

Automatic Facial Expression Interpretation: Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect

Christine L. Lisetti

Department of Information Systems
University of South Florida
Tampa, FL 33620

Diane J. Schiano

Interval Research Corporation
1801 Page Mill Road
Palo Alto, CA 94304

1. Motivation

Recently there has been an increasing incentive from researchers in Human-Computer Interaction (HCI, henceforth) – who study the mental and physical world of computer users, develop approaches and methods to predict properties of the interactions and to support the design of interfaces – to perform a fundamental shift in the way they think about input-output interactions with a computer: a shift toward a human-centered interaction architecture, away from a machine-centered architecture. The main motivating principle is that computers should be adapting to people rather than vice versa.

Since faces are at the center of human-human communication, it would seem natural and desirable to find faces at the center of human-computer interaction. Indeed, there has been considerable technical progress within Artificial Intelligence (AI, henceforth) – an area of research which has evolved to develop “intelligent agents” applying machine learning techniques to help people interact with computers – in the field of computer vision to open the possibility of placing faces, if not at the center of human-computer interaction, at least at a significant place within man-machine interaction.

It still seems to be the case, however, that computer vision researchers, interested in *how* to solve problems, are working in isolation to develop ever more sophisticated algorithms to recognize and interpret facial information without necessarily knowing *what* they would do with the information, were they to solve the vision problem. On the other hand, researchers in Human-Computer Interaction are not necessarily aware of the recent progresses in computer vision, which may have brought the possibility of using facial information in computer interfaces closer than ever. The question of *what* to do with facial information when it becomes available may actually motivate and foster ongoing research in HCI, in Artificial Intelligence and in Cognitive Science. The purpose of this paper is to attempt to bring together people, results and questions from these three different disciplines – HCI, AI, and Cognitive Science – to explore the potential of building computer interfaces which understand and respond to the richness of the information conveyed in the human face. Until recently, information has been conveyed from the computer to the user mainly via the visual channel, whereas inputs from the user to the computer have been made from the keyboard and pointing devices via the user's motor channel. The recent emergence of multimodal interfaces as our everyday tools might restore a better balance between our physiology and sensory/motor skills, and impact (for the better we hope), the richness of activities we will find ourselves involved in. Given recent progress in user-interface primitives composed of gesture, speech, context and affect, it seems feasible to design environments which do not impose themselves as *computer environments*, but have a much more natural feeling associated with them.

In particular, as we reintroduce the use of all of our senses within our modern computer tools via multimodal devices-- or at least the Visual, Kinesthetic, and Auditory (**V, K, A**) – the possibility to take into account *both* cognition *and* emotions while designing computer interfaces opens up. New theories of cognition, indeed, emphasize the tight interface between affect and cognition. Given the increasing use of computers which support the human user in many kinds of tasks and activities, issues in affective computing (Picard 1997) “computing that relates to, arises from, or deliberately influences emotions” –

necessarily begin to emerge. Indeed, there is now plenty of evidence in neuroscience and psychology about the importance of emotional intelligence for the overall human performance in tasks such as rational decision-making, communicating, negotiating, and adapting to unpredictable environments. As a result, the newest kinds of computer environments, aimed at adapting to the user's needs and preferences through intelligent interfaces and customizable interface agents, need to be able to recognize, acknowledge, and respond appropriately to affective phenomena. We might also find an opportunity to better understand ourselves by building multimodal tools with increased awareness of the user's states, as well as to assist psychologists in developing and testing new theories of the human cognition-emotion complex system.

Our research group is working on the construction of a computer system capable of recognizing and responding to cognitive and affective states of users while they are involved performing various tasks and activities. In this present article, we discuss one of our projects, which is aimed at developing an automatic facial expression interpreter, mainly in terms of signaled emotions. We present some of the relevant findings on facial expressions from Cognitive Science and Psychology that can be understood by and be useful to researchers in Human-Computer Interaction and Artificial Intelligence. We then give an overview of HCI applications involving automated facial expression recognition, we survey some of latest progresses in this area reached by various approaches in computer vision, and we describe the design of our facial expression recognizer. We also give some background knowledge about our motivation for understanding facial expressions and we propose an architecture for a multimodal intelligent interface capable of recognizing and adapting to computer users' affective states. Finally, we discuss current interdisciplinary issues and research questions which will need to be addressed for further progress to be made in the promising area of computational facial expression recognition.

