Panel on Affect and Emotion in the User Interface

Barbara Hayes-Roth, Chair
Stanford University
Gates Computer Science Building
Stanford, CA 94305 USA
+1 650-723-0506
bhr@ksl.stanford.edu

Gene Ball
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
+1 425-936-5653
geneb@microsoft.com

Christine Lisetti
Stanford University
Department of Computer Science and
Department of Psychology
Stanford, CA 94305 USA
+1 650-725-3046
lisetti@elrond.stanford.edu

Rosalind W. Picard
MIT Media Laboratory, E15-392
20 Ames Street
Cambridge, MA 02139 USA
+1 617-253-0611
picard@media.mit.edu

Andrew Stern
PF. Magic
501 2nd St., Suite 400
San Francisco, CA 94107 USA
+1 415-495-0400
AndrewS@pfmagic.com

INTRODUCTION

Intelligence. So much of our technology revolves around intelligence: technology in support of intellectual activities; the goal of engineering artificial intelligence; the need for intelligence in the user interface. And yet, so much of everyday life is really about affect and emotion: differences in performance under conditions that are supportive, threatening, or punishing; the challenges of conflict resolution and cooperation among heterogeneous groups of people; the implicit messages of body language and conversational style; the spirit-sustaining texture of our affective relationships with family and friends.

Over a century ago, William James ("Psychology," 1892) observed that people tend to put the "intellectual cart" before the "emotional horse":

"Our natural way of thinking about emotions is that perception of some fact excites emotion, and this gives rise to bodily expression. Common sense says, we lose our fortune, are sorry, and weep; we meet a bear, are frightened and run; we are insulted by a rival, are angry and strike. This sequence is incorrect. My theory, on the contrary, is that bodily changes follow perception of the exciting fact and our feeling of the changes IS the emotion. Thus, we feel sorry because we cry, afraid because we tremble, angry because we strike."

In other words, it is incorrect to suppose that cognition is the predominant operation, mediating affect as well as action. Instead, James exhorts us to recognize that affect is the predominant operation, mediating both cognition and action.

IUI 98 San Francisco CA USA

More recently, analysts from several different disciplines have turned their attention to affect and emotion and its place in the interstices between people and machines. In "The Media Equation," Cliff Nass and Byron Reeves document the many ways in which people's social response to technology mimics their social response to other human beings, where "social response" includes many dimensions of affect. In "Things that Make Us Smart," Donald Norman so ardently represents the emotional interests of human beings that he might just as well have called his book "Things that Make Us Feel Smart." In "Emotional Intelligence," Daniel Goleman argues that intelligence is not only intellectual, that people skills are just as important as the traditionally recognized cognitive skills, and by implication, that machines might also make similar good use of emotional intelligence. In "Hamlet on the Holodeck," Janet Murray suggests that believable affect and emotion will produce the "rounded characters" of the future for whom people will willingly suspend disbelief in order to immerse themselves in the characters' interactive comedies and dramas.

Several researchers have directly attacked the problem of creating synthetic creatures in which affect and emotion substantially determine or are expressively communicated in behavior (e.g., Clark Elliott, Joe Bates, Jim Lester, Elizabeth Andre, Bruce Blumberg, Michael Mauldin, Ken Perlin, Gene Ball, and Barbara Hayes-Roth). Other researchers are exploring new ways to sense and interpret the affective state of users (e.g., Christine Lisetti, Rosalind Picard). Finally, with toys like Bandei's Tamagotchi, Microsoft's Interactive Barney, Fujitsu's Fin Fin, and PF. Magic's Petz (Andrew Stern), affective creatures are finding their way into the commercial marketplace.

On this Panel, we will hear the perspectives of four illustrious pioneers in the understanding and creation of affect and emotion in the user interface. Our Panelists

outline their positions in the following four sections. The last section below gives a preview of our Panel Session Organization.

Emotional Interfaces. Gene Ball

In the coming decade, spoken interaction with computers will become increasingly common. Graphical user interfaces will be augmented by spoken interfaces that accept flexible descriptions of objects outside of the user's immediate workspace: "Find the notes that I took at the meeting with George just before I went to California last month." Descriptions of this sort are natural and powerful, but often result in ambiguities that require clarification: "Do you mean the group meeting or the meeting in George's office a few days earlier?" Therefore, spoken interactions won't usually consist of just isolated commands but will become conversational dialogues.

In many fundamental ways, people respond psychologically to interactive computers as if they were human. Nass and Reeves have demonstrated that strong social responses are invoked even if the computer doesn't use an explicitly anthropomorphic animated assistant. User interface designers should recognize that spoken conversation with a computer will generate social and emotional reactions in users, including expectations of appropriate emotional responses in the computer itself.

