A Platform for Affective Agent Research

Winslow Burleson, Rosalind Picard; Ken Perlin, Jon Lippincott
MIT Media Lab; New York University Media Research Laboratory
{win, picard}@media.mit.edu; {perlin, jal363}@nyu.edu

Abstract

Accurately interpreting and expressing affect is fundamental to empathetic relationships. A platform for sensing and interpreting several aspects of users' nonverbal affective information and responding through an expressive agent has been developed. The platform includes integration of multi-modal affective sensors with a real time inference engine, a behavior engine, and a 3d scriptable expressive humanoid agent within a graphical virtual environment. Currently the sensors include a pressure-sensitive mouse, a BlueTooth wireless skin conductivity sensor, a TekScan pressure sensor on a chair, and a stereo head tracking system as well as an IBM Blue Eyes infrared-sensitive camera. These sensors feed into custom algorithms for analysis of individual channels of information, such as postural and facial expressions, which in turn are combined with additional channels of information to make an inference about the user's affective state. The system further synchronizes this sensor data with the agent behaviors and with video of the user and his or her on-screen activity. This platform is seen as a general-purpose tool applicable to research in several areas, including how to design an affective learning companion, and how to further basic understanding of empathy and emotion contagion in human-agent interaction.

1. ACM classification keywords

Agent architectures, Agent programming languages and environments, Sensors, Emotion, Affective user interface, Agent and intelligent systems, E-Learning and education, children, Pedagogical agents.

2. Introduction

Most computational agents show expressive behaviors, often via facial movements or various gestures. Affective expressions have been argued to be useful to help make agents "believable" [1]. Expressive behaviors have additionally been associated with useful outcomes such as making agents likeable [2,3,4]. In more recent systems, agent expressions have been responsive to human expressions, contributing to making agents "relational," able to construct long-term social-emotional relationships with users. For example, the Laura agent, expanding on the empathic "frustration-handling" agent of Klein et al [5], received verbal (text-only) expressions of a variety of affective states from the user and responded with both verbal and nonverbal expressions of empathy based not only on what you expressed now, but also considering what you have expressed in the past [6]. Thus, the agent could respond to your statement "I'm feeling down" by moving in closer to you, displaying a facial expression of concern, and speaking an appropriate verbal response such as "Sorry to hear that." If day after day you continued to indicate these feelings, her wording would change to acknowledge the ongoing problem, and if things escalated, she would refer you for medical help. However, to date, there are no examples of agents that can sense natural (both verbal and non-verbal) human communication of emotion and respond in a way that rivals that of another person.[1]

Recognizing and responding to affective information is a vital part of natural intelligent interaction. These two skills are widely recognized as components of so-called "emotional intelligence" [7,8]. If an agent cracks a joke, and Bobby smiles while Cynthia frowns, then it would probably be fine for it to flash a smile back at Bobby, while the same expression back at Cynthia might be perceived as mean. Depending on the agent's goals, one response could be much more intelligent than another. If an agent winks and does a cute little dance that irritates you, and if it repeats that little dance and you show increased irritation and perhaps visible anger at it, then it might be wise for it to be able to see your response and subsequently to act in a way that acknowledges its failure, that is if the goal includes wanting you to have a favorable impression of it. How someone chooses to respond to your emotion greatly colors your opinions of their competence, trustworthiness, likeability, and more.

In our work we do not assume that human-human interaction is the same as human-computer interaction, nor do we see the need to limit the development of systems to this objective. However, we recognize that there is a lot to learn from many findings, e.g., those of Nass, Reeves and Moon, [10,11], that illustrate that results from studies of

[1] Note that there is significant progress in robotics in this area, namely the work of Breazeal and her robots that sense vocal affect and respond with affective facial and head movements [9].

