MIT Media Laboratory Affective Computing Technical Report No. 543.

To appear in: Proceedings from Emotional and Intelligent II: The Tangled Knot of Social Cognition, AAAI Fall Symposium, November 2001

Towards a Learning Companion that Recognizes Affect

Ashish Kapoor, Selene Mota, Rosalind W. Picard

MIT Media Lab
20 Ames Street
Cambridge, Massachusetts 02139
{ash, atenea, picard}@media.mit.edu

Abstract

This paper reports work in progress to build a Learning Companion, a computerized system sensitive to the affective aspects of learning, which facilitates the child's own efforts at learning. Learning related to science, math, engineering, and technology naturally involves failure and a host of associated affective responses. This article describes techniques and tools being developed to recognize affective states important in the interplay between emotions and learning.

Introduction

Learning the complex ideas involved in science, math, engineering, and technology and developing the cognitive reasoning skills that these areas demand often involves failure and a host of associated affective responses. These affective responses can range from feelings of interest and excitement to feelings of confusion and frustration. The student might quit if he is not able to recover from the 'feeling of getting stuck'. Expert teachers are very adept at recognizing and addressing the emotional state of learners and, based upon that observation, taking some action that positively impacts learning.

The goal of building a computerized *Learning Companion* is to facilitate the child's own efforts at learning. Our aim is to craft a companion that will help keep the child's exploration going, by occasionally prompting with questions or feedback, and by watching and responding to aspects of the affective state of the child—watching especially for signs of frustration and boredom that may precede quitting, for signs of curiosity or interest that tend to indicate active exploration, and for signs of enjoyment and mastery, which might indicate a successful learning experience. The *Learning Companion* is a player on the side of the student—a collaborator of sorts—to help him or her learn, and in so doing, learn how to learn better. It is a system that is sensitive to the learning trajectory of students.

Skilled humans can assess emotional signals with varying degrees of accuracy, and researchers are beginning to make progress giving computers similar abilities at recognizing affective expressions. Computer assessments of a learner's emotional/cognitive state can be used to influence how and when an automated companion chooses to intervene. For example, if a student appears to be engaged in the task and enjoying trying things, even if he or she is making mistakes, then it might not be good to interrupt. If, however, the student is showing signs of increasing frustration while making errors, then it might be appropriate to intervene. Affect recognition is thus a critical part of the process of determining how to best assist the learner. The challenge is to build computerized mechanisms that will accurately track and immediately recognize the affective state of a learner through the learning journey.

Kort et al. [2001] have developed a framework that models the complex interplay of emotions and learning. Inspired by that framework, we are trying to develop technology that is capable of recognizing some of the emotions involved in the learning process. The next section highlights the main ideas in the theoretical framework developed by Kort et al., namely the affective responses typically associated with science, math engineering and technology learning. The following sections describe tools and techniques we are developing to recognize some of these affective states – especially work focused on sensing gaze dynamics, facial expressions, and postural changes of the learner.

Guiding Theoretical Framework

Previous emotion theories have proposed that there are from two to twenty basic or prototype emotions (see for example, Plutchik [1980]; Leidelmeijer [1991]). The four most common emotions appearing on the many theorists' lists are fear, anger, sadness, and joy. Plutchik [1980] distinguished among eight basic emotions: fear, anger, sorrow, joy, disgust, acceptance, anticipation, and surprise. Ekman [1992] has focused on a set of seven emotions that have associated facial expressions and that show up in diverse cultures– fear, anger, sadness, happiness disgust, surprise, and contempt. However, none of the existing frameworks seem to address emotions commonly seen by teachers in learning experiences, some of which we have noted in Figure 1.