2. Facial Expressions and Facial Displays

Since Darwin (1872), the central preoccupation of researchers interested in the face has been to correlate movements of the face with emotional states. The advocates of this view, the "Emotion View", are not all homogeneous in all their opinions, but they do share the conviction that emotions are central in explaining facial movements (Ekman and Rosenberg 1997; Rinn 1984). The "Behavioral Ecology View", on the contrary, derives from accounts of the evolution of signaling behavior, and does not treat facial displays as *expressions of emotions*, but rather as *social signals of intent* which have meaning only in social context (Chovil 1991; Fridlund 1994). More recently facial expression has also been considered as *emotional activator*, contrary to being viewed solely as a response to emotional arousal (Zajonc 1994; Ekman and Davidson 1993; Camras 1992).

2.1. Emotion View: Expressions of Emotions

The Emotion View posits two basic kinds of facial actions. The first are the innate reflex-like facial actions that read out ongoing emotion, and display them with facial expressions of emotions. The second are learned instrumental facial actions that connote emotion that is not occurring, and reflect everyday social dissimulation such as the smile of politeness. In everyday life, the facial expressions observed are an interaction of emotional response and cultural convention.

The Emotion View proposes a small set of fundamental emotions that are reflexes or "affect programs" differentiated by natural selection, triggered by stimuli and accompanied by prototypical facial displays. Six basic emotions identified by their corresponding six universal expressions, and referred to with the following linguistic labels have been proposed (Ekman and Friesen 1975): *surprise*, *fear*, *anger*, *disgust*, *sadness*, and *happiness*. Recently *contempt* has also been counted among universal expressions as well. Variations from the prototypical emotional expressions are explained as reflecting the elicitation of blends of the fundamental emotions, or the effects of culture-specific conventions.

For example, there appears to be related emotions for each of the six basic emotions labeled above. The *surprise* group consists of a collection of possible expressions corresponding to different states: dazed surprise, questioning surprise, slight surprise, moderate surprise, etc. Different emotions can, moreover,

blend into one single facial expression to show sad-angry expressions, angry-afraid expressions (Ekman and Friesen 1975).

Discreteness of emotion, however, has not been universally determined: linguistic labels can be considered as somewhat artificial. One study indicates that facial expressions and emotion labels are probably associated, but that the association may vary with culture (Russell 1994). Furthermore, with her work on cross-cultural studies of emotions, Wierzbicka has found that what we refer to as universal emotions with labels such as *fear*, *surprise*, or *anger*, may well be culturally determined (Wierzbicka 1992). For example, Eskimos have many words for *anger*, but the Ilongot language of the Philippines, or the Ifaluk language of Micronesia do not have words corresponding in meaning to the English word *anger*.

The absence of universal emotion terms does not mean, however, that there cannot be any universal emotions, or that certain emotions cannot be matched throughout the entire world with universal facial expressions. According to Wierzbicka (1992b), “if there are certain emotions which can be matched, universally, with the same (identifiable) facial expressions, these emotions cannot necessarily be identified by means of English emotion terms, such as *sadness* or *anger*, because these terms embody concepts which are language- and culture-specific”.