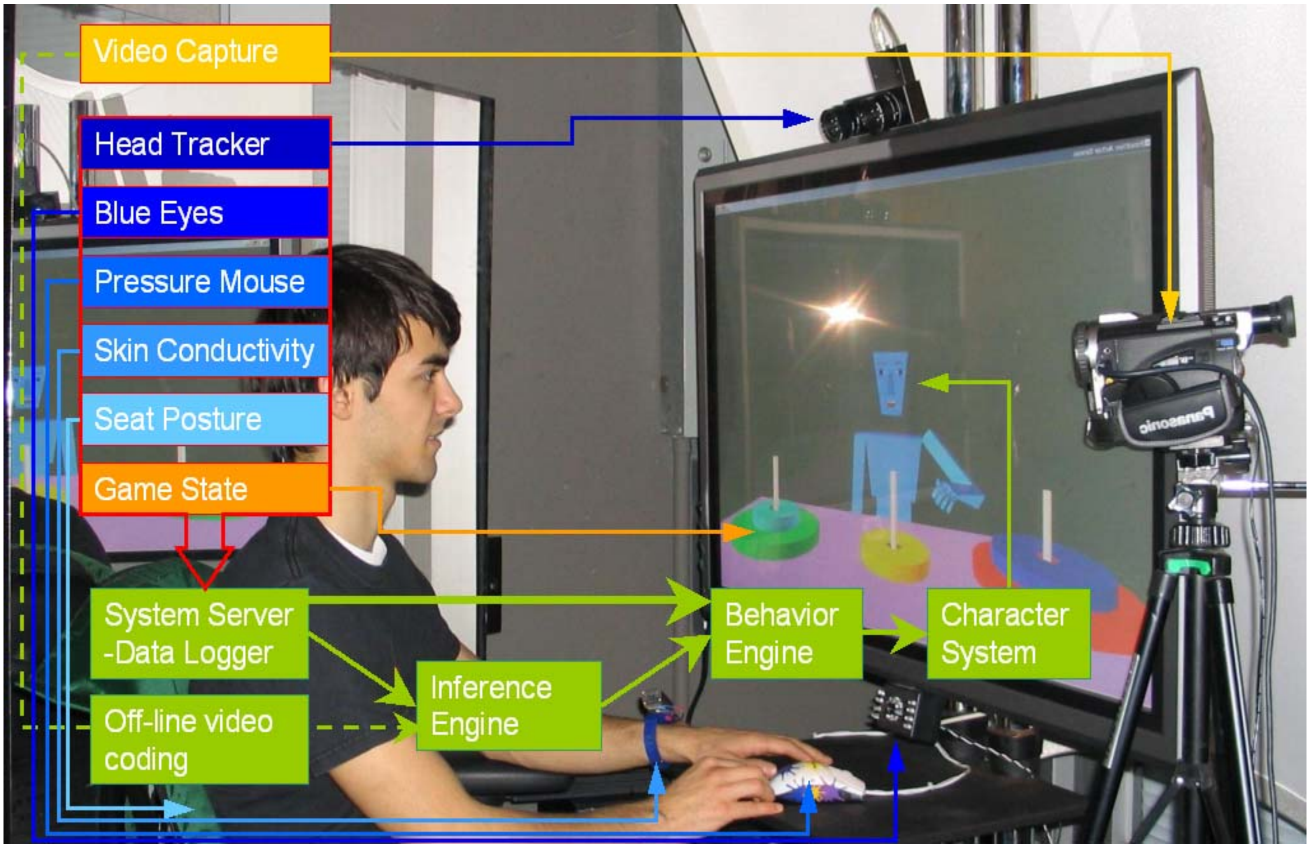

Figure 1: Learning Companion "Casey" being developed with new affective agent platform.

human-human interaction can usefully inform the design of human-computer interaction.

One example that illustrates the importance of an appropriate affective response is that of Robinson and Smith-Lovin, [12] who described how if a person responds positively to something bad happening, then you will like them less. Alternatively, if they respond in a way that is affectively congruent, then you will like them more. These findings sound supportive of the current approach in pedagogical agent research where the character smiles when you succeed and looks disappointed if you make a mistake or fail. However, it is also known that pre-school subjects smile as much after failure as they do after success [13]. We have also found that adults sometimes smile after failure, in our studies of task load in driving situations when a subject makes a simple math mistake (while talking on the phone while driving). In other words, the human expression is not necessarily affectively congruent with the task, so it is unclear what the agent expression should be if it is to be perceived empathetically. In human-human interactions it is important to realize that there are several different conditions for smiles, including nervous, humor, and success. We expect that such interactions will also hold true for human-agent interactions. However, to achieve deeper understanding of such interactions requires new advances, namely the development of technology that can recognize and respond in real time to affect.

Consider, for example, the scientific challenge of understanding the role of affect when coaching a learner to solve a difficult puzzle. One of our colleagues, Barry Kort, has been coaching children for nearly twenty years with puzzles such as the Tower of Hanoi (a stack of disks on one of three poles, where the goal is to move the stack to another pole, one disk at a time, without ever placing a larger disk on top of a smaller.). Through his interactions, he has developed several hypotheses about how emotion is communicated during learning and how it interacts with learning; however, when he is standing together with a learner, he is unable to precisely and repeatedly measure and respond objectively to students' affective expression to confirm his theory. For example, he is unable to control his own facial and postural expressions and whether or not they mirror those of the learner. Meanwhile, it is known that not only overt facial expressions, but also subtle aspects of body movement such as mirroring of body position have been shown to be an important part of interpersonal relationships, constituting one of the non-verbal channels of expression, interpretation, and communication [14]. In order to develop theories of how affect influences learning, advances are needed to enable researchers to sense and respond to such movements in precise and controllable ways. While most people cannot bring all these movements under precise control, a computational agent can. Thus, this new platform should enable fundamental scientific investigations of the role of affect in a wide array of interactions, including learning.

This paper describes a state of the art platform designed for such investigations. The paper is structured as follows. The next Section overviews the architecture of the platform. The subsequent two Sections will describe the affect recognition system, followed by the expressive *character system*. Finally, we will describe a few planned research applications of the platform.

3. Architecture

The research platform and architecture we are developing focuses on the sensing and analysis of signals related to affect, and on the ability to interpret and respond to these, in real time, with an expressive scriptable agent. While our first research objective is to construct a system that lets us explore the role of empathy in the development of affective learning companions, we also see that this system can enable new opportunities to extend and apply many important findings of the social sciences. Thus, we have in mind several general system features as well as specific particular system features that will appear to a user to be a learning companion.

The platform consists of an architecture for the integration of multi-modal affective sensors; a real time *inference engine*; a *behavior engine*; and a 3d scriptable expressive humanoid agent within a graphical virtual environment (see Figure 1). The user sits in front of a wide screen plasma display. On the display appears an agent and 3d environment. The user can interact with the agent and can attend to and manipulate objects and tasks in the environment. The chair that the user sits in is instrumented with a high-density pressure sensor array and the mouse detects applied pressure throughout its usage. The user also wears a wireless skin conductivity sensor on a wristband with two adhesive electrode patches on their hand and forearm. Three cameras in the system, a video camera for offline coding, the blue-eyes facial action units camera, and a stereo-head tracking camera, record and sense additional elements of human behavior.

Our approach to recognizing affect is a multi-modal one, not restricting the inference to any one channel (e.g. only facial expression) but rather sensing a broad spectrum of information and applying techniques from psychophysiology, emotion communication, signal processing, pattern recognition, and machine learning, to make an inference from this data. Since any given sensor will have various problems with noise and reliability, and will contain only limited information about affect, the use of multiple sensors should also improve robustness and accuracy of inference. Recent work on multi-modal affective sensing has shown this to be an effective strategy in the development of affective sensing [15].

Affect sensing is integrated into this system as follows. Each sensor sends its signal via a socket to the *system-server & data logger*. Upon receipt of a signal a time stamp is generated and the data is stored in a local text file. This text file is then submitted to the *inference engine*, which has previously been trained on pilot subject data for relevant affective state identification through semi-supervised machine learning. In the case of a learning companion the *inference engine* will be trained to identify states such as frustration, interest or boredom, and pleasure, among others. The *inference engine* generates the state information that along with the relevant sensor-data, is submitted to the *behavior engine*.