Axis	**-1. 0**	**-0. 5**	**0**		**+0. 5**	**+1. 0**
Anxiety-Confidence	Anxiety	Worry	Discomfort	Comfort	Hopeful	Confident
Boredom-Fascination	Ennui	Boredom	Indifference	Interest	Curiosity	Intrigue
Frustration-Euphoria	Frustration	Puzzlement	Confusion	Insight	Enlightenment	Ephipany
Dispirited-Encouraged	Dispirited	Disappointed	Dissatisfied	Satisfied	Thrilled	Enthusiastic
Terror-Enchantment	Terror	Dread	Apprehension	Calm	Anticipatory	Excited

Figure 1. – Some emotions relevant to learning (From Kort et al. [2001])

Figure 2 interweaves the emotion axes shown in Figure 1 with the cognitive dynamics of the learning process (Kort et al. [2001]). The horizontal axis is an Emotion Axis. It could be one of the specific axes from Figure 1, or it could symbolize the *n*-vector of all relevant emotion axes (thus allowing multi-dimensional combinations of emotions). The positive valence (more pleasurable) emotions are on the right; the negative valence (more unpleasant) emotions are on the left. The vertical axis is the Learning Axis, and symbolizes the construction of knowledge upward, and the discarding of misconceptions downward.

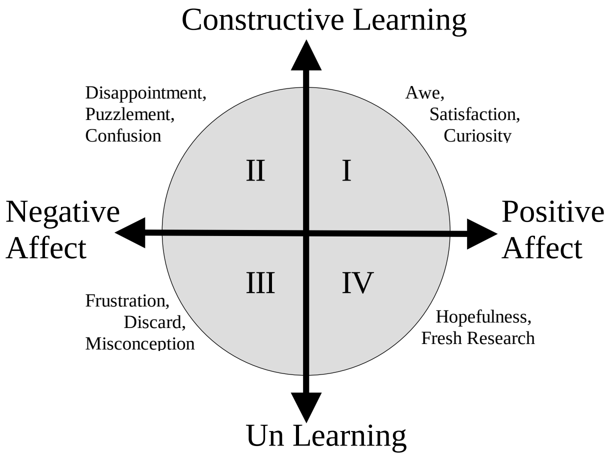
Constructive Learning
Disappointment, Puzzlement, Confusion
Awe, Satisfaction, Curiosity
II
I
Negative Affect
Positive Affect
III
IV
Frustration, Discard, Misconception
Hopefulness, Fresh Research
Un Learning

Figure 2. – Model relating phases of learning to emotions (From Kort et al. [2001]).

A typical learning experience involves a range of emotions, moving the student around the space as they learn (Kort et al. [2001]). Typically, movement would begin in quadrant I and proceed in a counter-clockwise direction. The student might be *curious/fascinated and/or interested* when he begins in quadrant I. He might be *puzzled /disappointed/ confused and/or motivated* to reduce confusion in quadrant II. In either case, the student is in the top half of the space if his focus is on constructing or testing knowledge. Movement happens in this space as learning proceeds. For example, when solving a puzzle in the software product *The Incredible Machine*, a student gets an idea how to implement a solution and then builds its simulation. When he runs the simulation and it fails, he sees that his idea has some part that doesn't work – that needs to be deconstructed. At this point it is not uncommon for the student to move down into the lower half of the diagram (quadrant III) where emotions may be *frustration/ hopelessness/ boredom* and the cognitive focus changes to eliminating some misconception. As he consolidates his knowledge—what works and what does not—with awareness of a sense of making progress, he may move to quadrant IV (*hopefulness/ excitement/ confident*).

Ideally, the *Learning Companion* should observe and try to understand the processes a learner experiences during all of these quadrants; however, this is currently beyond the capabilities of the technology. Our research is examining what aspects of the learner's affect can be reliably detected, with emphasis on those aspects that are most likely to be useful to a learning companion in determining when and how to intervene. Toward this aim, we are focusing initially on the sensing of states that may indicate a learner has drifted or is soon to drift from the task of learning (is *off-goal*). The aim of the initial intervention will be to help the learner return to being *on-goal*.

Affect Recognition in Learning

A lot of research had been done to develop ways and methods to infer affective states. Questionnaires have been used to infer affect from motivational and affective factors such as "curiosity, interest, tiredness, and boredom. (e.g., Matsubara and Nagamashi [1996]; de Vicente and Pain, [1999]). In a system by Klein et al. [1999], dialogue boxes with radio buttons were used for querying users about frustration. Although questionnaires can easily be administered, they have been criticized for being static and thus not able to recognize changes in affective states. Del Soldato [1994] had success in gathering information about the subject's affective state via face-to-face dialogue but studies of spoken assistance on demand (Olson and Wise [1987]) have revealed a serious flaw in assuming that young readers are willing and able to ask for help when they need it.