2.2. Behavioral Ecology View: Signals of Intent

Facial expressions can also be considered as a modality for communication, the face being an independent channel conveying *conversational signals*. It is interesting to note that although the human face is capable of as many as 250,000 expressions, less than 100 sets of the expressions constitute distinct, meaningful symbols (Birdwhistle 1970). When considered as conversational signals, three main categories of signals from the face have been identified (Chovil 1991):

1. *Syntactic Displays*: used to stress words, or clauses. For example, raising or lowering eyebrows can be used as an *emphasizer* of a particular word or clause, as well as a *question mark*.
2. *Speaker Displays*: illustrate the ideas conveyed. For example:
 - *interactive “you know”* can be expressed by raising the eye brows, whereas the
 - *facial shrug “I don’t know”* can be expressed by the corners of the mouth being pulled up or down.
3. *Listener Comment Displays*: used in response to an utterance. For example, the
 - *understanding level* corresponding to a head nod is confident, whereas if the eyebrows are raised the level of understanding is moderate, and lowered eyebrows indicate a lack of confidence in understanding,
 - *Agreement “yes”* can be expressed by raising eyebrows,
 - *Incredulity* can be expressed with a longer duration of eyebrow raising.

These communicative signals have been introduced with speech dialogue computer interfaces (Nagao and Takeuchi 1994).

In the Behavioral Ecology View (Fridlund 1994), there are no fundamental emotions nor fundamental expressions. This view does not treat facial displays as “expressions” of discrete, internal emotional states, nor as the output of affect programs. Facial displays are considered as “signification of intent”, evolving in a certain fashion in response to specific selection pressures. Furthermore, they necessarily co-evolve with others’ responsivity to them. From this view point, the fact that these signals serve social motives does not imply that they are learned: innate social cognition mediates juvenile or adult displays.

For example, instead of being six or seven displays of “basic emotions” (e.g., *anger*), there may be one dozen or more “about to aggress” displays appropriate to the relationship of the communicators, and the context in which the interaction occurs, without the inner feeling. The actual form of the display may depend on the communicator personality traits (dominant/ or nondominant), and context (defending the territory or young, access to a female, retrieving stolen property).

Facial displays have meanings specifiable only in their context of occurrence, and they are issued to serve one's social motives in that context. These motives have no necessary relation to emotion, and a range of emotions can co-occur with one social motive. Displays are serving intent, and they are issued when they will optimally improve cultural or genetic inclusive fitness. Facial displays, therefore, depend upon the intent of the displayer, the behavior of the recipient, and the context of the interaction (not on inner feelings).

2.3. Brain Plasticity: Emotional Activators and Regulators

Given some of the newest results in neuroscience emphasizing the plasticity of the human brain, facial actions have also recently been considered as emotional activators and regulators. It seems that facial actions such as muscle movements (as opposed to *expression* of emotion) can generate emotion physiology (Ekman 1993). In particular some results relating voluntary smiles with changes in regional brain activity (Ekman and Davidson 1993) suggest that by voluntarily smiling, it is possible to generate deliberately some of the physiological change which occurs during spontaneous positive affect. Furthermore voluntary facial action can generate significant levels of subjective experience of the associated emotion and emotion-specific autonomic nervous system activity (Levenson 1992; Levenson et. al. 1990).

Another approach, the vascular theory of emotion, holds that facial gesture have regulatory and restorative functions for the vascular system of the head (Waynbaum 1907). A revised view of the theory (Zajonc 1989) postulates that facial action can produce changes in brain blood temperature that, in turn, have significant hedonic consequences. These hedonic consequences are produced for a variety of reasons. Subjective changes can be obtained because changes in hypothalamic temperature can facilitate or inhibit the release and synthesis of various emotion-linked neurotransmitters. For example, if a certain movement of facial muscles (arrived at by a phonetic utterances or changes in pattern of breathing) results in raising hypothalamic temperature, and if consequently norepinephrine is released, the person might experience *excitation*. Similarly if the temperature is lowered, serotonin is released and the person might feel *depressed*.

In short, anything that a person can do to change hypothalamic temperature might have subjective effects (Zajonc 1994). These changes in brain temperature can be achieved via breathing patterns and facial efference, which would explain why Yoga, Meditation, and Qigong, as well as old deep-breathing techniques (Fried 1987; 1990) can influence hedonic experience: these techniques are all capable of altering brain temperature because they rely on variations in facial efference (as with vocal output).

