M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 515
Appeared: Actes Proceedings IMAGINA, March 1998, pp. 153-165, Monaco

Toward Agents that Recognize Emotion

Rosalind W. Picard
MIT Media Laboratory; 20 Ames Street; Cambridge, MA 02139
Phone: 617-253-0611; Fax: 617-253-8874
picard@media.mit.edu; http://www.media.mit.edu/~picard

Abstract

It is now easy to find examples of interactive software agents and animated creatures that have the ability to *express* emotion; this paper describes research for giving them the ability to *recognize* emotion. The ability to recognize a person's emotions is a key aspect of human "emotional intelligence," which has been described by a number of scientists as being more important to success in life than are the traditional forms of mathematical and verbal intelligence. This paper describes research underway in emotion recognition at the MIT Media Lab, especially research involving new *wearable* interfaces.

1 Introduction

People often laugh or express delight at something presented by a computer—a funny animation, a virtual pet, a piece of humor mail—even though computers, to date, have been unaware of these human reactions. It is perhaps even more frequent to see a person expressing frustration or irritation at a computer, especially when they feel that the system is hindering their goals and progress by failing to do what they need done. There is a recent Associated Press report of a man who got so upset at his computer that he shot it with a gun several times, revealing that human anger can be just as powerfully expressed toward a machine as it can be toward another person. Humans often express emotions in front of computers, and interact with them in a natural and social way [1].

To date, computers, and their agents, animated creatures, and other forms of potentially adaptive software, have been affect-impaired. Although many systems can express emotion, most do not have any internal mechanisms of emotion or the ability to recognize emotion in others. Although there has been a lot of talk and effort to make computers "friendly" and "intelligent," little has been done in terms of giving them the skills required for affective interaction. Existing attempts focus primarily on making animated agents and other characters expressive, via faces, gestures, posture, and other behaviors. Such work is an important part of developing more affective interactions, and for making characters more entertaining, more engaging, or more believable. However, expression is only one part of affective interaction.

Affective interaction can have maximal impact when emotion recognition and expression is available to all parties, human and computational. Not only would the human recognize the agent's emotions, but the agent could potentially recognize the human's. In healthy human-human interaction, this is the typical situation, where all parties can recognize and express emotions (even without necessarily "having emotions.") If one party cannot recognize or understand emotions, as is typically the case when communicating with an autistic person, then the interaction is impaired.

Once an agent has recognized the emotions of a person with whom it is interacting, what should it do with this information? The answer to this question depends on many things: the situation in which the emotions arise, the relationship of the computer to that person, knowledge about the person's goals, standards, and preferences, and many other factors. Although these issues are outside the scope of this paper, they are very important. We are engaged in research to investigate how computers should sensitively and appropriately respond to human emotions. This paper focuses on progress with the first aspect: emotion recognition.

1.1 Example: Tutoring agent

Let's consider one example—a learning tutor—where some simple emotion recognition, coupled with a small repertoire of possible responses, can lead to an engaging and productive interaction between a person and an agent.

Imagine you are seated with your computer tutor, and suppose that it not only reads your gestural input, musical timing and phrasing, but that it can also read your emotional state. In other words, it not only interprets your musical expression, but also your facial expression and perhaps other physical changes corresponding to your emotional feelings such as heart-rate, breathing, blood-pressure, muscular tightness, and posture. Assume it could have the ability to distinguish the three emotions that some scientists have argued we all appear to have at birth: distress, interest, and pleasure [2].

Given affect recognition, the computer tutor might gauge if it is maintaining your interest during the lesson, before you quit out of frustration and it is too late for it to try something different. "Am I holding your interest?" it would consider. In the affirmative, it might nudge you with more challenging exercises. If, however, it detects you are frustrated and making lots of errors, then it might slow things down and proffer encouraging feedback. Detecting user distress, without the user making mechanical playing errors, might signal to the computer the performance of a moving requiem, or the presence of a sticky piano key, or the need to ask the user afterward for more information. In any event, the system should not always just try to make the user happy. Nor should it simply make the lesson easier if the user is upset. There are intelligent responses which, if given information about what the user is experiencing, can improve the pupil's learning experience. Access to the user's affective expression is a critical aspect of formulating an intelligent response.