A more dynamic and objective approach for assessing changes in a person's affective state is via assessing sentic modulation (Picard [1997]), analyzing a person's emotional changes via sensors such as cameras, microphones, strain gauges, special wearable devices, and other. The computer assesses a constellation of such patterns and relates them to the user's affective state. Scheirer *et al.* [1999] have built *Expression Glasses* that discriminate between upward eyebrow activity indicative of expressions such as interest and downward eyebrow activity indicative of confusion or dissatisfaction. Healey [2000] has used physiological sensors to infer stress levels in drivers, and Picard et al. [2001] have reported 81% classification accuracy of eight emotional states of an individual over many days of data, based on four physiological signals. A survey of a variety of projects at the MIT Media Lab related to machine recognition of emotion is available (Picard [2001]).

The problem of automatic affect recognition is a hard one. Under restrictive assumptions in choosing from among about six different affective states, accuracy of from 60-80% is still state-of-the art in recognizing affect from speech. A lot of research has been directed at the problem of recognizing 5-7 classes of emotional expression on groups of 8-32 people from their Facial Expressions (e.g., Yacoob and Davis [1996]; Essa [1997]). Other recent studies indicate that combining multiple modalities, namely audio and video, for emotion recognition can give improved results ([DeSilva et al. [1997]; Huang et al. [1998]; Chen et al. [1998]). Most of the results are focused on deliberately expressed emotions posed in front of a camera (happy /sad /angry etc.), and not on those that arise in natural situations such as classroom learning.

Other facial expression analysis research has focused not so much on recognizing a few categories of "emotional expressions" but on recognizing specific facial actions—the fundamental muscle movements that comprise Paul Ekman's Facial Action Coding System, which can be combined to describe all facial expressions (Ekman, [1978]). These facial actions are essentially *facial phonemes*, which can be assembled to form facial expressions. Donato et al. [1999] compared several techniques, which included optical flow, principal component analysis, independent component analysis, local feature analysis and Gabor wavelet representation, to recognize eight single action units and four action unit combinations using image sequences that were manually aligned and free of head motions. Yingli Tian et al. [2001] have developed a system to recognize sixteen action units and any combination of those using facial feature tracking.

The techniques mentioned above were not aimed at reliably recognizing all the affective states in learning like *interest/ boredom/ confusion/ excitement.* The *Learning Companion* aims to sense truly felt emotional and cognitive aspects of the learning experience in an unobtrusive way. Cues like posture, gesture, eye gaze, facial expression etc. help expert teachers to recognize whether the learner is on task or off task. Rather than identifying exact emotional state continuously throughout the learning experience we aim to able to identify the surface level behaviors that suggest a transition from an *on-goal* state to *off-goal* state or vice versa.

Surface Level Behaviors to Infer Affect

Affective states in learning (like interest/ boredom/ confusion /excitement) are accompanied by different patterns of postures, gesture, eye-gaze and facial expressions. Rich et al. [1994] have defined symbolic postures that convey a specific meaning about the actions of a user sitting in an office which are: interested, bored, thinking, seated, relaxed, defensive, and confident. Leaning forward towards a computer screen might be a sign of attention (*on-task*) while slumping on the chair or fidgeting suggests frustration/ boredom (*off-task*).

The direction of eye gaze is an important signal to assess the focus of attention of the learner. In an *on-task* state the focus of attention is mainly toward the problem the student is working on, whereas in an *off-task state* the eye-gaze might wander off from it. The facial expressions and head nods are also good indicators of affective and motivational states. Approving head nods and facial actions like smile (AU 6+12), tightening of eyelids while concentrating (AU 7), eyes widening (AU 5) and raising of eyebrows (AU 1+2) suggest interest/ surprise/ excitement (on task), whereas head shakes, lowering of eyebrows (AU 1+4), nose wrinkling (AU 9) and depressing lower lip corner (AU 15) suggests the state *off-task.* Similarly appropriately directed activity on the mouse and keyboard can be a sign of engagement whereas no activity or sharp repetitive activities may be a sign of disengagement or irritation.