The *behavior engine* in turn further interprets the state information and the sensor-data to generate real-time interactions from a repertoire of pre-scripted behaviors as well as through the direct generation of "serendipitous/real-time/new interactions". The pre-scripted behaviors are composed of a sequence of actions and timing commands, and together represent the "behavior space" [16]. Since multi-threaded action procedures, where actions can occur simultaneously, are also supported by the *character system* these can be initiated and controlled through variables by the *behavior engine*. This approach enhances Lester's "dynamic behavior sequencing", by incorporating inter-sequence restructuring of pre-scripted behaviors. This allows for the character to pursue an action such as looking from the user to an object and back while nodding or shaking their head concurrently. In a pre-scripted sequence, the decision to nod or shake the character's head is determined by variables controlled by the *behavior engine*. An obvious time for the use of a pre-scripted behavior is when a learning companion initiates the "introduction" script when a user first sits in the chair. The introduction script has the character introduce his/herself engage in small talk and present the Towers of Hanoi activity. Scripted behaviors mediated by the *behavior engine* are seen as "empathic interactions" but are likely to be less rich than "serendipitous interactions".

One example of a "serendipitous interaction" is the development of "shared attention" for an action or object. This can occur when a learning companion looks at the same place, in the virtual or physical environment as the user is looking. The companion will look at the user and then return its gaze to the point of shared interest. The *behavior engine* will decide if developing "shared attention" is likely to be appropriate based on its previous behaviors, the current task-state, and affective sensor readings. Another example of an unscripted interaction is the generation of "reciprocal facial expressions" based on the facial action unit analysis of blue-eyes. When the *behavior engine* is presented with information from the *inference engine* that there has been a smile, frown, or indication of frustration it will decide if and for how long the character's expression should mirror the user. Both of these are seen as "empathic interactions" that are facilitated to a greater extent by the multi-modal sensor system than by a more traditional/impoverished system.

The character is an approachable blockish humanoid. It is situated in an OpenGL environment that can be manipulated both by the character and by the user. For the learning companion scenario the objects in this environment are a table and a Towers of Hanoi puzzle. The user mouse interactions and the task-state of this puzzle are treated as sensor input provided to the *behavior engine* for the construction of further behavioral directives. In this description we have tried to emphasize that the agent and environment are both scriptable and under dynamic control at the same time. There are general macro-scale scripts such as “introduction”, “first-intervention”, “response-to-frustration”, “intervention-during-frustration”, “response-to-progress” and “celebration-of-success”. There are also meso-scale decisions made within these scripts, based on *behavior engine* variables. This structure provides the ability for real-time subtle interactivity concurrent with the macro-level scripts. This ability is useful in the development of interactions, such as the “shared-attention” and “reciprocal facial expressions” described above and in the direct mirroring by the agent of posture as sensed by the seat and head tracking systems (slumping or leaning forward). This shared level of control is discussed further in a later section of this paper.

While non-verbal interactions are important the system does support verbal interactions through pre-generated TTS dialogue. Wav files generated using NextUp.com’s TextAloudMP3 product, [17] are called upon by the *behavior engine* and through the presentation of text bubbles presented above the character. It allows for users to respond through clicking on multiple-choice text responses and by clicking on objects in the environment such as the puzzle disks and poles in response to questions (see Figure 2). This type of asymmetric text and voice based interaction, while not as ideal as a (future) symmetric natural language processing system, has been shown to be effective for “Laura” the physical trainer agent [18].

The current architecture is modular, and in the future could be augmented in various ways, several of which we’ll describe briefly here. The control and dynamic generation of appropriate behavior is an ongoing focus of the development of agents [16]. In terms of real time interaction, the BEAT System [19], focusing on dynamic generation of appropriate physical rapport (body language) synchronized with verbal discourse, and the COLLAGEN system’s discourse models emphasize the importance of immediacy and responsiveness [20]. While the character does not have discourse planning or BEAT-Shaping for its gestures and rapport, these are important future considerations. As our system develops we hope to integrate elements of these approaches into the modular platform.

The current architecture treats the user’s affective and other behavioral information as an input that, over various time scales, is mapped to one of several outputs via a set of probabilistic as well as rule-based mechanisms. The machine learning that it uses is currently via offline processes that require designer intervention. One key area of future augmentation would involve giving the system on line learning mechanisms, so that it can modify its display rules based on what appears to be working for a particular user, and improve its responses while interacting with that user.

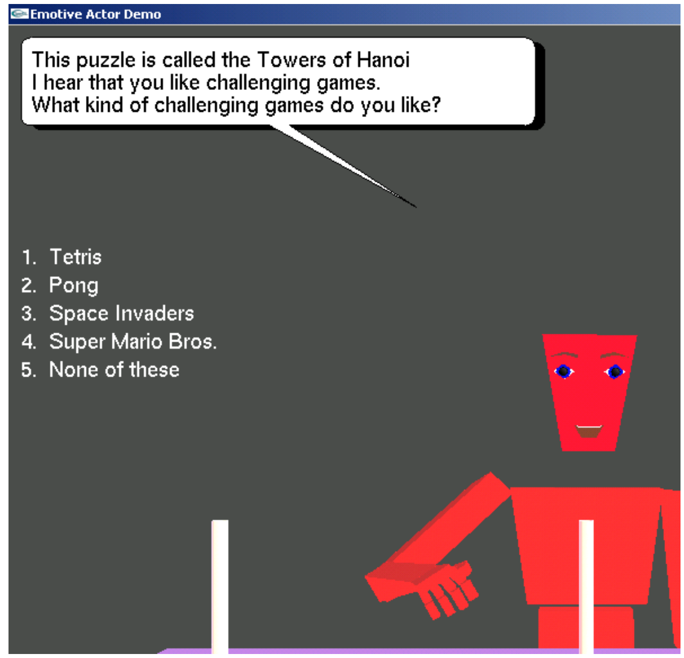

Figure 2. Character dialogue.

