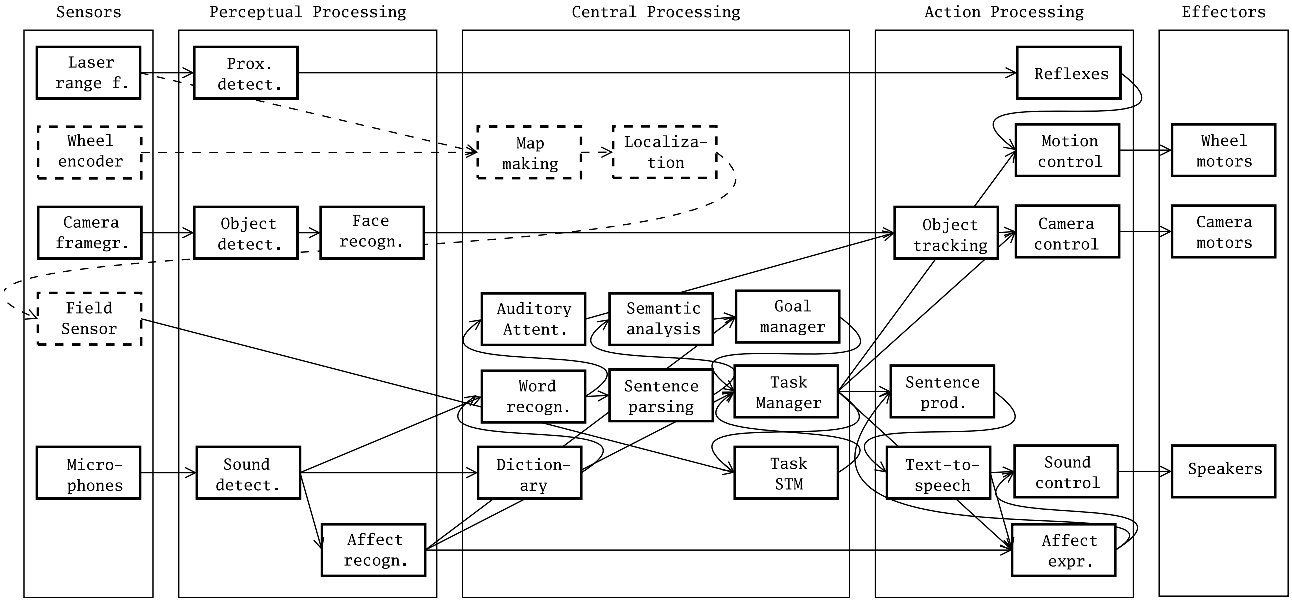

**Figure 1: A partial view of the proposed DIARC robotic architecture for HRI consisting of only those components that were used in the experiment described in this paper. Boxes depict concurrently running components of varying complexity and arrows indicate the information flow through the architecture. Dashed items are related to the simulated field sensor, and are not part of the architecture *per se* (see the experiment description).**

will reduce the activation level over time (in the absence of any triggerings). *Priority*, then, is the product of urgency and importance ($P = U \cdot I$). The goal stack is resorted periodically according to the priorities of its goals, and the goal on the top is executed. This priority mechanism allows the robot to focus on goals that are of importance to its well-being (as determined by the affective evaluation of the goals utilities and costs), while being able to keep multiple other goals around and adapt their priority dynamically based on environmental and internal changes.