These surface level behaviors and their mappings are loosely summarized in table 1. Whether all of these are important, and are the right ones remains to be evaluated, and it will no doubt take many investigations. Such a set of behaviors may be culturally different and will likely vary with developmental age as well. The point we want to make is that we are examining a variety of surface level behaviors related to inferring the affective state of the user, while he or she is engaged in natural learning situations.

	On Task	***Off Task***
Posture	Leaning Forward, Sitting Upright	Slumping on the Chair, fidgeting
Eye-Gaze	Looking towards the problem	Looking everywhere else
Facial Expressions	Eyes Tightening(AU7), Widening(AU5), Raising Eyebrows (AU 1+2), Smile(AU6+12)	Lowering Eyebrow(AU1+4), Nose Wrinkling(AU9), Depressing lower lip corner(AU15)
Head Nod/ Head Shake	Up-Down Head Nod	Sideways Head Shake
Hand Movement	Typing, clicking mouse	Hands not on mouse/keyboard

Table 1. Surface Level Behaviors

Recognizing Surface Level Behavior

The detection of the surface level behaviors is critical to the performance of the *Learning Companion.* We have been working on mechanisms to sense posture, eye-gaze and facial expressions in an unobtrusive manner so that they don't interfere with the natural learning process.

Facial Features and Gaze Tracking. We have built a version of the IBM Blue Eyes Camera (http://www.almaden.ibm.com/cs/blueeyes) that tracks pupils unobtrusively. The pupil tracking system is shown in Figure 3. The system has an Infrared (IR) sensitive camera coupled with two sets of IR LEDs. One set of LEDs is on the optical axis and produces the *red eye* effect. The two sets of LEDs are switched on and off to generate two interlaced images for a single frame (Haro et al. [2000]).

The image where the on-axis LEDs are on has white pupils whereas the image where off-axis LEDs are on has black pupils. These two images are subtracted to get a difference image,which is used to track the pupils. Figure 4 shows a sample image, the de-interlaced images and the difference image obtained using the system.

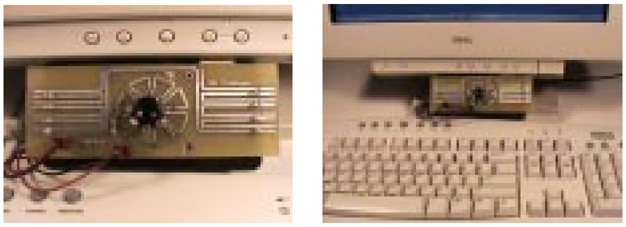

Figure 3. – Camera to track pupils, placed under monitor.

The facial action recognition system developed by Yingli Tian et al. [2001] requires the fitting of templates for facial features manually in the first frame. We are developing techniques using the pupil tracking system to automatically detect facial features like eyes, eyebrows, etc. in real time. The IR pupil tracking system is quite robust in different lighting conditions as well and can be used extensively to normalize the images and to determine the gaze-direction. We are building a system to detect head-nods and head-shakes using the position of pupils in the image. Also the physiological parameters like pupillary dilation, eye-blink rate etc. can be extracted to infer information about arousal and cognitive load.

Despite all these advantages the capability of the system is limited to the instances where the eyes are visible in the image. In our preliminary experiments we have observed that for a person using a computer in a normal manner, at least one pupil is visible (hence trackable) to the IR system for over 85 % of the time, whereas both the pupils are present for over 75% of the time. As the tracking of pupils depends upon the difference of two frames separated by approximately 1/60th of a second, a very quick movement by the person or a fast enough temporal source that can change the image in the odd field from the image in the even field (for example flicker of monitor) affects the tracker. Pattern recognition also needs to be added to disambiguate the pupils from other bright spots that show up due to IR reflections bouncing off earrings or eyeglasses.