The piano tutor scenario holds also for non-musical learning tasks – learning a software package, a new game, a foreign language, and more. The topic can vary, but the problem is the same: how should the computer adapt the pace and presentation to the user? How can it know when to provide encouraging feedback or to offer assistance? Certainly, the user should have the option to ask for this at any time; however, it has also been demonstrated that systems which proactively offer suggestions can provide a better learning experience.[1] The tutor probably should not interrupt a user who is doing well, but it might offer help to one who has been getting increasingly frustrated. Human teachers know that a student's affective response provides important cues for discerning how to help the student.

1.2 Affect and Intelligence

The ability to recognize emotion is one of the hallmarks of intelligence in people [4] and has been argued, together with several other affective abilities, to be even more important than IQ in determining success in life [5]. However, the role of emotion in machine intelligence has been largely ignored by researchers in artificial intelligence, who have focused instead on logical rule-based reasoning, which tends to be brittle and limited to well-defined problems. This focus can be expected to shift given new neurological findings about the critical role of human emotion for even rational decision making. Such findings have shown that people who, because of certain kinds of brain damage, essentially do not have enough emotions, do not behave more rationally or intelligently, but actually are severely impaired when it comes to ordinary day-to-day decision making [6].

I should emphasize that although I do not think computers will achieve true intelligence without mimicking almost all the mechanisms of human emotion, I do not think that all computers will need a full set of emotional abilities, just like all animals do not need the same set of emotions humans need. Certain computers will need more emotions than others. In particular, computational agents that attempt to function in human roles, assist humans in complex decision-making, manage large numbers of complex tasks with limited resources in an unpredictable environment, and interact intelligently with humans will need more sophisticated emotional abilities than will an agent that is dispatched to simply find an automobile at its lowest price. In a recent book, I give several illustrations for why we might want to give agents actual internal mechanisms of emotion that bias their own behaviors and cognitive functions [7]. This paper focuses on giving agents the subset of emotional skills for recognizing emotions in people.

Intelligent interaction involves the skills associated with what has come to be called "emotional intelligence," including the ability to recognize the emotions of others, to effectively communicate emotions, and to deftly manage and utilize emotions in social and motivational situations [8]. For example, sometimes *what* was said, is not as important as *how* it was said, and the intelligent person can figure this out. It is my hypothesis that many of the important skills involved in affect recognition and communication can be accomplished by a computer that does not have emotions, but does have the ability to recognize and communicate them. Although there are many reasons we may want to give agents actual mechanisms of "having" emotion, there are also many affective skills that are useful without actually having internal emotional states and their corresponding motivations.

Consider how an agent could behave more intelligently if given the ability to recognize valenced affective cues such as approval or disapproval. Valenced information—either positive or negative—is usually subtly expressed, but carries immense weight in human interaction. When the assistant does something you mildly dislike, you may not verbally elaborate about the situation, but you might subtly frown. If he is intelligent, he notices the frown, figures out what behavior you are frowning about, and tries to do things differently the next time.

Positive and negative reinforcement is important to systems that try to learn and adapt; however, it has been hard for humans to communicate this feedback to computers in a way that is natural because computers have not recognized emotional expression. Most software agents that try to adapt to people watch only their direct input. They observe human preferences from frequency of clicks on certain items, time spent reading certain kinds of articles, lists of keywords, and responses to online questionnaires. All of these are valid and useful, but none of them capture the indirect and natural forms of bodily expression that modulate so much of human communication. The computer records what you typed, but not if you typed happily or angrily. It records what you clicked, but not if you clicked with delight or with irritation. And what if there is no menu or keyword readily available to communicate how you feel about what the computer did?

In contrast, people learn about your affective state and preferences by watching your facial expressions, listening to your intonation, and observing your overall behavior and affective cues. Did she lean toward that item with interest, or scowl at that behavior? Did he furrow his brow in confusion, and did his hands tremble or become clammy with anxiety? This is not to say that affective cues are never articulated, but only that it is natural to communicate affect subtly, as it continuously modulates behavior.

Consider a final example based on a common debate about agents and emotions: should agents have faces and expressions or not? Suppose an agent with a face is irritating to its user. If the computer could see that its user was irritated, it might try reducing the agent's expressiveness, and if that didn't alleviate some of the irritation, it would offer to the user to remove this feature. If on the other hand, the computer saw that its user really enjoyed the agent's presence, then it would continue to be expressive, and might, over time, try expanding its repertoire of expressive interaction. The basic idea is that the computer, or agent, would make an effort to adapt to the user's expressed preferences, instead of always requiring the user to

[1] An example study that demonstrated this was conducted by Ted Selker, with his "COACH" system for teaching the computer language LISP [3].

directly manipulate system parameters.