## 3.2 Affect Detection in Spoken Language

We only describe the extraction of "stress" in a speaker's voice from the auditory stream (even though the algorithm can be extended to detect other affective features), which was used in the experiment. In [16], empirical studies show that stress in the voice is marked by an increase in the mean of the fundamental frequency ($F_0$) mean and intensity. Because the fundamental frequency is inversely proportional to the pitch period, this means stress can be determined by a decrease in the pitch period. For pitch period estimation, the algorithm implemented in [11] was followed with slight modifications. First, segments of 20 msec sampled at 16 KHz (320 samples) are selected and filtered using a lowpass filter. After that, each speech sample $x$ passes through a three-level clipper *f(x)*, which is defined as 1 if $x > CL$, -1 if $x < -CL$, and 0 otherwise. $CL$ is the clipping level of the speech segment. Given the first 100 samples ($x_1$) and the last 100 samples ($x_2$) of the segment, $CL$ is defined as $0.68 * min(max(x_1), max(x_2)$. The autocorrelation of the clipped result is used to determine the pitch period [28]. The energy of the raw speech signal is calculated and if it falls below an experimentally-determined threshold, the segment is considered "unvoiced" and no further action is taken. Otherwise, if the end of the word is reached (as marked by silence, or after 600 msec), the average frequency of that word is computed and the word is marked as "stressed" if the pitch is higher than the cumulative average pitch.[8]

While this method's stress detection will be speaker-dependent (because the average will be determined by the speaker's voice), the stressed/unstressed state of the speaker will actually be independent of the voice; an external system uses the ratio of stressed words to total words detected over a period of time, and compares it to a threshold. If the ratio exceeds the threshold , then the speaker is classified as "stressed", or "not stressed" otherwise.

This thresholding method is different from methods discussed in [27], because those methods are focused on learning schemes. Rather, it is similar to earlier systems (e.g., [23, 38]), which use general comparisons of properties of the input speech signal to those of a "calm" state (in our case, an increase in pitch correlating to stress). The advantage of the employed system is that it is speaker-independent and requires no training corpus nor specific underlying training algorithm (e.g., statistical learning algorithms as in [6], [19], [27]). It only requires the speaker to speak naturally (i.e. without stress) at the beginning of the program, so the baseline can converge to a true representation of the user's

[8] While word lengths of 600 milliseconds may not generalize to English as a whole, the chosen boundary is acceptable for most current interactions with the robot–a more general system would depend solely on word boundaries.

neutral state. Afterwards, this baseline is locked so further utterances can be measured for affect.

## 3.3 Affect Modulation of Speech

A modified version of the University of Edinburgh's *Festival* system was used for speech synthesis. Based on [8], an emotion filter was applied to the speech output of Festival, altering various speech parameters based on affective state. In particular, we defined various degrees of intensities of emotions for the four categories "sad", "angry", "frightened", and "happy". For example, to give the robot a "frightened" voice, $F_0$ and speech rate were increased, as was the range of $F_0$, and jitter was added to give the voice a quivering sound. These match the results of [16], which states that in fear/panic, $F_0$ mean and range will increase from the normal, as will the speech rate.

# 4. AFFECT-INDUCTION EXPERIMENT

While it seemed clear from the beginning that expressing affect (e.g., via facial expressions, voice, gestures, etc.) would make robots more believable to human observers, there was already some early recognition of the potential utility of affective control for influencing the behavior of people (e.g., [5]). Moreover, studies with robots and simulated agents showed that affect mechanisms can facilitate task performance of artificial agents and may be cheaper than other, more complex non-affective mechanisms (e.g., [22, 33]).

Encouraged by recent findings from usability studies in HRI about facilitatory effects of affect recognition (e.g., that recognizing affect can help to improve speech recognition results [17]), we set out to test the main hypothesis that *affect expression based on internally generated affect or affect generated in response to affect in humans can help improve the performance of mixed human-robot teams on tasks that have to be performed together.*

To be able to test the hypothesis, a task with (at least) the following characteristics is required:

- at least one robot and one human are needed for the task and neither robot nor human can accomplish the task alone
- robot and human have to exchange information in order to accomplish the task (in our case via *spoken natural language*)[9]
- there is a performance measure (in our case *time-to-task-completion*) that can be evaluated objectively on task performance alone rather than being dependent on subjective ratings
- the task must include aspects of human affect, which can be influenced by the robot (in our case *affective modulation* of robot speech output)
- these aspects of human affect (in our case *stress*) must be triggerable (e.g., via cognitive tasks, time pressure, etc.) before or during the task (in our case we induce stress as described below via *time pressure*)
- a control condition is needed where the same aspects of human affect are not influenced by the robot (in our case no affective modulation of robot speech output)

Note that while the first three items are common to many joint human-robot tasks, the second three are specific to testing the utility of affect for task performance.