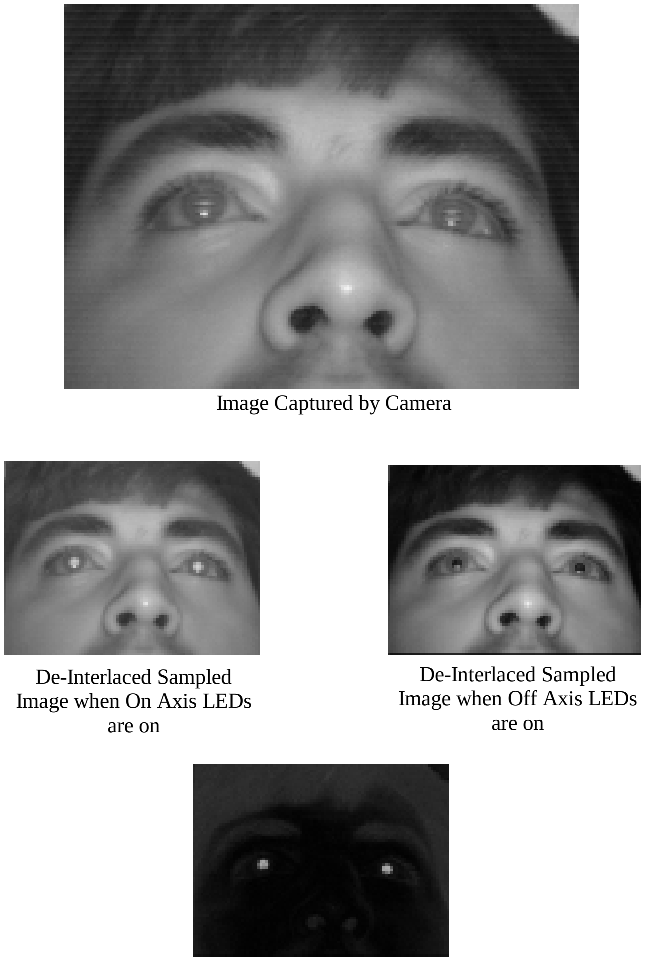

Image Captured by Camera

De-Interlaced Sampled Image when On Axis LEDs are on

De-Interlaced Sampled Image when Off Axis LEDs are on

Difference Image

Figure 4. – Pupil tracking with the Blue Eyes camera.

Recognizing Postures. Different postures are recognized using a sensor chair that uses an array of force sensitive resistors and is similar to the *Smart Chair* used by Tan et al. (1997). It consists of two 0.10 mm thick sensor sheets, with an array of 42-by-48 sensing units. Each unit outputs an 8-bit pressure reading. One of the sheets is placed on the backrest and one on the seat. The pressure distribution map (2 of 42x48 points) sensed at a sampling frequency of 50Hz is used to infer about the posture. The sensor chair is shown in Figure 5.

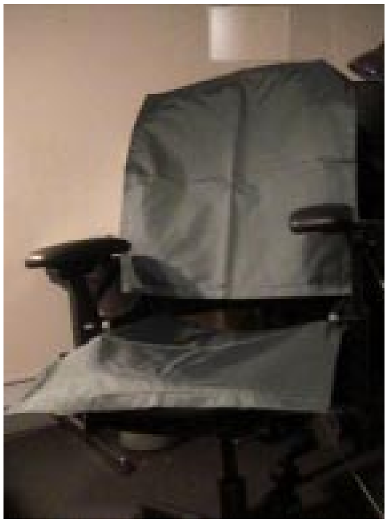

Figure 5. – The Sensor Chair

The real advantage of using this kind of system is that it supports very fast real time acquisition of data and does not depend upon the persons, surroundings etc. Furthermore the ergonomic requirements are minimal. The system can easily detect postures like whether the person is leaning forward or backward and whether he is slumped toward his side. Figure 6. shows some of the associated patterns with the different postures. It can track the joint positions of the lower body and detect swinging of feet as well.