2 Emotion Expression Recognition

Can a computer be taught to recognize our emotions? Emotion recognition is very difficult for humans, and some researchers have argued that emotion is primarily cognitive—just another kind of thought—which would imply that it is impossible to truly recognize emotions, since it is impossible to recognize thoughts. Increasingly, however, scientists have shown that emotion involves a combination of bodily and mental events, so that it is both physical and cognitive. An internal emotional state may be invisible to outsiders, but often this state is accompanied by forms of expression that are observable. Your felt emotion is expressed through a mixture of verbal and nonverbal channels, such as vocal intonation, facial expression, posture, behavior, skin color and temperature, and other physiological changes. Such forms of emotion expression can potentially be recognized by computers.

Emotion recognition is thwarted by factors such as the fact that emotion is ill-defined. After decades of discussion, psychologists still argue about a definition for emotion, and nearly a hundred definitions have been collected and analyzed for common themes [9]. In this paper I do not propose another definition, but assume the most common folk usage of the term emotion, which includes states such as anger, fear, joy, distress, interest, confusion, and other commonly experienced kinds of feelings, but does not include feelings that are simply sensory, like the feeling of a pinprick or of a greasy surface. The latter sensations might *elicit* emotions such as disliking or disgust, but psychologists usually distinguish these simple sensations from emotions.

2.1 Sensing Expressive Signals

Expressions are communicated by many modalities: face, voice, posture, gesture, muscle tension, skin temperature, and more. The most work with computers recognizing affect has been done with respect to facial expression recognition, which is also the easiest of the kinds of expression for people to control. Many people can "put on a poker face" and suppress the display of feelings. Cameras, microphones, bio-sensors, and a variety of tools from signal processing and pattern recognition can be used together to try to recognize a person's emotions. Additionally, reasoning about the situation that a person is in can assist in recognition of emotions. Although a person may not look sad, if she just got some bad news about something that meant a lot to her, then you can reason that she might feel sad. It is likely that the most robust emotion expression recognition will use a combination of modalities, from low-level signal processing of vocal, facial, and other bodily signals, to high-level reasoning about situations. We are interested in all of these ways of understanding human emotion, including not only direct expression and situation analysis, but also human behavior. The work described below focuses on a subset of these modalities inspired by what is possible with new wearable computing technology.

One of the distinguishing features of wearable computers, as opposed to merely portable computers, is that they can be in physical contact with you in a long-term intimate way. A wearable may not just hang on your belt,

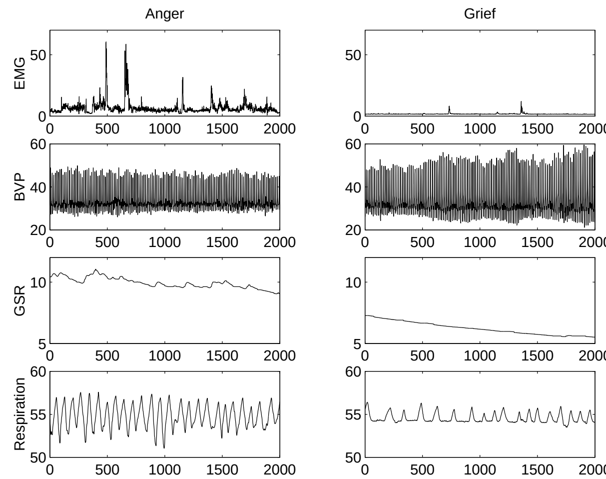

Figure 1: Examples of physiological signals measured from an actress while she consciously expressed anger (left) and grief (right). From top to bottom: electromyogram (microvolts), blood volume pulse (percent reflectance), galvanic skin conductivity (microSiemens), and respiration (percent maximum expansion). All of these signals can be gathered from sensors on the surface of the skin, without any pain or discomfort to the person. These signals were sampled at 20 Hz, using a ProComp system from Thought Technology Ltd. Each box shows 100 seconds of response.

but it may also reside in your shoes, hat, gloves, jewelry, or other clothing, providing a variety of kinds of physical contact beyond the traditional kind of contact, viz., fingertips touching the keyboard and mouse. In particular, when equipped with special sensors and tools from signal processing and pattern recognition, a wearable computer can potentially learn to recognize physical and physiological patterns–especially those that correspond to affective states–such as when you are fearful, stressed, or happily engaged in a task.