If such a plasticity of the human brain can be exercised (Camras 1992), the number of HCI applications that could spread from it are numerous. Imagine your computer application getting you to take a break when it perceives you as too tense. It could suggest facial movements, breathing exercises, or posture changes to help you "switch gears" and move to a more comfortable state.

2.4. Measuring Facial Signals

There exist various precise methods for measuring facial movements which require training. Fourteen of these have been reviewed (Ekman 1982). The most commonly used are FACS and MAX which have been compared (Hager 1985), and which we explain briefly here.

The Facial Action Coding System (FACS) (Ekman and Friesen 1976) is a widely used anatomically based system for measuring all visible facial movements. FACS describes facial activity in terms of muscle "action units" (AU). For example, an "inner brow raiser" corresponds to AU1, whereas a "jaw drop" corresponds to AU26 in FACS. FACS is free of theoretical bias about the possible meaning of facial behaviors. The analysis of the groups of muscles underlying all facial movements allowed to define 44 AUs to account for changes in facial expression, and 12 AUs to account for changes in gaze direction and head orientation. Using FACS, a trained coder can identify specific AUs present in an observed expression, and their duration and intensity is also recorded. The scores for a facial expression consist of the list of AUs which produced it. Among other uses, FACS has been used to define groupings underlying

the expression of emotions on the face. FACS has also been used to build an automated analysis of facial actions (Bartlett et al. 1998) explained later, which outperformed human non-experts on the task of classifying six upper facial actions.

The Maximally Discriminative Affect Coding System (MAX) (Izard 1979), is a theory based system, which also measures visible changes in the face. It does so by categorizing facial components based on the theory about what areas of the face should be involved in certain emotions. It uses units formulated in terms of appearances relevant (and only relevant) to eight specific emotions. It does not score facial movements which are not related to these eight emotions, contrary to the categorization of all possible movements of groups of muscles in FACS.

Using electromyography (EMG) – a technique to measure electrical activity by attaching electrodes to the surface of the face – it is also possible to measure activity that might not be visible (Cacioppo et al. 1990), and which is therefore not a social signal.

We would like to point out that research on human perception and measure of facial expressions has somewhat been plagued with inconsistent use of stimuli and response measures. For example, Ekman's largely used list of six basic or universal expressions of emotion derives largely from *forced-choice* responses, which has been pointed out by some as possibly constraining results. In *forced-choice* conditions, participants are shown facial images (or sequences of), choose one term from a list of emotion terms (*anger, disgust, fear, happiness, sadness, surprise, plus neutral*). Participants rate, on a scale of 0 (not present at all) to 6 (extremely high), the degree to which each of the listed emotions are displayed in the presented facial image. Another experimental procedure uses *multiple-choice*. In multiple-choice conditions, participants choose one term from the same list of emotions used in the forced-choice condition, and rate them on the same scale. In addition to the forced-choice options, participants can also type in any emotion term they think is displayed by the facial image, along with its intensity rating. Finally in the *open-ended* procedure, participants respond freely to each facial image by choosing a term of their choice which best describes the expression, and rate the degree of expressiveness found in the expression.

A study has carefully examined a *single* set of stimuli using the forced-choice, multiple-choice, and open-ended measures (Ehrich et al. 1998). The results of this study show that there is little evidence for response method bias in identifying Ekman's six basic expressions. A possible exception can be found in the case of *fear* (which is sometimes displayed by ambiguous expressions) and seemed constrained by the forced-choice procedure.

Finally another one of the controversial issues still disputed, is whether facial expressions are perceived by humans as varying continuously along certain underlying *dimensions*, or as belonging to qualitatively discrete *categories*.