Lastly, the current architecture mechanisms do not construct a persistent model of the user’s affective state or other characteristics; instead, the system uses fixed methods for responding to affective cues. For future real-world application development, the platform will benefit from these and possibly other more complex adaptive mechanisms.

Although we are very interested in these augmentations, the aim of this version of the platform is to enable careful control of system responses to test various theories about affective responses in learning. Thus, for the first version we want the mappings to be fixed, in order to provide explicit and direct, repeatable control at a level that can facilitate answering basic scientific questions about affect communication.

4. Affective sensing and inference

Existing agent systems typically infer human affect by sensing and reasoning about the state of a game or an outcome related to an action taken by the user within the computing environment. Use of such an approach is illustrated by the pedagogical agent COSMO, who applauds enthusiastically and exclaims “Fabulous!” if the student takes an action that the agent infers as deserving of congratulations [21]. One can imagine cases where this would be warmly received and perhaps reciprocated with a smile by the user, and cases where it would not.

While reasoning based on a user’s directly input behaviors is important and useful, it is also limited. For example, COSMO has no ability to see how the user responds nonverbally to its enthusiasm: did the user beam with pride, or did she frown and perhaps roll her eyes, as if to say that COSMO’s response was excessive, or otherwise inappropriate. If the latter, it might be valuable for

COSMO to acknowledge its gaffe, thus making it less likely the user will hate it or ignore it in the future. Thus, we wish to advance agent capabilities to include perceptual sensing of nonverbal affective expressions together with the channels that are traditionally sensed in interactive agent systems.

Affect can be expressed in many ways –- not just through facial expressions and gestures, but also through the adverbs of pretty much any aspect of the interaction. Affect modulates how you type and click, what words you choose to speak and how you speak them, as well as how you fidget in your chair and how you move your head and facial muscles. Our approach is one of integrating many channels of information in order to better understand how affect is communicated.

The multi-modal sensor system that we are using in the current version of this platform consists of a Pressure Mouse, a Wireless BlueTooth skin conductivity sensor, a Posture Analysis Seat, a Facial Action Unit analysis using the Blue Eyes camera system, and Head Tracking. This system expands upon the earlier work of Kapoor, Mota, and Picard [22], which used only facial and postural information. Through the combination of all these modalities we not only seek to provide the agent with a better understanding of the affect and interactions of the user but also to determine the contribution of each of the sensors to the modeling of affect [15].

4.1 Pressure mouse

The Pressure-Mouse has eight force-sensitive-resisters that capture the amount of pressure that is put on the mouse throughout the activity [23]. Users who have been administered a frustration inducing online application form have been shown to produce increasing amounts of pressure related to their level of frustration [24].

4.1 Wireless BlueTooth skin conductivity

In collaboration with Gary McDarby at Media Lab Europe and Carson Reynolds at the MIT Media Lab, we have developed a wireless version of our earlier "glove" that senses skin conductivity. While the skin conductivity signal does not tell you anything about valence – how positive or negative the affective state is – it does tend to be correlated with arousal – how activated the person is. High levels of arousal tend to accompany significant and attention-getting events [25].

4.2 Posture analysis seat

The Posture Analysis Seat utilizes the TekScan sensor pad system developed for medical and automotive applications [26]. The system uses pattern recognition techniques, while watching natural behaviors, to "learn" what behaviors tend to accompany states such as interest and boredom [27]. The system thus detects the surface-level behaviors (postures) and their mappings during a learning situation in an unobtrusive manner so that they don't interfere with the natural learning process. Through the chair, we have demonstrated significant detection of nine static postures and four temporal patterns associated with learner interest.

4.2 Blue Eyes camera system

Kapoor and Picard [28] have been developing automatic tools for computer vision and machine learning that are capable of detecting facial movements and head gestures used as conversational cues and communications of emotion. The system currently tracks upper facial features, eyes and eyebrows, their motion and action (eyes squinting or widening, eyebrows raised, head nod and shake). These techniques are being extended to include lower facial features, cheeks and mouth, which express smile, fidgets, and tension. The data logging includes full frame synchronized capture of the Blue-Eyes [29] camera images at 20 hz giving the opportunity to code for additional facial action units as they are identified.

4.3 Head Tracking

The Head Tracking System [30,31] is built upon the Small Vision System developed by SRI International and the MEGA-DCS stereo camera [32]. This system also incorporates real-time head nod and head shake algorithm [33]. This system provides information on the intersection of the user's gaze and the screen plane. This plane can be shifted to various reference depths within the environment to ascertain the virtual object that a user is directing their head toward. This type of sensing helps to facilitate shared attention behaviors.

We have the wide screen plasma screen that provides greater spatial resolution between objects. This causes users to move their head, to a greater extent than they would on a smaller screen, to attend to different objects and points of interest. This facilitates the use of the head tracker.

4.4 Video capture

The video camera records the user and the onscreen activity. It is positioned so as to acquire both an image of the user and an image of the screen that is reflected in a mirror positioned behind the users head. This system was chosen so as not to miss any of the features of the user/character interaction and provide true (same image) synchronization. When the system is initialized a datagram signal is sent to start the DirectX video capture and the time is noted in the log.

4.5 Open architecture for new sensors

The sensor system is constructed as an open architecture that can be added upon simply by sending sensor output to the Server-System. The data will be logged for future offline analysis for its incorporation to the *inference engine* and the data can be passed on to the *behavior engine* and coded for direct interpretation and association with behavioral actions.

4.6 Inference Engine

The simplest inference engine for an empathetic agent would be to recognize movements that the character was capable of mirroring or mimicking, such as leaning toward the system, smiling, nodding, and so forth. These can be recognized with current pattern recognition tools that we have developed. At a more advanced level, we wish to discern states such as "is the person interested?" Here, we have found that coupling different channels leads to improvement over any one channel [15]. While most of the learning conducted by the inference engine is done offline, in parallel with algorithm development, our aim is to enable inference to occur in real time, and to eventually learn continuously online while interacting with people. In the case of a learning companion the inference engine will be trained to identify states such as frustration or distress, interest or boredom, and pleasure, among others. The inference engine generates the state information that along with the relevant sensor-data, is submitted to the behavior engine.