An "affective wearable" is a wearable system equipped with sensors and tools which enables recognition of its wearer's affective patterns [10]. Affective wearables have the potential not only to contribute to affective interaction with computational agents and other adaptive software, but also to help with advances in emotion theory and user modeling. One of the big problems in modeling people's emotions is determining what physiological and behavioral patterns accompany each emotion. Wearables help in data collection for a variety of situations.

Almost all studies trying to determine which responses occur with which emotions have been done on large numbers of people over a short amount of time, instead of trying to get to know one person over an extended period of time, in many situations. Behavior that might be consistent for one individual might be different for another, weakening the findings for a large group. Furthermore, most studies have involved artificially eliciting emotions in a lab setting, where there is good reason to believe that people might not feel the emotions in the same way as when they are naturally elicited. These problems have held back progress in emotion understanding. A wearable system addresses these problems by allowing an individual to wear sensors and gather patterns in natural situations, outside the laboratory, over a long period of time. This should aid in answering fundamental questions related to affect and human behavior.

In particular, a wearable computer can gather signals such as the four shown in Fig. 1: electromyogram (EMG), blood volume pressure (BVP), galvanic skin response (GSR), and respiration. The short segments shown in this figure illustrate very different responses obtained while an actress expressed two different negative emotions. Although clear differences can be seen in the signals for the two different emotions, we obtained data from the actress over 20 days, and sometimes found that the variations in the signals for the same emotion over different days were greater than the variations between the different emotions on the same day. In other words, the examples shown in Fig. 1 are some of the cleanest and most illustrative of the differences; in practice, it is very hard to build a system to recognize emotions from just these signals. The ultimate recognition accuracy that can be obtained is presently unknown.

Another issue is that people may differ in how they express emotions, not just as a function of context or situation, but also as a function of temperament and personality. Our current best recognition results are obtained when a system learns an individual person's pattern of expression over a period of time. The state of this research is like the early days of research in speech recognition, when the best results were obtained only on speaker-dependent systems. My expectation is that, just like performance improved dramatically over the last two decades for speaker-independent systems, performance will also improve dramatically for person-independent affect recognition. For now, however, much research is needed to discern which signals are most useful in each interaction, for each individual. This is true for facial expression, vocal inflection, bio-signals, behavioral signals, and so forth.

2.2 Example Interaction

In some early tests using a physiological monitoring system, we wired up a student who was playing the rather violent video game DOOM. We expected to see signs of distress during especially violent events in the game, such as when the student was brutally murdered with a machine gun. Over the several games where we observed the student, we saw many small responses in his bio-signals, such as when he "found the rocket launcher" or when he "was killed." However, the biggest response we found, throughout all of our recordings, was not when the student was being brutally murdered, or during any particular acts of violence. The biggest response of all occurred at a point where the software didn't work properly.

A short segment of three of his physiological signals can be seen in Fig. 2, which shows his GSR, BVP, and EMG over 5 minutes of time. His stress is initially signaled by the jaw-clenching peak in the EMG, which remains high during the minute and a half where the software controlled navigation keys failed to work as he expected. The point labeled "give up" is where he stopped the game and started over. After the game gets going, we see a constriction in his BVP, indicative of lowered blood flow to his extremities, and an increase in the GSR, indicating a state of higher arousal.

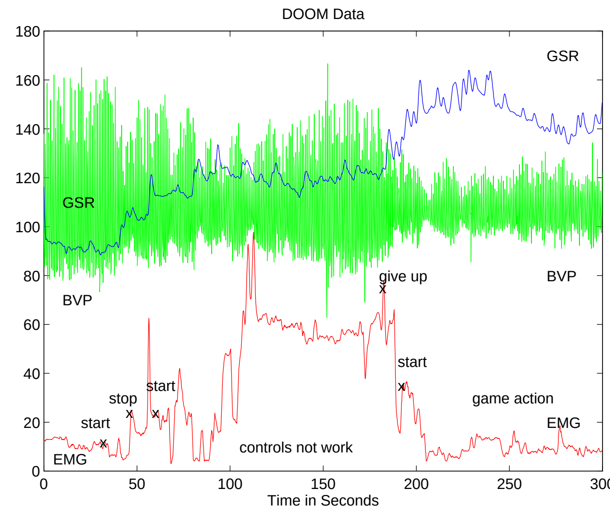

Figure 2: Three physiological signals measured from a student playing the game DOOM on the computer. The most significant response is shown in the center, where EMG peaked during loss of use of the controls.