To keep the interaction as natural as possible (e.g., no handheld microphones or tethering to the robot), we let subjects freely interact with the robot (even during training phase we only suggested to them the kinds of commands the system would understand without actually pointing to limitations about what it would not understand). We also forfeit any speaker-dependent adaptation of the employed voice recognition system (at the expense of the overall recognition rate) to keep training phase to a minimum.[10] This was partly possible because the task-specific vocabulary was very small and thus the speaker-independent recognition rate acceptable.

## 4.1 The Task

We decided on a task that is relevant to NASA's envisioned future space explorations with joint robot-human teams [14]. The task takes place against the backdrop of a hypothetical space scenario. A mixed human-robot team on a remote planet needs to determine the best location in the vicinity of the base station for transmitting information to the orbiting space craft. Unfortunately, the electromagnetic field of the planet interferes with the transmitted signal and, moreover, the interference changes over time. The goal of the human-robot team is to find an appropriate position as quickly as possible from which the data can be transmitted. The specific goal of the human is to steer the robot using natural language commands until it has found a viable transmission location.

**Experimental Setup:** This envisioned space scenario is simulated in a room of approximately 5m x 6m (see Figure 2). During the experiment, the robot maintains an internal map of the area, with a set of six fixed points representing locations of local peaks for potential transmissions.[11] Each peak has a strength $S_P$ ranging between 200 and 500, decreasing proportionally with the distance of the robot from the peak at a rate of one unit per cm. For overlapping fields, the maximum is chosen. The location of these points is unknown to the subjects, but the same across all subjects (similarly, the initial location of the robot is the same across all runs). Only two locations have sufficient $S_P$ for transmission. To learn about the field strength at the current location ($S_C$), subjects request a reading from the robot.

[9] This is necessary to exclude trivial "team tasks" such as situations where the robot has to find a target while the human has to solve a mathematical problem and the "joint task" is accomplished if each individual subtask is accomplished.

[10] In retrospect, we believe that our results might have been even more pronounced had we used online speaker adaptation during the training phase to improve the recognition. Moreover, a wireless microphone could be attached to the subject to reduce noise and further improve recognition.

[11] Note that the map in this experiment is not a proper part of the robot's architecture.

The robot checks a "simulated field sensor," which effectively returns $S_C$ for the current location. To successfully transmit, $S_C$ must be greater than 400 units.

**Equipment:** The robotic platform for the experiment is a Pioneer ActivMedia Peoplebot (P2DXE) with a pan-tilt-zoom camera, a SICK laser range finder, two microphones, two speakers, three sonar rings, and an onboard 850 MHz Pentium III. In addition, it is equipped with two PC laptops with 1.3 GHz and 2.0 GHz Pentium M processors. All three run Linux with a 2.6.x kernel and are connected via an internal wired ethernet; a single wireless interface on the robot enables system access from outside the robot for the purpose of starting and stopping operation. Obstacle detection and avoidance is performed on the onboard computer, while speech recognition and production, action selection, and subject affect recognition are performed on the laptops.

**Method:** For the purposes of this experiment, we employ three test conditions: control, self, and other. The *control* condition utilizes no affect expression. The robot's voice remains neutral throughout the task. In the *self* condition, voice affect is modulated by the robot's inner affect states. Specifically, the stress internally generated by the urgency of the top-most goal on the goal stack is expressed by increasing the "fearfulness" of the robot's voice as time passes. In the *other* condition, voice affect is modulated whenever stress is detected in the subject's voice (i.e., negative affect is triggered, leading to an increase in the activation level of $A_N$, which, in turn, causes a modulation of the affective speech output). Since our current affective speech production system can only produce discrete modification to voice output, the continuous affective states are mapped onto discrete affective voices (depending on the intensity levels of $A_N$), e.g., "half-frightened" and "frightened" to indicate stress levels.