Explicit attention to the emotional aspects of computer interaction will be necessary in order to avoid degrading the user's experience by generating unnatural and disconcerting behaviors. For example, early text-to-speech systems generated completely monotonic speech, which conveyed a distinctly depressed (and depressing) emotional tone. Therefore I would argue that the initial goal for emotional interfaces should be to simulate appropriate emotional reactivity by demonstrating an awareness of the emotional content of an interaction. In order to approach this goal, interfaces need to:

- recognize and/or predict the emotional state of the user, and then
- synthesize and communicate an appropriate emotional response from the computer.

At Microsoft Research we are beginning to investigate the recognition and synthesis of various emotional cues, including:

- word choice and syntactic framing of utterances,
- speech pace, rhythm, and pitch contour,
- gesture, expression and body language, and
- the choice of actions to perform.

Over the last decade, user interface design has focused primarily on system capability and usability. The coming years will see greatly increased attention to the subjective user experience, including the aesthetic and emotional impact of computer use. My expectation is that the experience gained from modeling the emotional impact of spoken interfaces will also be used to inform the design (and possibly the dynamic behavior) of conventional graphic interfaces, in order to improve user satisfaction.

Acknowledging the Interface between Affect and Cognition. Christine L. Lisetti.

Because emotion generation is associated with three phenomena -- autonomic nervous system (ANS) arousal, expression, and subjective experience -- some of the latest progress in AI such as machine vision, speech recognition, and haptic processing have made it within our reach to observe computationally the arousal of emotional phenomena.

We are working on the development of an adaptive system which (1) observes the user's affective states via multiple modalities, and (2) acquires ``awareness" and interprets the most probable user's current affective state. The system is designed to observe physiological components associated with emotions. These components are to be identified and collected from observing the user via the three main sensory systems of machine perception *visual, kinesthetic, auditory,* and via *natural language processing*. By processing the observed features, the system can then recognize the most likely experienced emotional state corresponding to the sensory observations. This recognition enables the interface to change adaptively some appropriate aspects of its interaction with the user, using multiple modalities.

Interestingly, while facial expressions have been traditionally considered as the *result* of emotional arousal, more recent views, which emphasize the plasticity of the brain, associate facial actions as a *source* of arousal. This challenging view adds potential to developing interfaces which can acknowledge and induce various affective phenomena. We are working on automatic recognition of facial expressions both off-line, and in real-time while the user is interacting with the computer application. Relevant expressions and their interpretations may vary depending upon the chosen type of application. From a computer-interaction perspective, for example, it is interesting to work with expressions corresponding to *surprise, fear, confusion,* and *satisfaction,* and with user-specific expressions corresponding to some of the user's most frequent affective states experienced while interacting with the environment.

We are also interested in collecting the observations about the user's affective state into a meaningful and accessible data structure. Our approach is designed to encode observations about the user's state into a data structure with various emotion components. For example, it includes components such as *current facial expression, emotion valence*} and *emotion intensity* which can be set with the results of directly observing the user's state. Other components are derived from what is known about that the causal events associated a specific emotion. They describe the subjective cognitive experience components which are associated with the emotion, the beliefs, the goals, and their achievement.

As the field of Human-Computer Interaction evolves in its structure and spread, researchers realize that the experience of a person interacting with the computers is not limited

to its cognitive aspect, as explored in mainstream literature. The user's experience is also aesthetic, affective, and emotional. Since the recent integration of software into new technologies such as entertainment, and games, the need to concentrate on the affective dimensions of human response has become more obvious. Yet, even for the more conventional computer users in laboratories and offices, the non-cognitive aspect of the user's experience may have been undermined for lack of better knowledge about the primacy of emotional phenomena and its interface with cognition. Indeed, whenever people experience a piece of software, they have natural human responses such as beauty, satisfaction, frustration, or fun. There is not even a definite consensus in psychology about exactly where the separation between emotion and cognition occurs! From either perspective, however, the importance of affect in the human experience is considered salient.

As further progress is made in machine perception of human affect, applications of intelligent interfaces are likely to explode. It is expected, for example, that three to ten years from now, the price of cameras will have dropped considerably. This will make visual awareness of affect for intelligent interfaces a very affordable feature for most computers on the market. As applications will continue to spread throughout our everyday activities, involving users of increasingly wider age groups, monitoring of inexperienced users could be facilitated by the computer's awareness of their affective reactions. Visual awareness will eventually be coupled with auditory processing of speech intonation which will add some of the crucial contextual information often missing in human-computer interaction. Ideally, an interface agent could: (1) have a database of emotion states for each of the most commonly experienced emotions of a given user; (2) provide feedback to the user about his or her state; (3) record and remember the user's states during an interaction; (4) adapt its feedback dynamically using multi-modal devices (expression, posture, vocal inflection); (5) change the pace of a learning environment based upon the identified user's state (e.g. bored, overwhelmed, frustrated, satisfied); (6) guide the user to avoid undesirable cognitive-emotional paths; (7) instantiate its own motivational state; (8) initiate some appropriate adaptive action from that state. Much work is still needed to model affect in a computer systems, yet affect processing might very well be a determining factor for the future of the next generation of intelligent systems.