Building tools to recognize stress would clearly be useful in a variety of interactive situations. Systems that proactively offer help might choose to adapt their timing based on signs of distress from a user. The tutoring system might modify its interaction with the pupil to allow just enough stress that a student learns how to overcome difficulty, but not so much that she gives up before she has met her goals. A character in an interactive game might change its behavior and offer some comic relief if it recognized that the human player was exhibiting a significant level of stress. It might also reward a player who exhibited less stress, especially if the human indicated interest in relaxing. Human factors experts evaluating new products could be aided by this technology to help gather data indicating precisely where subjects found interactions most stressful or annoying, helping pinpoint places where the product most needs improvement. For example, a copy machine might note what state it was in when people got most frustrated using it. Although the use of physiological and other affective signals may not replace the ubiquitous questionnaire, it can provide valuable information that a user may not easily be able to recall or articulate upon request.

2.3 Example Recognition Results

We have been running experiments to try to teach computers how to recognize human emotion. These experiments take two basic forms to date: (1) A person deliberately expresses emotions, using techniques such as visualization to increase the probability of simultaneously feeling the emotion being expressed; (2) The person is placed in a situation that is intended to induce an emotion. Clearly, the means by which an emotion is elicited can greatly influence the results. Hence, it is important to consider many different kinds of eliciting factors, including not only the stimulus for eliciting an emotion (watching a movie, listening to music, giving of a gift, impeding a goal, etc.) but also the environment in which it takes place (public setting, professor's lab, etc.)

Consider two situations we have investigated that illustrate the two kinds of experiments just mentioned. In the first situation, an actress, sitting in her office alone, deliberately expressed one of eight different emotions. She expressed all eight emotions in series, over the course of about half an hour. She repeated this once a day, for several weeks. The actress used a combination of techniques to try to arouse an actual experience within herself of each emotion. Not only did she use visualization, but also she was aided by the use of a "sentograph," a device designed by Manfred Clynes, together with a tape created by Clynes, that helps a person quickly experience a broad range of emotions [11]. The latter also assists a person in re-generating each emotion over a period of time, so that it builds in intensity over several minutes before the next emotion is begun. The eight emotions she expressed were: no emotion, anger, hate, grief, platonic love, joy, romantic love, and reverence, in that order, each day. While the actress went through the process of generating the eight emotions, she was wired to a system that measured the four physiological signals of EMG, BVP, GSR, and respiration, examples of which were shown above. Hence, 32 signals were gathered every day for twenty days. Note that this experiment is different from what is usually done by emotion theorists who gather data from groups of people over a short amount of time, instead of gathering data from an individual over a long period of time.

We have developed several approaches to analyzing this data and conducted numerous tests with these approaches. In the best cases, where the system works with a subset of the data such as three emotions, we are able to obtain recognition rates of up to 87%, which is significantly higher than the rate of 33%, which would be expected with random guessing [12]. Although the rates drop when more emotions are included, they do remain significantly higher than random guessing, indicating that although there is much room for improvement, these early results still show that there is important affective expression information in the signals we gathered.

A second situation in which we have run experiments aims to approximate a more natural human-computer interaction. In this case, a user is given a task, and a goal, and is told he will win a 100 dollar prize if he is the fastest and most accurate of our subjects in achieving the goal. We set up a "visual perception" task, consisting of many small simple puzzles, and asked the person to complete them as accurately and quickly as possible. Meanwhile we had rigged the mouse so that it would break in a controlled fashion, intending to frustrate the user who had to rely upon the mouse's performance to achieve his goal. We assumed that these events would cause an emotion of mild distress, one of the common kinds of frustration occurring during human-computer interactions.