Future Directions

We are in the process of further refining the sensors and algorithms to detect affective cues like posture, gaze direction, facial expressions etc. The functionality of the IBM Blue Eyes camera is being extended to real time tracking of gaze and recognition of FACS (Ekman 1978). Pattern recognition techniques are being used on the data gathered by the sensor chair to determine the posture in real time. Also we are analyzing the data collected by the IBM Blue Eyes Camera and the sensor chair to verify the mapping between the surface level behavior and the affective state. Ultimately our goal is to develop a multi-modal system for informing the *Learning Companion*, so that it is capable of recognizing in real time whether the student is *off-task*, and whether or not its intervention succeeds in helping the student return to being *on-task.*

Acknowledgments. We are indebted to Nancy Alvarado, Barry Kort and Rob Reilly for the insightful discussions. We thank Ted Selker, Surj Patel and Yue Hann Chin for helping with IBM Blues Eyes setup and Hong Tan for the help with the sensor chair. This research was supported by NSF ROLE grant 0087768.

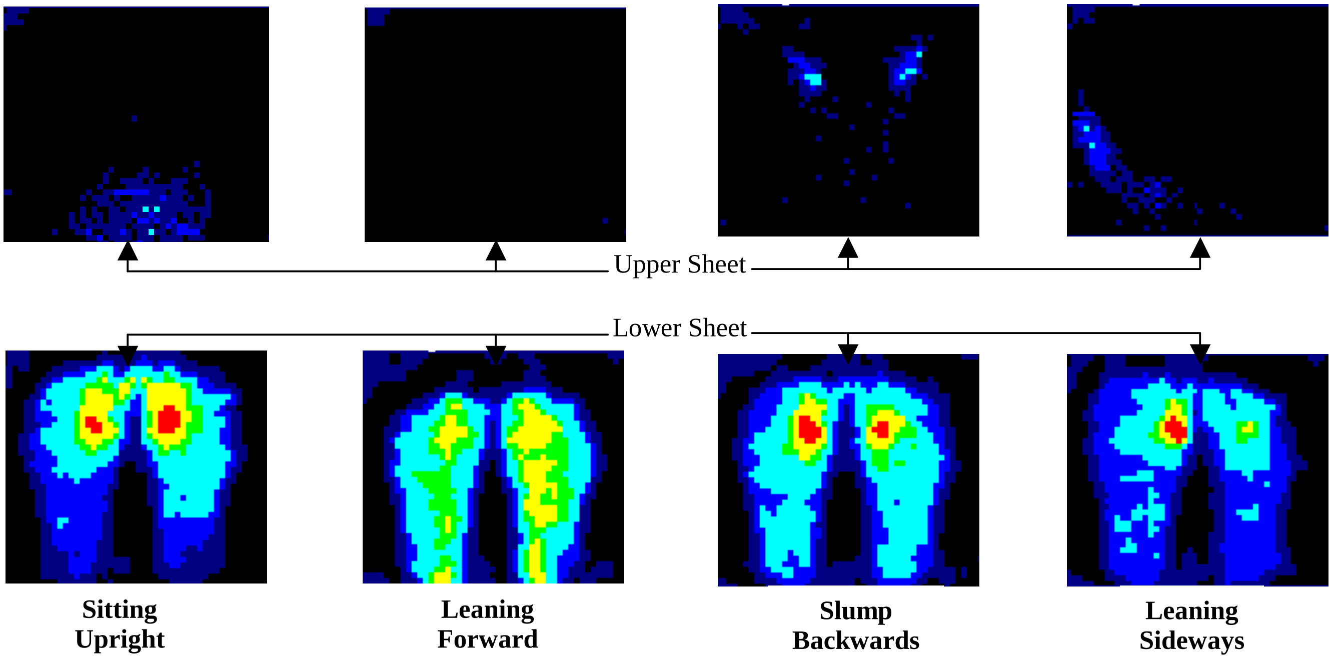

Figure 6. – Sensor chair patterns associated with postures

References

Chen, L.S., Huang, T. S., Miyasato, T., and Nakatsu, R. 1998. Multimodal Human Emotion/Expression Recognition. In *Proceedings of the Third International Conference on Automatic Face and Gesture Recognition*. Nara, Japan.