5. Character and environment

The animation system allows explicitly separate control of *(i)* actions and *(ii)* affect of the character as the character performs actions, in a rough analogy to verbs and adverbs. This architecture allows characters to perform the same actions with varying body language and affect. For the current project, this structure allows the same actions to be repeated with differing affects. Actions include walking, jumping, reaching, blinking, grabbing and looking at selected objects.

The actors can follow sequential scripts, which have access to variables that monitor the state of the simulation world. These scripts can be run in parallel and can contain control structures such as "if" conditionals, "while" loops, and both sequential and parallel calls to procedures.

The actors can respond to events in their environment, either by conditionally switching between scripts in response to changes in environment variable values, or else under explicit control of an external C++ program.

Actors also contain internal scalar variables or "knobs" that can be modified over time by the actor's script. These knobs include posture (stooped versus erect), knees more bent or unbent, thrusting the pelvis forward or back, thrusting the pelvis from side to side, limping, rate of eye blink, face coloration, sidling (side-stepping), energy level (snappy, quick movement versus slow movement), involvement (body follows gaze direction more or less), and jitteriness (for creating more or less nervous appearance).

It is very easy to combine these controls to create somewhat higher level controls. For example, thrusting the pelvis forward/back can be used together with thrusting the pelvis left/right, in order to create hip gyration for dancing. The same low level forward/back hip control can also be used to move the pelvis backward while bending the knees, to make the character sit down.

Face affect knobs constitute an integrated subsystem. We use the same facial affect controls that were previously described in (Perlin 1997) [34]. These include head turn, nod and tilt, eyebrows up/down, eye gaze direction, eyes open/closed, eyelid-centers up/down, mouth open/closed, mouth corners up/down, mouth narrow/wide, sneering. That subsystem has previously been effectively used to help children with Autism to learn how to accurately recognize human facial affect [35].

Each of these controls can either have the same value for the left and right sides of the face, or can be given left/right asymmetric values. The latter case is used for such gestures as winking and one-sided sneering or smiling (See Figure 3).

Rather than provide only a high level emotional API, we chose to provide lower level physical affect knobs, which the script writer can combine to create the appearance of higher level or more subtle emotional affects. In particular, by providing lower level controls, such as mouth corners raised, rather than "smile", we enable script writers to create the appearance of a very rich set of emotional states, including even self-contradictory emotional states. For example, a character's mouth can be smiling while his eyelids can convey sad or neutral affect.

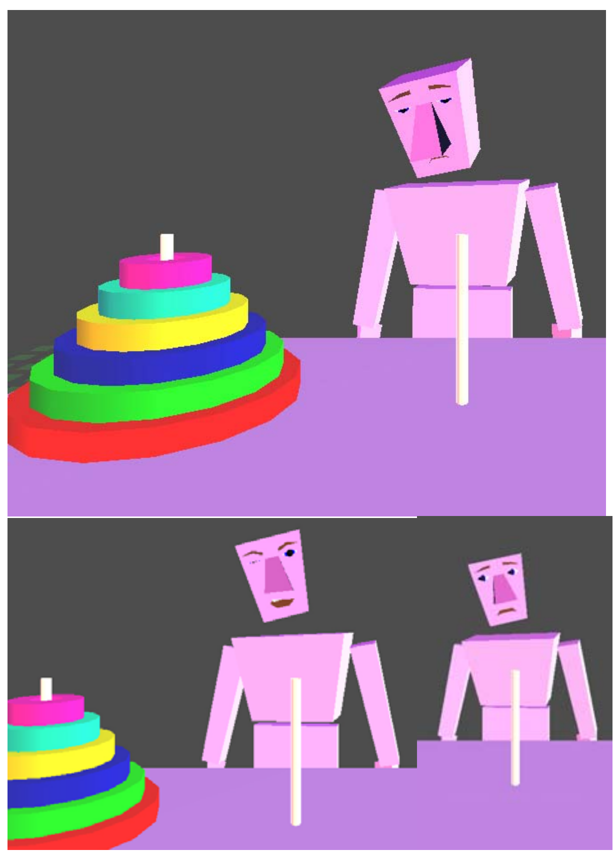

Figure 3. The agent is capable of a continuum of different expressions.

For this project, the architecture is used for "mirroring": features of the user's emotional state are inferred, and the actor is instructed to visually mirror aspects of that state. In this way, a virtual actor can appear to mirror users' emotional state without the virtual actor itself needing to have an extensive internal emotional model. Also, the "mirroring" can be conducted with natural appearing variation, so that it does not appear to be an exact duplication of what the user does.

Other knobs that control physiological appearance are also being used for this project. The system allows the programmer to control, at run-time, physical attributes such as height, girth and gender. In the current project these are being used to roughly match the physical characteristics of the user, in order to maximize the empathic effect of mirroring. We also expect these capabilities to prove useful in other contexts.

6. Research investigations

The platform that has been described is a general-purpose tool, which is currently adapted for exploration of affective interactions in learning situations where a synthetic character serves as a companion or coach to accompany a human learner. We also see potential uses in the exploration of non-verbal social-emotional communication that arises in areas as diverse as business negotiation and therapeutic care giving. Affect communication is believed to play a key role in job interviews, persuasive interactions, creative problem solving, behavior-change motivation, tutoring, and more. For example "thin slices" of video (30s or just 6s long) when rated for non-verbal behavior such as warmth, enthusiasm, and likeability has significant predictive value of perceived teacher effectiveness by college and high school students [36]. Likewise, Judges' belief about the guilt of a defendant is transmitted affectively in their brief, "unbiased", instructions to the juries and predicts jury outcome [37]. This range of findings shows that many aspects of relational studies may not require longitudinal study, that the formative elements of relationships can be very strong from the outset and that these elements are strongly influenced by nonverbal communication. A few examples are described below, with emphasis on their applications to the science of learning.