From the *dimensional perspective*, or the circumplex model (Russell 1980), more extreme degrees of an emotion fall around the edge of a two-dimensional "emotion space" encoding orthogonal bipolar dimensions of pleasure and arousal, with milder emotions falling more toward the center. The consequences of dimensional accounts is that linear transitions from one expression to another will be accompanied by characteristic changes in identification. For example, a transition from a *happy* face to an *angry* face will need to pass through a neutral face. Any transition between expressions lying at opposite points in the emotion space will need to pass through a neutral expression, whereas transitions between expressions which do not involve entering the region of any another emotion in the emotion space can be relatively abrupt. Dimensional accounts can be used to predict the consequences for identification of physically transforming one facial expression to another, changing one facial expression to another by artificial computer "morphing" of the facial images (as many computer vision researchers have done to increase the size of their database of facial images).

From the *categorical perspective*, on the other hand, quite a few studies indicate that emotional expressions, like colors and speech sounds, are perceived categorically, not as a direct reflection of their continuous physical properties (Etkoff and Magee 1992; Young et.al 1997). These studies are founded on

the phenomenon that linear physical changes in a stimulus can have non-linear perceptual effects, with changes occurring near or across category boundaries being easier to detect. These results have also influenced computer vision researchers (Padgett and Cottrell 1998) described later.

However, using the multiple-choice method for rating perception of facial expressions described earlier, a study has found that many facial expression transitions are perceived continuously rather than categorically (Ehrlich et al. 1998). In other words, opposites in emotion space (e.g. transition from *happy* face to *angry* face) do pass through a *neutral* face.

All these issues are highly relevant to AI researchers in computer vision because many AI-based systems aimed at automatic facial expression interpretation are implemented and tested sometimes without comparison of the system performance to human experts and non-experts. These ad-hoc evaluations of computer systems performance could lead to some misleading claims about computational recognition of facial expressions. To be carried out appropriately, these performance tests often require the expertise of psychologists and cognitive scientists, which AI computer vision research do not often have. Similarly, psychologists and cognitive scientists will strongly benefit from progress in computer vision because facial expression recognition is a very timely process when carried out by experts: it requires more than one hour to measure each minute of facial expression (Ekman et al. 1992). This means that the nature of facial expression perception and recognition is inherently interdisciplinary. This article is an attempt at bridging some gaps between these disciplines. We shall see in the following section how computer vision researchers have (and have not) made use of these findings from Psychology and Cognitive Science.

3. Automatic Facial Expression Recognition and Its Applications

3.1. Human-Computer Interaction Applications and Relevant Facial Expressions

It is expected that three to ten years from now, the price of digital cameras will have dropped considerably. This will make visual awareness research for artificially intelligent systems an interesting alley for developing new computer environments. In particular, there exist a number of applications for HCI to make use of automatic facial expression recognition. The main motivating principle for such applications is to give the ability to computers to adapt to the people's natural abilities rather than vice versa.

Relevant expressions and their interpretations may vary depending upon the chosen type of application. As we have explained earlier, facial expressions can indeed be considered as expressing communicative signals of intent, or expressing emotional inner states, or even as emotion activators. We list here some of the most "common-sense" type of applications which we hope will motivate further research in facial expression interpretation. We will discuss some of the psychological issues associated with the different ways to view facial expressions and the privacy/ethical questions that some of these applications may raise in a later section:

- *User coaching*: It would be interesting to work with expressions corresponding to *surprise*, *confusion*, *frustration* and *satisfaction*, for example, while monitoring a user's learning process. A program like COACH (Selker 1994) which assists learners of the high-level programming language Lisp by adapting its feedback, details of explanations and suggestions based on the level of expertise of its current user, would strongly benefit from being able to adapt its feedback to the level of emotional intensity as well (Lisetti 1999).
- *Distance learning/tele-teaching assistant*: Instructors who have experimented with distance learning via tele-teaching, often report not being able to grasp the level of understanding of their distant students. As many of us who teach will probably agree, we often rely on feedback from our students' faces to pace the lecture, or to repeat a portion of the lecture if too many faces express *confusion*. We also know how to move on to the next point when students' faces change to express *understanding* and *confidence*. With tele-teaching, however, this information is not available to the instructor. Being able to have such feedback from a facial expression recognizer operating on the recorded videos from the

distant learners would seem to be an invaluable tool. For example, the possibility of generating averages about the global level of understanding of the class by simply displaying to the teacher a label, say “Tele-class is 95% confused at present” might be welcomed by instructors wishing to be able to gauge the level of understanding of their remote students.