**Procedure:** Subjects are first asked to fill out a pre-survey with five basic questions about their views on different aspects of robots used for HRI (see Table 1). The same five questions are also included on the post-survey to test whether the experiment would have any influence on their perceptions of interactive robots (similar to [18]). Then an experimenter reads the "background story" (summarized in the above task description). The subjects are told that their goal is to control the robot to find a transmission location as quickly as possible. Before attempting the actual task, subjects go through a *practice period* during which they become acquainted with the robot by interacting with it in natural language. In particular, they are asked to help the robot explore its environment using commands such as "go forward", "turn right", "take a reading", etc. During practice, the robot does not employ affective speech modulation. This practice phase lasts at most ten minutes.

To ensure that subjects' affective states will be altered during the experiment, we artificially induce stress in subjects by having the robot issue a battery warning: "I just noticed that my battery level is somewhat low, <name>, we have to hurry up." After another minute, the robot issues another warning: "<name>, my battery level is very low, we have only one minute left." When a total of three minutes has elapsed, the robot indicates that its battery has died and the task has failed. Subjects may not reach all of these interaction points if they achieve transmission early enough. In both affective conditions, the robot's voice remains neutral for the first minute of the task. Thereafter, the voice is modulated to express elevated stress starting with the first battery warning in the *self* condition, and again to express even more stress at the second battery warning. Voice modulation remains elevated for all interactions (e.g., field strength reports). In the *other* condition, the robot's voice only changes temporarily after the first minute, when the affect recognition module detects stress in the subject's voice (otherwise the voice remains neutral). Note that it is, therefore, possible that some subjects in this condition will never hear a modulated voice if the robot never detects stress (such subjects are classified as "control", see also footnote 12). The performance of the team is measured in in terms of the time it takes the team to find a valid transmission location and transmit the data. Throughout the experiment, the robot's motion trajectory, speech produced and detected, and the state of the affect recognition module were recorded.

After the experimental run, subjects are asked to fill out the post-survey, which, in addition to the 5 pre-survey questions, also has questions about whether subjects felt "stressed" at the beginning of the experiment and after the robot announced that it was running low on battery power.

**Participants:** 24 male subjects were recruited from the pool of Computer Science and Engineering students and randomly assigned to the three groups.[12]

## 4.2 Results

First, we compared the results of two questions on the survey related to the stress that subjects experienced during the experiment to make sure that affect induction in subjects worked as expected: "How stressed did you feel at the beginning of the task?" and "How stressed did you feel after the robot announced for the first time that its batteries were running low?" We found a high statistical difference in subjects' self-assessed stress levels before ($\mu = 3.70, \sigma = 2.01$) and after ($\mu = 5.67, \sigma = 1.74$) the robot had announced that its battery was running low (F(1,22)=17.773,p<0.001). We conducted an additional ANOVA to confirm that there was no difference among the three groups with respect to the change in self-reported stress levels (F(2,21)=0.8,p=0.451).

A 3-way ANOVA with *affect* ("control", "self", and "other") as independent and *time* (to task completion) as dependent variable shows only a slight trend towards significance, but no significant effect (F(2,21)=2.508, p=0.106). This is due to (1) the relatively small number of subjects and (2) the fact that the performance time of subjects who were not able to complete the task is taken to be the time when the experiment was ended (i.e., slightly over 180 sec.) rather than

[12] Originally, seven subjects were assigned to each group, but since some subjects in the "other" group ended up finishing the task either before the robot was allowed to express affect or without the robot having detected any stress, they were added to the "control" category and 3 additional subjects were recruited to be able to have about the same number of "self" and "other" subjects without having too many "control" subjects.