During this experiment we gathered two signals: BVP and GSR. Our goal was to analyze these two signals during times of rest, and during time periods shortly after the mouse "broke," to see if they gave any indication of when the subject was "frustrated" by the mouse breaking. Although we do not claim that this situation reliably generated human frustration, or that frustration can generally

Figure 3: The shoe provides a convenient location for sensor placement. Here, a skin conductivity sensor is placed in the arch of the shoe and a pressure sensitive resistor is placed on the heel. Sensors look unobtrusive when worn. (Photograph courtesy of Fernando Padilla.)

be recognized by these two signals, we did find that patterns of the two signals gave significant discrimination of "non-frustrated" vs. "frustrated" for twenty of twenty-four users [13]. The average recognition rate was 75%, whereas chance guessing would be 50%. I should mention that it took a lot of work to obtain these results, as many of the methods of analysis we tried initially did not discriminate the two cases. The significant results we finally obtained used relatively sophisticated hidden Markov models, a doubly stochastic model often used in speech recognition systems. These models also worked best when trained for each user, individually. The patterns learned by the system for recognizing frustration in one person did not, in general, work for recognizing frustration in another person.

Both situations—the actress situation, and the case of users who we attempted to deliberately frustrate—approximate situations that might occur while a person is interacting with a computer, software agent, or computational characters in a virtual world. In all these cases, it is possible that something could go wrong in a way that causes a user to feel like the users did when the mouse stopped working. It is also possible, especially in games and role-playing situations, that a user might arouse within himself certain emotions in a similar means as those generated by the actress in our experiment. As a user interacts repeatedly with a system, the system has increased opportunity to learn how that person expresses different emotions and what situations tend to give rise to such emotions. Although the recognition of human emotions by computers is a new area, and far from being solved, we have already obtained results illustrating the promise of combining pattern recognition and signal processing for helping a computer recognize expressions.

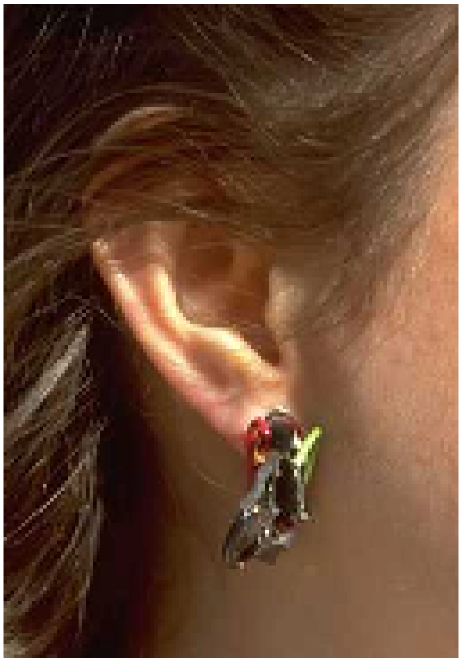

Figure 4: Analog BVP sensor incorporated into an earring. (Photograph courtesy of Frank Dabek.)

3 New Wearable Interfaces for Sensing Affect

Physiological signals are traditionally gathered with medical monitoring equipment, which tends to be non-stylish and awkward to wear. This section describes our ongoing effort to develop clothing, jewelry, and other wearable accessories as an opportunity to more comfortably sense affective information from people.

Figure 3 illustrates the use of a sandal for sensing GSR. We have found that the GSR signal on the foot is highly correlated with that on the hand [10], the place where GSR is usually measured. Sometimes it is desirable to move the sensors off the hands so that the hands can be free for other tasks, in which case the sandals offer a comfortable alternative.

One of the other signals commonly measured using a bulky device on the fingertip is the BVP, for which we have developed an alternate sensor that can be used on the earlobe. A prototype earring incorporating this sensor is shown in Fig. 4.

We are very interested in wireless sensing, and have begun to try to remove the wires connecting sensors to the computer, replacing them with a "Personal Area Network" (PAN) communication system such as that described by Tom Zimmerman [14]. We have augmented the PAN technology to work with multiple worn devices, allowing for many sensors to share approximately 2400 baud of bandwidth through a person's body. PAN uses short-wave FM to send signals through the body, obviating the need for wires to run from each sensor to the wearable computer. In this way the wearable computer can gather a constellation of patterns without a wearer having to tangle with a web of wires.

Figure 5: Glasses equipped with EMG sensors for detecting muscle movements used to recognize confusion. (Photograph courtesy of Fernando Padilla.)

There are remote learning situations where a computer is used to deliver a lecture or present new information, and where a videoteleconference system delivers a picture of an audience to a lecturer, and the lecturer cannot make out the facial expressions of the audience. This makes it difficult for the lecturer to gauge if the audience is interested, confused, bored, or ready to fall asleep. If the lecturer had this information, she could better adjust her presentation for a more effective interaction. If each member of the audience was willing to wear a networked piece of jewelry and special eyeglass frames that sensed their confusion, interest, or arousal, then this information might be integrated and presented to the lecturer in a way that nicely supplemented the tiny low-resolution video image of the remote audience.