6.1. Facilitating relationships and learning experiences

We see several opportunities for this platform to be used in the development of learning companions, intelligent tutors, virtual peers, or groups of virtual friends to support learning, creativity, playful imagination, motivation, and to pursue the development of meta-cognitive skills that persist beyond interaction with the technology. Relationships have been shown to be effective in many learning situations: they help learners to develop responsibility, and increase the belief in children's ability for mastery; and caring relationships have also been shown to be predictors of performance [38,39,40,41]. Embodied conversational agents are capable of: developing trusting and beneficial relationships with humans; sharing combined physical and virtual space with children; and helping children develop literacy skills [18, 42]. The ability to recognize and respond to affective information in an empathic way plays a role in all of these relationships. It has been argued that being attuned to the child's emotional state through affective sensing will be important to the development of Intelligent Tutoring Systems and learning companions [43,44,45].

We are particularly interested in the nonverbal emotional signals sent by the agent, and how these are timed relative to those expressed by the learner. For example, if the agent subtly "mirrors" the leaning forward posture, facial expressions, and other expressive moves made by the learner, will this increase rapport and liking for the agent? (Note that we do not mean that the agent should perfectly imitate such expressions, which we think would be distracting.) Such an outcome might in turn facilitate increased desire by the learner to continue with a difficult task and possibly even a task perceived as boring, enabling them to gain valuable experience in persevering through difficulty. Perseverance makes it more likely the learner will achieve success in problem solving, which is often its own reward. Having "great learning experiences" is widely believed to facilitate intrinsic motivation toward learning [46]. Thus, we'd like to use this platform to help understand the role of affect communication in facilitating great learning experiences.

6.2. Can you catch enthusiasm from an agent?

Another line of investigation is not to test mirroring effects where the agent responds to the learner, but emotion contagion effects, which have the aim of getting the learner to respond to the agent. Many people have had the experience of becoming interested in a topic because of the "infectious" enthusiasm of a teacher or peer. We are curious if an agent can express emotion in a way that is infectious to its human companion – perhaps encouraging a desire in the learner to look more deeply into a topic than the learner is otherwise motivated to do. By carefully varying parameters that control how the agent expresses affect, either in response to the learner or proactively, we can use this platform to test various hypotheses about the role of visible and verbal affect communication in learning.

6.3. Enabling better responses to failure

Our main focus at this time is to see if we can improve the learners' response to failure through empathetic interaction. Failure is important to learning and instrumental to the development of multiple points of view required for deep understanding [47,48,49]. To facilitate deep understanding of a new concept -- to facilitate learning -- learners must have the opportunity to develop multiple and flexible perspectives. The process of becoming an expert involves failure, understanding failure, and the motivation to move onward. Meta-cognitive awareness and personal strategies can play a role in developing an individual's ability to persevere through failure. However, failure and repeated failure can also have a negative impact on motivation, affect, and learning. Therefore, learners' response to failure is very important to their continued learning. While most agents treat positive events with positive expressions and negative events with expressions such as disappointment or confusion, without regard for what the learner himself is expressing, our approach examines the expressions of the learner together with these other events that make up the interaction. For example, a traditional system that sees a learner making repeated errors would respond negatively; however, our system would also look to see what the user is expressing: Is she making mistakes because she's curious and exploring different approaches? Or is she making mistakes and showing increased frustration and perhaps distress? Our system will explore different ways to respond to these conditions, taking into account the affective state expressed by the learner. Additionally, by watching the learner's own response, the system can try to gauge how positive it should appear when the learner succeeds, celebrating big successes even longer if it can get the learner to share its smiles, while attenuating its positive expressions if the learner's affective expressions oppose those of the agent.

One hypothesis is that a learning companion with the ability to sense and respond empathetically to the child's affect, will have a greater impact toward learning than one that lacks this ability. Our plan is to test this hypothesis through experiments that use the Towers of Hanoi activity as the learning setting. The Towers of Hanoi is an engaging and challenging puzzle that has been the subject of considerable mathematical and psychological study [50]. A principle benefit of the choice of the Towers of Hanoi as a learning scenario is that it is recursive -- a procedure that includes itself and therefore is repeated for each successive operation. It therefore presents the important opportunity not only for failure and recovery but repeated failure and recovery. Repeated failure and recovery has been advocated as being fundamental to the development of deep understanding and multiple viewpoints [47]. In the Towers of Hanoi, persevering through repeated failure, with the assistance of a Learning Companion Architecture, may result in affective awareness, thinking strategies, and learning that significantly contributes to expertise during a very few sessions.

To incorporate the Towers of Hanoi into the system the task-state information is treated as an additional sensor input and a module is incorporated in the *behavior engine* that recognizes the significance of task-states and interactions, such as progression, regression, completion of a recursive cycle and completion of task. We are collecting data with 12-13 year olds as subjects, varying strategies of affective mirroring and responding, and testing the impact on outcome measures of interest, perseverance, caring, enjoyment, and performance. While these findings propose that it is quite possible to find a direct effect of interaction with a character during the interaction, what we would like to show is the persistence of that affect beyond the interaction with the character such that it effects a change in approach, meta-cognition, and behavior. The latter is a longer-term effort, but one in which the current platform plays a supporting role.

6.4. Cognitive-affective explorations

There are also many additional research questions that the new platform facilitates. For example, we have long been interested in the role that emotional awareness plays in persevering through difficult situations. When you become aware that you are frustrated, and recognize that frustration means a goal has been impeded and it is time to search for a new strategy, then your "frustration awareness" can be used productively to trigger a search for alternate means of reaching your goal. Recognizing that frustration is a common and natural part of learning can help a learner to not get discouraged so easily. Can an agent help a learner to become more aware of these states, and perhaps even model better ways to respond to them, inspiring the learner to do likewise? Agent expressions that subtly or overtly mirror or respond to a learner's affective state might be used to facilitate improved emotional awareness.