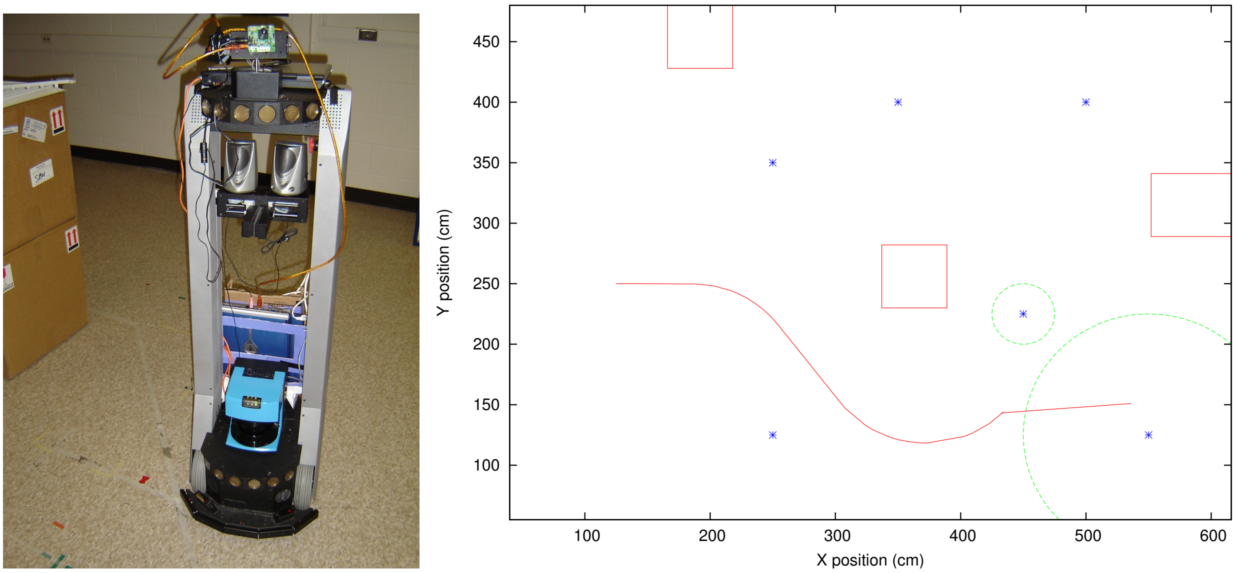

**Figure 2: The robot used in the experiment (left) and a typical trajectory of the robot in an run (right–green circles indicate transmission regions with sufficient field strength, blue crosses indicate indicate local peaks of the field, red boxes indicate "rocks", i.e., obstacles).**

including a "time penalty" for failing the task. With regard to (1), we get a significant effect if we compare the combined affective to non-affective groups (F(1,22)=4.882,p=0.038). With regard to (2), we get a significant difference between all three groups if we use *success* as dependent variable instead of *time* (F(2,21)=5.958,p<0.01). Hence, the results confirm our main hypothesis that the expression of affect (at the right time) both based on internally generated affect as well as affect generated in response to affect in humans can improve the performance of mixed human-robot teams on tasks that have to be performed together.

We also compared the five identical pre- and post-survey questions in order to determine whether the experience and interaction with the robot had any influence on the subject's views on basic questions about HRI (Table 1). We conducted ANOVAs for all five questions with *pre* and *affect* as independent, and *post* as dependent variable. In all cases we found a significant effect of *pre*, but no significant effects of *affect*. However, for questions 2 and 5 we found significant interactions between *pre* and *affect* indicating that subjects in the "self" affect group changed their ratings more so than the other groups. While the difference between pre- and post-survey ratings is not significant in either case for the "self" group ($\mu = 5.00, \sigma = 1.73$ vs. $\mu = 6.72, \sigma = 0.95$ for question 2, and $\mu = 4.71, \sigma = 2.36$ vs. $\mu = 6.57, \sigma = 1.52$ for question 5), this is only due to the small number of subjects in that group (N=7) and we expect this difference to become significant with a larger number of subjects. Interestingly, subjects' views on question 4 did not change based on the experiment, which suggests that for them "detecting human emotions and to reacting to them" is separate from "having emotions and expressing them".