We have prototyped a pair of eyeglasses with EMG sensors (Fig. 5) that can sense changes in the facial muscles involved in expressions of confusion. Plans are underway to build analyzers and integrators for the data they collect, to enable better communication of affective information through computers. Although in the long run it might be less burden on the user if this information were gathered from non-contact field sensing or video cameras, the glasses provide an immediate and relatively simple solution compared to these alternatives, and provide an advantage of privacy. The glasses can optionally cover up the user's expression and only reveal it anonymously. Anonymous communication can be desirable for timid students or for situations where people do not wish to reveal that they are confused. Glasses, as opposed to video cameras, provide a familiar and comfortable paradigm when it comes to a person feeling "in control" of the sensing technology: it is easy to take one's glasses off, or to physically disconnect the sensor from the glasses; it is not necessarily easy to figure out how to turn off a surveillance camera in the environment.

There are dozens of applications of affective computing in addition to the ones above [7]. For example, emotions are known to provide a keen index into human memory; therefore, a computer that pays attention to your affective state will be better at understanding what you are likely to recall on your own, and what is of interest to you. This is potentially very useful in helping people deal with information overload, an important application area for software agents. For example, instead of a system recording *everything* you hear, see, or click on, the system might learn to record (or play back) just those places where you were interested. Or, it might play back just those places in a lecture that you missed, perhaps because your mind wandered or you were bored. Augmenting a system like Steve Mann's WearCam [15] with affective sensing and pattern recognition could help it learn when to "remember" the video it collects, as opposed to always relying on the user to tell it what to remember or forget. Of course the user can still directly manipulate the system if that is desired; that function does not go away. The goal is simply to begin to automate those functions that the user typically applies, especially when they are predictable with affective information.

Suppose for example that you let the WearCam roll in a continuously learning mode while playing with a cute little child. It might notice that you always save the shots when the child makes you laugh, or smile. By detecting these events, it could become smarter about automatically saving these kinds of photos in the future. Moreover, by labeling the photos with these affective events, you can later ask the system to retrieve data by its affective qualities, "Computer, please show us the funny images." If the wearable learns continuously, by watching what the wearer chooses, it should help reduce some of the users workload and enable the wearer to offload repetitive tasks.

We have built a prototype of an affective WearCam that includes a small camera worn as a pendant around the wearer's neck, together with skin conductivity sensors and pattern recognition software running in a wearable computer. The camera continuously records and buffers images in a rotating buffer, deleting the oldest images as the buffer fills. Simultaneously, the system uses small electrodes to sense skin conductivity in the wearer's skin, either across two fingers or across the arch of the foot. Pattern recognition software has been trained to recognize the wearer's "startle response," a skin conductivity pattern that occurs when the wearer feels startled by a surprising event. Unlike many affective signals, the human startle response is fairly robust and easy to detect. With a matched filter and threshold detector, the startle pattern in the wearer's skin conductance signal is detected in real time. The skin conductivity response occurs with a typical latency of three seconds after the startling event. When the pattern is detected, the images leading up to the startle event are extracted from the buffer. The buffer can be set to hold arbitrary amounts of imagery, typically in the range of 5 seconds to 3 minutes of data. When the startle is detected, the images extracted from the buffer can then be saved into a more permanent memory for your later perusal, or automatically sent back to a remote location to be analyzed by a "safety net" [15], community of friends or family with whom you felt secure, to see if the event warranted any action on your behalf.

Applications of affective wearables extend to other forms of information management beyond image and video. An intelligent web browser responding to the wearer's degree of interest could elaborate on objects or topics that the wearer found interesting, until it detected the interest fading. An affective assistant agent could intelligently filter your e-mail or schedule, taking into account your emotional state or degree of activity. The agent might also be given the ability to recognize forms of affective expression in text, perhaps highlighting in red certain lines you typed in an email that might be read as "angry" by a recipient. "Are you sure you might not want to re-word this potentially angry-sounding mail?" it might ask you before it forwarded the mail.