- *Automobile driver alertness/drowsiness monitor*: Many car manufacturer companies are currently working on adding sensors to the exterior of cars to improve the cruise control option. Sensors will soon be able to perceive the ravine or wall approaching, and stop or slow down the engine to avoid accidents. Sensors might also be added to the inside of the car to monitor the *sleepiness, drowsiness* or *alertness* level of the driver. With such information available, the car equipment could either send out a nice spray of cold water to the driver’s face, or issue a visual or vocal warning. Lastly, if the sensor perceives “driver is sleeping”, it could issue a wise message to the engine to stop the car!
- *Stress detector*: A similar situation to the drowsiness/alertness monitor described above can be envisioned with air plane pilots and other workers potentially exposed to highly stressful situations, where any indication of *panic* or *distress* is source of concern and needs to be monitored. A facial expression recognizer could act as a *stress* detector in a similar way as lie detectors can assist at detecting lies. It is important to note that of course there are accuracy issues in both cases.
- *Lie detector*: There are indications that micro-expressions can reveal whether one is telling the truth or lying (Ekman et al. 1988). These subtle differences among “true” and “false” expressions seem to be found among the different forms of smiles that people portray. When people are actually enjoying themselves, the smile of their lips is accompanied with muscular activity around the eye, whereas when enjoyment is feigned to conceal negative emotions the muscle that orbits the eye shows no activity (Ekman et al. 1990). The current progresses in computer vision indicate that it is in the domain of the feasible to train a system to identify and detect only selected facial actions with high accuracy. A lie detector based on some of the psychological results mentioned might find more applications than the current technology available for lie detection such as the polygraph. Indeed, such a system would be able to operate in real-time in court rooms, police head-quarters, or anywhere truthfulness is of crucial importance. Again ethical issues that such applications might raise need to be addressed.
- *Computer/phone conversation facial data encoder*: With the mass appeal of Internet-centered applications, it has become obvious that the digital computer is no longer viewed as a machine whose main purpose is to compute, but rather as a machine (with its attendant peripherals and networks) that provides new ways for people to communicate with other people. The excitement of posting cameras on top of a personal computer which displays real-time the face of a loved one or of a co-worker while exchanging email is surely an indication of such a shift from computing to communicating. Technically speaking however, the amount of information that needs to be transferred back and forth might exceed the reasonable bandwidth for good timely communication. Instead it is also possible to envision decoding the actual images collected with the digital cameras, interpreting the expressions displayed, and rather than sending the entire images, communicating just the result of the interpretation, say “Annie looks *happy* now”. A further option involves animating a chosen avatar able to display an array of previously stored expressions. In that manner, the avatar on the receiving end of the communication, could be animated based upon the result of the interpretation of the current expression of the sender. The animation might still be as informative in terms of the sender’s intended expression and as pleasant as the real images, and it would solve the bandwidth problem of sending large data files. Having information on facial expressions would also resolve many misunderstandings observed during communicative exchanges in which facial expressions are unavailable.
- *User personality type recognizer*: The field of User-Modeling has been growing and would benefit from identifying the user’s personality type, such as *introvert/extrovert*. The question of whether facial expressions can reveal or are related to personality types is an interesting one which needs to be studied. Access to this type of information would enable personalizing interfaces to match user’s personality preferences. For example, while some individuals might prefer working with people with very pro-active personalities, others would rather work with people with a more laid-back

personalities. These various personality types could be encoded in computer agents as well. Discreet digital assistants are more likely to be preferred by introverted users, while extraverted users might like assistants to be more interactive and visible.