## 5. RELATED WORK

The two closest affective robotic architectures in terms of using emotions for internal state changes and action selection are [22, 18] and [4]. [22] implement emotional states with fixed associated action tendencies in a service robot as a function of two time parameters ("time-to-refill" and "time-to-empty" plus two constants). Effectively, emotion labels are associated with different intervals and cause state transitions in a Moore machine, which produces behaviors directly based on perceptions and emotional states. This is similar to the way *urgency* is calculated in our action interpreter, but different from the explicit goal representation used in our architecture, which allows for the explicit computation of the *importance* of a goal to the robot (based on positive and negative affective state), which in turn influences action selection (e.g., urgency alone may or may not result in reprioritization of goals and thus changes in affective state). Moreover, none of the robots in [22, 18] use (spoken) natural language to interact with humans nor do they detect human affect.

The architecture in [4] extends prior work [6] to include natural language processing and some higher level deliberative functions, most importantly, an implementation of "joint intention theory" (e.g., that allows the robot to respond to human commands with gestures indicating a new focus of attention, etc.). The system is intended to study collaboration and learning of joint tasks. One difference is that our robot lacks the ability to produce gestures beyond simple nodding and shaking by the pan-tilt unit (although it is mobile and fully autonomous as opposed to the robot in [4]). More importantly, the mechanisms for selecting subgoals, subscripts, and updating priorities of goals seem different in our affective action interpreter, which uses a dual representation of positive and negative affect that is influenced by

**Table 1: Comparison of pre- and post-survey questions for all three groups (from 1=*strongly disagree* to 9=*strongly agree*).**

| Question | Pre $\mu(\sigma)$ | Post $\mu(\sigma)$ |
|---|---|---|
| Would you prefer robots that understand natural language over robots that can be controlled via the keyboard? | 6.21 (1.96) | 6.46 (1.72) |
| Do you think it will be useful for robots to detect and react to emotions in humans? | 5.54 (1.59) | 6.25 (1.45) |
| Do you think it is a good idea for robots to have their own personality? | 5.42 (1.91) | 5.08 (1.77) |
| Do you think it will be useful for robots to have emotions and express them? | 4.58 (1.82) | 4.67 (1.76) |
| Do you think it is a good idea for robots to have their own goals and be somewhat autonomous rather than fully controlled by people? | 5.42 (2.41) | 6.17 (2.36) |

various components in the architecture and used for the calculation of the importance, and consequently the priority, of goals.[13]

The experiment closest to ours in spirit is that of [29], where physiological sensors are attached to subjects to measure cardiac, electrodermal, and electromyographic responses. They are combined via a fuzzy logic system to obtain an overall "anxiety level" in real-time, which is then fed as input into a simple subsumption-based robotic control architecture, where it can cause the robot to interrupt its exploratory wandering behavior if it reaches a certain threshold. The results demonstrate that anxiety levels of humans (performing cognitive tasks of varying difficulty) can be detected in real-time. Different from our experiment, the robotic system–except for detecting human affect–seems decoupled from the human and the two tasks performed by the robot and the human are unrelated.