Not only can posture be used as a communication tool, but also it can have a direct effect on affect and on outcome measures. In "Stoop to Conquer: Guiding and Self – Regulatory Functions of Physical Posture after Success," Riskind finds that a congruency between posture and recent performance is not only an expression of that performance but is also in fact a beneficial strategy for future success. If an "appropriate" posture, e.g. slumping after failure, is taken, then it can be part of a beneficial self-regulatory process which minimizes helplessness, depression and motivational deficits [51].

In contrast, encouraging the learner to "buck up" or keep the "chin up" and sit up proud in the face of failure can have harmful effects. It would be interesting to see if an empathetic agent, designed with such findings in mind, could encourage slumping after failure, and sitting up proudly after success, and replicate and further illuminate the results of Riskind in this more controlled environment.

Isen and her colleagues have found evidence that mild positive affect improves negotiation processes and outcomes; promotes generosity and social responsibility; self-efficacy; motivation toward accomplishment; openness and flexible manipulation of new information. "Positive affect is a source of human strength… promoting thinking that is not only efficient, but also careful, open-minded and thorough [52]." It is important to realize that the staying power of negative affect tends to outweigh the more transient experience of positive affect. This is a phenomenon known as "negative asymmetry [53]." Unfortunately for the purposes of motivating learners this negative asymmetry means that negative affect experienced from failure will persist disproportionately to the positive affect experienced from success. Educators and innovators must try extra hard to create motivating learning environments which celebrate achievement and provide sustaining inquiry opportunity at times of failure. At the same time, negative affect has also been shown to play an important role in facilitating certain kinds of focused analytical thinking [54]. Learning technologies that monitor affect can examine many more of the effects of positive and negative affect, perhaps promoting positive affect particularly at times of great challenge, while not shying away from negative affect when it might be useful to induce.

7. Conclusion

A new platform for affective agent research has been developed. It integrates an array of affective sensors in a modular architecture that drives a system server and data logger, *inference engine*, *behavior engine* and *character system*. The *character system* includes dynamically scripted character attributes at multiple levels. This approach is particularly suited to affective expression. This platform will be used to explore several affective findings in the social, behavioral, and learning sciences. Preliminary tests show that multi-sensor logging and generation of affective response to children's interactions are possible in real-time. We expect to have conducted a number of experiments on affective interactions and to be able to present results at the workshop.

8. Acknowledgements

We acknowledge contributors to the system development: Ashish Kapoor, David Merrill, Rob Reilly, Barry Kort, Danielle Martin, Devin Neal, Louis-Philippe Morency, Carson Reynolds, Phil Davis, Karen Liu, Marc Strauss and Salene Mota. This material is based upon work supported in part by the MIT Media Lab's Things That Think consortium and in part by the National Science Foundation under Grant No. 0087768. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation or the official policies, either expressed or implied, of the sponsors or of the United States Government.

8. References

1. J. Bates, A.B. Loyall, and W.S. Reilly, An architecture for action, emotion, and social behavior; Proceedings of the Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World, S. Marino al Cimino, Italy, 1992.

2. T. Koda, and P. Maes, Agents with Faces: The Effect of Personification, Proceedings of HCI'96, 1996.

3. C. Elliott, J. Rickel, and J. Lester, Lifelike Pedagogical agents and affective computing: An exploratory synthesis, In M. Woodridge and M. Veloso, eds., Artificial Intelligence Today, Volume 1600 of *Lecture Notes in Computer Science*, Springer-Verlag. pp. 195-212, 1999.

4. L. Johnson, J. Rickel, and J. Lester, Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments, International Journal of Artificial Intelligence in Education,11, pp. 47-78, 2000.

5. J. Klein, Y. Moon, and R.W. Picard, This Computer Responds to User Frustration: Theory, Design, Results, and Implications, Interacting with Computers, 14, pp. 119-140, 2000.

6. T. Bickmore and R.W. Picard, Towards Caring Machines, Proceedings of Conference on Human Factors in Computing Systems (CHI), 2004.

7. P. Salovey and J. D. Mayer, Emotional intelligence. Imagination, Cognition and Personality, 9(3):pp. 185-211, 1990.

8. D. Goleman, *Emotional Intelligence*, Bantam Books, New York, 1995.

9. C. Breazeal, *Designing Sociable Robots,* MIT Press, 2002.

10. B. Reeves and C. Nass, *The Media Equation*. Cambridge University Press, Cambridge, 1996.

11. Y. Moon, Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers, Journal of Consumer Research, pp. 323-339, 2000.

12. D. Robinson and L. Smith-Lovin, Emotion Display as a Strategy for Identity Negotiation, Motivation and Emotion, Vol 23. No. 2. 1999.

13. K. Schneider and I. Josephs, The expressive and communicative functions of preschool children's smiles in an achievement-situation, J. Nonverbal Behavior 15: pp. 185-98, 1991.

14. P. Bull, *Body Movement and Interpersonal Communication*. John Wiley & Sons, New York, 1983.

15. A. Kapoor, R.W. Picard and Y. Ivanov, Probabilistic Combination of Multiple Modalities to Detect Interest, International Conference on Pattern Recognition, Cambridge, U.K; August 2004.

16. J. Lester, C. Callaway, J. Gregoire, G. Stelling, S. Towns, and L. Zettlemoyer, Animated Pedagogical Agents in Knowledge-Based Learning Environments, In *Smart Machines in Education*. K. Forbus and P. Feltovich (eds.), AAAI press, Menlo Park, CA, pp. 269-298, 2001.

17. NextUp.com website, TextAloudMP3, www.nextup.com, 2004.

18. T. Bickmore, "Relational Agents: Effecting Change through Human-Computer Relationships", MIT Ph.D. Thesis, February 2003.

19. J. Cassell, H. Vilhjálmsson, and T. Bickmore, BEAT: The Behavior Expression Animation Toolkit, SIGGRAPH '01, Los Angeles, CA. 2001.

20. C. Rich, C.L. Sidner, and N. Lesh, COLLAGEN: Applying Collaborative Discourse Theory to Human-Computer Interaction, AI Magazine, 2001.

21. J. Lester, S. G. Towns, and J. Fitzgerald, Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents, International Journal of Artificial Intelligence in Education, 10, pp. 278-291; 1999.