- *Software product testing analyzer*: Too often data about whether a piece of software is liked by and considered helpful to users, or whether is simply too confusing, are collected by interrupting users and asking them to fill out questionnaires about how the product satisfies their needs. This practice disrupts the workflow and is often unwelcomed by the subjects involved with this type of questionnaire. Automatic facial expression recognition could provide analyzed data about product satisfaction. A record of the user's facial expressions would need to be kept and associated with statistical data about, for example, how many times did the users show *confusion* or *frustration*, or *content* at a specific stage during the interaction. Inferences would be drawn from such data about the software usability, without the need to interrupt the person interacting with the software to collect the data.
- *Entertainment and computer games*: Perhaps the most obvious realm of applications that can benefit from automated facial expression recognition is the field of entertainment and computer games. A game that can know whether its player is *happily surprised*, *fearfully surprised*, *confused*, or *intrigued* and can respond to the player's states would most certainly be more entertaining than one that is oblivious to them.
- *Health and family planning*: Working closely with psychologists and social workers, a system could assist detecting emotions of particular importance to psychopathology, and generate pre-diagnostic information as well keep a record of how facial action patterns change after treatment. For depression detection, for example, reoccurring and frequent displays of *sadness* could be an indication that the patient might be depressed. After treatment, facial displays that showed an increase in a *happy* face, would be an indication that the patient is doing better. Furthermore, the degree to which people can dissociate emotion expression from emotion experience (Gross 1998). Software products capable of monitoring the degree of dissociation throughout various activities could lead to beneficial information on emotion regulation.
- *Human-human communication trainer*: When facial expressions are considered as associated with inner feelings and emotions, they can help one understand the feelings of others, and they can help one to become more aware of one's own feelings. They can also assist in determining people's expressive style e.g., the *facial withholder* who does not express much on one's face; the *unwitting expressor* who expresses without even knowing or wanting to express ones' feelings; or the *substitute expressor* who expresses one emotion while actually experiencing another. Very often people are less aware of their own expressive style than their mates and coworkers for the obvious reason that they most of time do see their own facial expressions. By providing facial images with their interpretation during a communicative exchange (say over the Internet), the trainer system would enable people to become aware of their own expressions, possibly learning from them, and having the possibility of adjusting their expressive style, or disambiguating their expressions to the people they communicate with.
- *Ubiquitous computers*: The number of facial expressions relevant to HCI is likely to explode with the number of "invisible" computer applications still to come (Norman 1998). The time when rooms, walls, desks, and blackboards will all be invisibly given computing power is not far (Hiroshi, 1997; Mozer, M. 1998). Imagine your bedroom walls being able to perceive the moment when you actually fall *asleep*, and know that this is the time to turn off your lamp. Older people and insomniacs have often reported losing their opportunity to sleep, by having to move (say to turn the switch off) at this very precious moment between sleep and wake states.

While we are aware that all these applications raise important issues about privacy, precision, and error correction (covered in a later section), the advantage of some applications makes the research effort worthwhile.

3.2. A Review of the Some AI-based Approaches to Automatic Facial Expression Recognition.

Quite a lot of research has been done already in the field of face and gesture recognition with some outstanding results which led to the creation in 1995 of an international conference dedicated specifically to face and gesture recognition: The *International Conference on Automatic Face and Gesture Recognition*. Some of the major accomplishments in the field range from tracking the face, to segmenting the face, to recognizing the face for identification, among other projects.

While there has been considerable research done over the past century about facial actions signaling emotions, strategies for enabling their recognition and their memory store/recall have received little attention compared to the question of whole face recognition.

Similarly, most of computer vision research on the human face has concentrated on issues of identification of a person by name, recognition of a known face by true/false answer, or categorization of a face by gender, race, or age. Recent approaches are studying the importance of facial expressions for these three tasks, or how systems can be trained for face recognition regardless of facial expressions (Dror et al. 1996). Indeed, for a number of applications involving face identification and recognition, such as real-life settings in banks or airports, the facial expressions and orientations of the face on the pictures are multiple and unpredictable. Not only is automatic facial expression recognition of interest on its own (as illustrated previously by the numerous applications that can derive from it), but it is also expected to assist automatic facial identification processing.