## 6. CONCLUSIONS

In this paper, we have proposed an architecture for HRI tasks involving joint human-robot teams, which can detect, generate, and express affect in novel ways. Since we share the belief of [15] that "peer-to-peer HRI will enable more effective and productive human-robot teams for space exploration", our research attempts to elucidate the potentially facilitatory roles of affect recognition and expression for task performance in joint human-robot teams. As we have demonstrated in the HRI experiment, in which success critically depended on collaboration between human and robot, it is not only critical to recognize human non-verbal, affective cues to improve the interaction between robots and people, but affect generated by mechanisms within the robot's architecture can *actually* improve the task performance of joint human-robot teams. And while these are clearly early results, we believe that they nevertheless point towards the potential of affect-aware and affect-generating architectures for HRI as an important direction for future research in human-robot collaboration.

## 7. ADDITIONAL AUTHORS

Additional authors: Christopher Middendorff (A.I. & Robotics Lab, Notre Dame), email: `cmidden1@cse.nd.edu`.

[13] The details for reprioritization of goals were not provided in [4].

## 8. REFERENCES

[1] R. C. Arkin and T. R. Balch. Aura: principles and practice in review. *JETAI*, 9(2-3):175–189, 1997.

[2] H. Bless, N. Schwarz, and R. Wieland. Mood and the impact of category membership and individuating information. *European Journal of Social Psychology*, 26:935–959, 1996.

[3] C. Breazeal. Regulating human-robot interaction using 'emotions', 'drives', and facial expressions. In *Proceedings of Autonomous Agents 1998*, pages 14–21, Minneapolis, MO, 1998.

[4] C. Breazeal, G. Hoffman, and A. Lockerd. Teaching and working with robots as a collaboration. In *Proceedings of AAMAS 2004*, 2004.

[5] C. Breazeal and B. Scassellati. How to build robots that make friends and influence people. In *IROS*, pages 858–863, 1999.

[6] C. L. Breazeal. *Designing Sociable Robots*. MIT Press, 2002.

[7] J. Burke, R. Murphy, E. Rogers, V. Lumelsky, and J. Scholtz. Final report for the DARPA/NSF interdisciplinary study on human-robot interaction. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 34(2):103–112, 2004.

[8] F. Burkhardt and W. Sendlmeier. Verification of acoustical correlates of emotional speech using formant-synthesis. In *Proceedings of the ISCA Workshop on Speech and Emotion*, 2000.

[9] R. J. Davidson. Cerebral asymmetry and emotion: Conceptual and methodological conundrums. *Cognition and Emotion*, pages 115–138, 1993.

[10] E. Diener and R. A. Emmons. The independence of positive and negative affect. *Journal of Personality and Social Psychology*, 47:1105–1117, 2002.

[11] J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner. Real-time digital hardware pitch detector. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 24(1):2–8, 1976.

[12] P. Ekman. Facial expression and emotion. *American Psychologist*, 48(4):384–392, April 1993.

[13] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42:143–166, 2003.

[14] T. W. Fong and I. Nourbakhsh. Interaction challenges in human-robot space exploration. *Interactions*, 12(2):42–45, March 2005.

[15] T. W. Fong, I. Nourbakhsh, R. Ambrose, R. Simmons, A. Schultz, and J. Scholtz. The peer-to-peer human-robot interaction project. In *AIAA Space 2005*, September 2005. AIAA-2005-6750.

[16] T. Johnstone and K. Scherer. Vocal communication of emotion. In M. Lewis and J. Haviland, editors, *Handbook of Emotion, 2nd ed.*, pages 220–235. Guilford, 2000.

[17] T. Kanda, K. Iwase, M. Shiomi, and H. Ishiguro. A tension-moderating mechanism for promoting speech-based human-robot interaction. In *IROS*, pages 527–532, 2005.

[18] C. L. Lisetti, S. Brown, K. Alvarez, , and A. Marpaung. A social informatics approach to human-robot interaction with an office service robot. *IEEE Transactions on Systems, Man, and Cybernetics–Special Issue on Human Robot Interaction*, 34(2):195–209, 2004.