22. A. Kapoor, M. Selene and R.W. Picard, Towards a Learning Companion that Recognizes Affect, Proceedings from Emotional and Intelligent II: The Tangled Knot of Social Cognition, AAAI Fall Symposium, November, 2001.

23. C. Reynolds, The Sensing and Measurement of Frustration with Computers, MIT Masters Thesis, May, 1999.

24. J, Dennerlein, T. Becker, P. Johnson, C. Reynolds, and R.W. Picard, Frustrating Computers Users Increases Exposure to Physical Factors, Proceedings of the International Ergonomics Association, Seoul, Korea Aug 24-29, 2003.

25. W. Boucsein, Electrodermal Activity. *The Plenum Series in Behavioral Psychophysiology and Medicine*, Plenum Press, New York, 1992.

26. Tekscan, Tekscan Body Pressure Measurement System User's Manual, Tekscan Inc., South Boston, MA, USA. 1997.

27. S. Mota and R.W. Picard, Automated Posture Analysis for Detecting Learner's Interest Level; 1st IEEE Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, CVPR HCI, 2003.

28. A. Kapoor and R.W. Picard, Fully-Automatic Upper Facial Action Recognition, IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003), held in conjunction with ICCV 2003, Nice, France, October 2003.

29. A. Haro, I. Essa, and M. Flickner, Detecting and Tracking Eyes by Using their Physiological Properties, Proceedings of Conference on Computer Vision and Pattern Recognition, June, 2000.

30. L. P. Morency, P. Sundberg, and T. Darrell, Pose Estimation using 3D View-Based Eigenspaces, ICCV workshop on Analysis and Modeling of Face and Gesture, Nice, France, October 2003.

31. L. P. Morency, A. Rahimi, and T. Darrell, Adaptive View-based Appearance Model CVPR 2003, Wisconsin, June 2003.

32. MEGA-DCS Stereo Camera, Videre Design website, www.videredesign.com/sthmdcs.htm, 2004.

33. A. Kapoor and R.W. Picard, A Real-Time Head Nod and Shake Detector, Workshop on Perceptive User Interfaces, Orlando, FL, 15th-16th, November 2001.

34. K. Perlin, Layered Compositing of Facial Expression, ACM SIGGRAPH 97 Technical Sketch, 1997.

35. G. Brandt, Using Virtual Agents, Avatars and Worlds for Treatment and Healing, Medicine Meets Virtual Reality Conference, Jan, 2004, Long Beach, CA, 2004.

36. N. Ambady and R. Rosenthal, Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness, Journal of Personality and Social Psychology, 64, pp. 431-441, 1993.

37. P.D. Blanck, R. Rosenthal, A.J. Hart, and F. Bernieri, The measure of the judge: An empirically-based framework for exploring trial judges' behavior. Iowa Law Review, 75,pp. 653-684; 1990.

38. J. Tudge and D. Caruso, Cooperative Problem-Solving in the Classroom, US, Illinois, 198, EDO-PS-88-1, 1988.

39. P. Strain, (Ed.) *The Utilization of Classroom Peers as Behavior Change Agents*, New York: Plenum Press 1981.

40. M. Webb, Peer Helping Relationships in Urban Schools. ERIC Digest, US, 1987-12-00. ERID/CUE-D-37, New York, 1987.

41. K. Wentzel, Student Motivation in Middle School: The Role of Perceived Pedagogical Caring, Journal of Educational Psychology, 89, 3, pp. 411-419, 1997.

42. K. Ryokai, C. Vaucelle, and J. Cassell, "Virtual Peers as Partners in Storytelling and Literacy Learning", Journal of Computer Assisted Learning 19(2): pp. 195-208, 2003.

43. R.W. Picard, *Affective Computing,* MIT Press, Cambridge, MA, 1997.

44. A.B. Adcock and R. Van Eck, The development of an affective measure for use in pedagogical agent tutoring environments, Poster presented at The 15th Annual University of Memphis Student Research Forum, Memphis, TN, April, 2003.

45. C. Conati, Probabilistic Assessment of User's Emotions in Educational Games. Journal of Applied Artificial Intelligence, special issue on " Merging Cognition and Affect in HCI", vol. 16 (7-8), pp. 555-575, 2002.

46. M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*, New York: Harper & Row, 1990.

47. A. Kay, Computers, Networks and Education, Scientific American, September, 1991.

48. S. Papert, *Mindstorms: children, computers, and powerful ideas*, Basic Books, Inc. New York, NY, 1980.

49. M. Minsky, Why People Think Computers Can't, AI Magazine, vol. 3 no. 4, Fall 1982, Reprinted in Technology Review, Nov/Dec 1983, and in *The Computer Culture*, Donnelly (eds.) Associated University Presses, Cranbury NJ, 1985.

50. C. Lock, M.C. Welsh, C. Adams, and A. Kurtz, Tower of Hanoi: Influence of Strategy Instruction and Extended Practice on Performance and Knowledge, International Neuropsychological Society, Toronto, Ontario, Canada, February, 2002.

51. J. H. Riskind, They stoop to conquer: Guiding and self-regulatory functions of physical posture after success and failure, *Journal of Personality and Social Psychology*, 47, pp. 479-493. 1984.

52. A. Isen, A Role for Neuropsychology in Understanding the Facilitating Influence of Positive Affect on Social Behavior and Cognitive Processes, (Snyder, C.; Lopez, S. Eds.) *Handbook of Positive Psychology*, New York, NY, 2002.

53. L. Giuseppe and D. Brass, Exploring the Social Ledger: Negative Relationships and Negative Asymmetry in Social Networks in Organizations, Academy of Management Review, Special issue: Building Effective Networks, Academy of Management, Chicago, IL. April 4, 2003.

54. N. Schwarz, Situated Cognition and the Wisdom in Feelings, Cognitive Tuning; in *The Wisdom in Feeling, Psychological Processes in Emotional Intelligence*; Eds. P. Salovey, L. Barrett, Guilford Press, New York, 2002.