Agents involved in preference selection, such as musical preference [16] [17], might not only take into account your explicitly stated musical tastes, but might also learn how your selection of music depends upon your "mood." Music is perhaps the most popular and socially-accepted form of mood manipulation. Although it is usually impossible to predict exactly which piece of music somebody would most like to hear, it is often not hard to pick what *type* of music they would prefer—a light piano sonata, an upbeat jazz improvisation, a soothing ballad—depending on what mood they are in. As wearable computers gain in their capacity to store and play music, to sense the wearer's mood, and to analyze feedback from the listener, they have the opportunity to learn patterns between the wearer's mood, environment, and musical preferences.

4 Conclusions

It is becoming important for agents and other computational creatures to develop certain emotional skills, especially the ability to recognize human emotions. This paper has highlighted recent research in emotion recognition based on signals that can be sensed from wearable computers. New designs for sensing physiological signals of expression have been constructed in jewelry, shoes, eyeglasses, and other wearable devices. Prototypes have been developed for gathering expressive signals, and work is underway to develop better algorithms for analyzing the data using tools from signal processing and pattern recognition. Once a computer learns to recognize various affective patterns for an individual, this information can be included in its decision-making for how to better adapt its behavior in service to a user. These are steps toward the ultimate goal of making human-computer interaction more human-centered, enabling computational agents and other forms of technology to better adapt to people and their expressed preferences.

Acknowledgments

I am grateful to the students and colleagues who assisted with the creation and implementation of the ideas in this paper. Jennifer Healey oversaw the physiological data gathering for all the experiments described, developed the prototype of the first affective wearable, and coordinated the development of the affective jewelry, shoes, and StartleCam. Raul Fernandez, Jennifer Healey, Tom Minka, and Elias Vyzas collaborated in the development of signal processing, pattern recognition, and learning algorithms for teaching computers to recognize emotions. Nicholas Negroponte suggested putting sensors in eyeglasses. Jocelyn Riseberg proposed the confusion glasses and developed the prototype shown here. Grant Gould developed the BVP earring and PAN protocol for multiple communicating devices. Matt Norwood, Katy Riley, and Rob Gruhl helped run experiments. Fernando Padilla assisted with the development of the StartleCam. This work has been sponsored in part by BT, P.L.C., by NEC and by Hewlett Packard Labs.

References

[1] B. Reeves and C. Nass, *The Media Equation*. Cambridge University Press; Center for the Study of Language and Information, 1996.

[2] M. Lewis, "Self-conscious emotions," *American Scientist*, vol. 83, pp. 68–78, Jan.-Feb. 1995.

[3] T. Selker, "Coach: A teaching agent that learns," *Communications of the ACM*, vol. 37, pp. 92–99, July 1994.

[4] H. Gardner, *Frames of Mind*. New York: BasicBooks, 1983.

[5] D. Goleman, *Emotional Intelligence*. New York: Bantam Books, 1995.

[6] A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Gosset/Putnam Press, 1994.

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

[8] P. Salovey and J. D. Mayer, "Emotional intelligence," *Imagination, Cognition and Personality*, vol. 9, no. 3, pp. 185–211, 1990.

[9] P. R. Kleinginna, Jr. and A. M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition," *Motivation and Emotion*, vol. 5, no. 4, pp. 345–379, 1981.

[10] R. W. Picard and J. Healey, "Affective wearables," in *Proceedings of the First International Symposium on Wearable Computers*, (Cambridge, MA), pp. 90–97, Oct. 1997.

[11] D. M. Clynes, *Sentics: The Touch of the Emotions*. Anchor Press/Doubleday, 1977.

[12] J. Healey and R. W. Picard, "Digital processing of affective signals," Media Laboratory, Perceptual Computing TR 444, MIT, Cambridge, MA, 1997. To appear ICASSP '98.

[13] R. Fernandez and R. W. Picard, "Signal processing for recognition of human frustration," Media Laboratory, Perceptual Computing TR 447, MIT, Cambridge, MA, 1997. Appears in Proceedings of ICASSP '98.

[14] T. Zimmerman, "Personal area networks (PAN): Near-field intra-body communication," *IBM Systems Journal*, vol. 35, pp. 609–618, 1996.

[15] S. Mann, "Wearable computing: A first step toward personal imaging," *Computer*, pp. 25–31, February 1997.

[16] M. E. Metral, "Motormouth: A generic engine for largescale, real-time automated collaborative filtering," Master's thesis, MIT Media Lab, Cambridge, MA, June 1995. Media Lab.

[17] Y. Z. Lashkari, "Feature guided automated collaborative filtering," Master's thesis, MIT Media Lab, Cambridge, MA, September 1995.