[19] S. McGilloway, R. Cowie, E. Cowie, S. Gielen, M. Westerdijk, and S. Stroeve. Approaching automatic recognition of emotion from voice: a rough benchmark. In *ISCA Workshop on Speech and Emotion*, 2000.

[20] F. Michaud and J. Audet. Using motives and artificial emotion for long-term activity of an autonomous robot. In *Proceedings of the 5th Autonomous Agents Conference*, pages 188–189, Montreal, Quebec, 2001. ACM Press.

[21] L. Moshkina and R. Arkin. On TAMEing robots. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pages 3949 – 3959, 2003.

[22] R. R. Murphy, C. Lisetti, R. Tardif, L. Irish, and A. Gage. Emotion-based control of cooperating heterogeneous mobile robots. *IEEE Transactions on Robotics and Automation*, 18(5):744–757, 2002.

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

[24] I. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto. An affective mobile educator with a full-time job. *Artificial Intelligence*, 114((1-2)):95–124, 1999.

[25] R. Pfeifer. Artificial intelligence models of emotion. In V. Hamilton, G. H. Bower, and N. H. Frijda, editors, *Cognitive Perspectives on Emotion and Motivation, volume 44 of Series D: Behavioural and Social Sciences*, pages 287–320. Kluwer Academic Publishers, Netherlands, 1988.

[26] R. W. Picard and J. Healey. Affective wearables. In *ISWC*, pages 90–97, 1997.

[27] O. Pierre-Yves. The production and recognition of emotions in speech: features and algorithms. *International Journal of Human-Computer Studies*, pages 157–183, 2002.

[28] L. R. Rabiner. On the use of autocorrelation analysis for pitch detection. *IEEE Trans. Acoust. Speech and Signal Proc.*, AASP-25(1):24–33, 1977.

[29] P. Rani, N. Sarkar, , and C. A. Smith. Affect-sensitive human-robot cooperation-theory and experiments. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pages 2382–2387, 2003.

[30] M. Scheutz. The evolution of simple affective states in multi-agent environments. In D. Cañamero, editor, *Proceedings of AAAI Fall Symposium*, pages 123–128, Falmouth, MA, 2001. AAAI Press.

[31] M. Scheutz. Affective action selection and behavior arbitration for autonomous robots. In H. Arabnia, editor, *Proceedings of the 2002 International Conference on Artificial Intelligence*, page 6 pages. CSREA Press, 2002.

[32] M. Scheutz. Useful roles of emotions in artificial agents: A case study from artificial life. In *Proceedings of AAAI 2004*, 2004.

[33] M. Scheutz and B. Logan. Affective versus deliberative agent control. In S. Colton, editor, *Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing*, pages 1–10, York, 2001. Society for the Study of Artificial Intelligence and the Simulation of Behaviour.

[34] M. Scheutz and B. Römmer. Autonomous avatars? from users to agents and back. In A. de Antonio, R. Aylett, and D. Ballin, editors, *Intelligent Virtual Agents, Third International Workshop, IVA 2001, Madrid, Spain, September 10-11, 2001, Proceedings*, volume 2190 of *Lecture Notes in Computer Science*, pages 61–71. Springer, 2001.

[35] M. Scheutz, P. Schermerhorn, C. Middendorff, J. Kramer, D. Anderson, and A. Dingler. Toward affective cognitive robots for human-robot interaction. In *AAAI 2005 Robot Workshop*, 2005.

[36] A. Sloman, R. Chrisley, and M. Scheutz. The architectural basis of affective states and processes. In J. Fellous and M. Arbib, editors, *Who needs emotions? The Brain Meets the Machine.* Oxford University Press, New York, forthcoming.

[37] J. Velásquez. When robots weep: Emotional memories and decision-making. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pages 70–75, Menlo Park, 1999. AAAI, CA, AAAI Press.

[38] U. Williams and K. Stevens. Emotions and speech: some acoustical correlates. *JASA (52)*, pages 1238–1250, 1972.