Computers that Recognize a User's Emotional Expressions. Rosalind W. Picard

I lead a research effort to develop computers that have the skills involved in emotional intelligence. A key part of this research is focused on giving computers the ability to recognize human emotions and to respond intelligently to them. For example, if a computer recognizes its user is confused then it might offer an alternative explanation, or possibly even connection to an expert human user. If it notices that several users get frustrated when accessing a similar sequence of events, it might not only offer alternative means of help to these people, but it might make a note of the details of the situation and report these to its designers. Designers would get information such as "When the computer was in this state, 160 users last week got frustrated," helping them to prioritize which aspects of the design most need evolution.

Currently, there are few efforts to recognize human expressions of emotion or emotional preference. Sometimes software agents employ menus asking for feedback, but this is an unnatural interruption when a person is involved in a task. In particular, if the emotion is a positive one, such as an expression of interest, it would be distracting to interrupt the user to inquire about this information. On the other hand, people communicate this information via subtle expressions all the time: facial, vocal, gestural, and so forth. Although people are not 100% reliable in their communication, the imperfect information is nonetheless very powerful information.

There is evidence that a felt emotional state modulates the entire body; hence, it can be expressed through many channels. We are developing a variety of means to access human emotion. Some of the devices we have prototyped include: wearable jewelry and shoes for sensing states such as frustration or stress, "expression glasses" that can sense expressions such as confusion or interest, and a "sentic mouse" that can discern slight variations in finger pressure that tend to be indicative of valenced information: liking or disliking. These novel interfaces are a few of many possible examples for ways to augment our interaction with computers, facilitating greater bandwidth communication without demanding any extra effort or training on the part of the user.

In five to ten years, I predict that there will be many interfaces capable of recognizing *deliberately expressed* human emotions. These systems will work best when the user freely communicates her emotional state, that is, when it is in her best interests to do so, and when she feels that the information is secure and she has something to gain from it. If a user does not want this information communicated, he should know how to inhibit its communication to the system. I also expect that there will be many applications that try to use this information to adapt to people, enabling much better learning of human preferences, and, ideally, vastly improved feelings on the part of users.

Virtual Petz as an Intelligent User Interface. Andrew Stern

At PF.Magic we are striving to create the most life-like, believable and expressive computer characters in the world. Over the past two years we have developed a series of Virtual Petz called Dogz and Catz. They are autonomous characters with real-time 3D animation and sound. With a mouse the user moves a hand-shaped cursor to directly touch, pet, and pick up the characters, as well as use toys and objects. Virtual Petz live on your PC computer desktop and grow up over time from

puppies or kittens into mature adults, and develop personalities based on attention, training,and interaction with their owners.

The fundamental premise of the Petz experience is to build an intimate relationship between the Petz and the user, their owner. Therefore the pet's primary motivation is to receive attention and affection. They feed off of this interaction. Without it they become lethargic, depressed, and if ignored long enough, they will run away. The interface supports this in three very important ways. First, users can touch and pet the characters, which is critical for intimacy and creating the illusion of life. Second, the toys and objects in the pet's environment have direct object-like interaction for both the user and the Petz. Petz have full access to the toy shelf, and if they really want something, they will just go and get it themselves, giving the Petz equal footing in their relationship with the user. Third, the Petz communicate their feelings through highly expressive animation, behavior and sound, and never with sliders or bar graphs. Users are not offered a "happiness value" or "hunger level" or a numeric score of any kind -- they must interpret their pet's feelings and thoughts naturally from body language, facial expression, and behavior.

Interface is one of the biggest issues for the future development of life-like computer characters. We can make the Petz very expressive and communicative towards the user, but we have too few ways to allow users to communicate and express themselves back. Communication is the key to any relationship, even virtual ones. Once PC computer interfaces offer true language capability, such as good speech recognition and/or gesture recognition, then computer characters can take a quantum leap forward.

Preview of the Panel Discussion

Our 90-minute panel will be organized roughly as follows:

The Chair will open the panel by introducing our topic, Affect and Emotion in the User Interface. Then she will introduce the Panelists, giving each one an opportunity to present a brief position statement. (30 minutes)

The Chair will then moderate a series of round-robin mini-discussions, in which she presents an intentionally provocative proposition and gives each panelist an opportunity to respond. (25 minutes)

Although the list of propositions is not yet final, illustrative candidates include:

1. Non-verbal affective channels will enable the creation of new and improved functional relationships between people and machines.

2. People will feel exposed under the scrutiny of machines that monitor and interpret their emotional states.

3. People will feel emotionally manipulated by machines that feign empathetic responses.

4. Affective bonds will be the basis for deep and lasting relationships between people and machines.

5. People will tolerate and appreciate empathetic machines in a greater number and variety of roles than are currently filled by their affectless predecessors.

Next, the Chair will open the Panel to questions and comments from the audience. (20 minutes)

The panel will close with final statements from each of the Panelists and the Chair. (15 minutes)