Del Soldato, T. 1994. Motivation in Tutoring Systems. Technical Report, CSRP 303, School of Cognitive and Computing Science, The University of Sussex, UK.

DeSilva, L.C., Miyasato, T., and Nakatsu, R. 1997. Facial Emotion Recognition using Multi-Modal Information. In *Proceedings of IEEE International Conference on Info., Communications and Signal Processing.* Singapore.

de Vicente, A. and Pain, H. 1999. Motivation Self-Report in ITS. In Lajoie, S. P. and Vivet, M. editors, *Proceedings of the Ninth World Conference on Artificial Intelligence in Education,* 651-653. Amsterdam, IOS Press.

Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P. and Sejnowski, T. J. 1999. Classifying Facial Actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21(10): 974-989.

Ekman, Paul, and Friesen, W. V. 1978. *Facial Action Coding System: A technique for the measurement of facial movement.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, Paul 1992. Are there Basic Emotions? *Psychological Review*, 99(3): 550-553.

Essa, I. and Pentland, A. 1997. Coding, Analysis, Interpretation and Recognition of Facial Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19(7):757-763.

Haro, A., Essa, I., Flickner, M. 2000. Detecting and Tracking Eyes by Using their Physiological Properties, Dynamics and Appearance. In *Proceedings of IEEE Computer Vision and Pattern Recognition*. SC.

Healey, J. 2000. Wearable and Automotive Systems for Affect Recognition from Physiology. Ph.D. thesis, MIT Media Lab.

Huang, T. S., Chen, L. S. and Tao, H. 1998. Bimodal Emotion Recognition by Man and Machine. *ATR Workshop on Virtual Communication Environments*. Kyoto, Japan.

Klein, J. 1999. Computer Response to User Frustration. Master's thesis, MIT Media Lab.

Kort, B., Reilly, R., Picard, R. W. 2001. An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion. In *Proceedings of IEEE International Conference on Advanced Learning Technologies*. Madison.

Leidelmeijer, K. 1991. *Emotions: An Experimental Approach*. Tilburg University Press.

Matsubara, Y. and Nagamachi, M., 1996. Motivation Systems and Motivation Models for Intelligent Tutoring. In Claude Frasson et al., editors, *Proceedings of the Third International Conference in Intelligent Tutoring Systems*.

Olson, R.K. and Wise, B. 1987. Computer Speech in Reading Instruction. In Reinking D., editors, *Computers and Reading: Issues in Theory and Practice*. New York: Teachers College Press.

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

Picard, R. W., 2001. Towards Computers that Recognize and Respond to User Emotions. *IBM Systems Journal*, vol. 39: 705-719.

Picard, R. W., Vyzas, E. and Healey, J. 2001. Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Forthcoming.

Plutchik, R. 1980. A General Psychoevolutionary Theory of Emotion. In Plutchik R. and Kellerman H., editors, *Emotion Theory, Research, and Experience: vol. 1, Theories of Emotion*. Academic Press.

Rich, C., Waters, R. C., Strohecker, C., Schabes, Y., Freeman, W. T., Torrance, M. C., Golding, A., Roth, M. 1994. A Prototype Interactive Environment for Collaboration and Learning. Technical Report, TR-94-06. http://www.merl.com/projects/emp/index.html

Scheirer, J., Fernandez, R. and Picard, R. W. 1999. Expression Glasses: A Wearable Device for Facial Expression Recognition, In *Proceedings of CHI*. Pittsburgh.

Tan H. Z., Ifung Lu and Pentland A. 1997. The Chair as a Novel Haptic User Interface. In *Proceedings of the Workshop on Perceptual User Interfaces*. Banff, Alberta, Canada.

Yacoob, Y. and Davis, L. 1996. Recognizing Human Facial Expressions from Long Image Sequences Using Optical Flow, *IEEE Transaction on Pattern Analysis and Machine Intell*igence, vol. 18(6): 636-642,

Yingli Tian, Kanade, T. and Cohn, J. F. 2001. Recognizing Action Units for Facial Expression Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23(2).