For these reasons, there has recently been a growing interest in the field of computer vision to specifically recognize facial expressions. A number of systems have dealt with issues relevant to facial expression recognition using three different technical approaches such as: (1) image motion (Mase 1991, Black and Yacoob 1995, 1995b; Rosenblum, Yacoob, and Davis 1996; Essa, Darrell and Pentland 1994; Essa and Pentland 1997); (2) anatomical models (Mase 1991; Terzopoulos and Waters 1993; Essa and Pentland 1997); (3) neural networks, eigenfaces and holons (Cottrell and Metcalfe 1991; Padgett and Cottrell 1997; Padgett and Cottrell 1998; Lisetti and Rumelhart 1998); (4) hybrid systems (Rosenblum et al. 1996; Bartlett et al. 1999). We give a brief overview of the latest progress in the field, by separating two major technical currents: non-connectionist approach and the connectionist or neural network approach.

3.2.1. The Non-Connectionist Approaches to Facial Expression Recognition

Most of the non-connectionist approaches to processing facial information (surveyed in Samal and Iyengar 1992) typically assume a *feature-based representation* of the faces. Faces are represented in terms of distances, angles, and areas between features such as eyes, nose, or chin, or in terms of moment invariant. These parameters are extracted from full facial views or profiles, and each face is then stored as a “feature vector” in a location in memory. These models usually carry the following the processes: (1) find individual features (eyes, nose, chin), (2) measure statistical parameters to describe those features and their relationship, (3) extract descriptors. These approaches are usually efficient when the selected features are appropriate, but appropriate feature selection is not always an easy task. Furthermore, these types of models are limited to matching a target face with a face already present in the database, which means that the models cannot perform well on new face images.

Since the survey of computational approaches to facial processing (Samal and Iyengar 1992), a number of attempts have been made to recognize facial expressions in particular. One rule-based system, the JANUS system (Kearney and McKenzie 1993), has pursued machine interpretation of facial expressions in terms of signaled emotions. In this memory-based expert system, the machine interpreter identified the six basic expressions: *anger*, *happiness*, *surprise*, *disgust*, *fear*, and *sad* (Ekman and Friesen 1975). The model adopted a representation in line with Ekman’s earlier Facial Affect Scoring Technique (FAST) (Ekman, Friesen and Tomkins 1971) which is less anatomically precise than FACS (Ekman and Friesen 1976).

JANUS first converted face geometry into a static facial action format (such as “brows raised”, “eyes open”, “jaw dropped”, etc). It then classified an expression by matching it to the typical muscle

movements (i.e. the facial action units) found in the six basic expressions. The output of the system was in the form of an emotion label such as “*happy*”, or “*sad*”. One of the strength of this study is that it also measures the performance of the model by comparing it thoroughly to human experts and non-experts in facial expression recognition. The results suggest that the interpretations achieved by JANUS are generally consistent with those of college students without formal instruction in emotion signals.

Other models have used anatomical models of the face (Terzopoulos and Waters 1993) to interpret facial expressions (Essa and Pentland 1997), and to synthesize them as well. One of the limitations of such approaches is that the accuracy of their results depends entirely on the validity of the anatomical model of the face that they use. However, building a deterministic model of the human face with all of its possible motions is a very challenging task, considering the range of differences that can exist across individual faces (sizes, relative locations, etc.)

Other approaches analyze *motion* and extract dynamic muscle actions from sequences of Images, usually using optical flow. This method estimates motion vectors as an array of points over the face, which is called the *flow field*. It is usually represented by an array of arrows (or flow vectors) indicating the direction and the magnitude of the displacement at each image location.

In an early experiment (Mase 1991), facial actions forming a 15-dimensional feature vector was used to categorize four expressions: *happiness*, *anger*, *surprise*, and *disgust*. Nineteen out of twenty-two test data were identified correctly. Another rule-based system guided by psychological findings (Yacoob and Davis 1994) was based on the analysis of optical flow fields computed from image sequences of human subjects, combined with feature-based representation of the face. It recognized six expressions by constructing mid- and high- level representations of facial actions (e.g., “mouth opening”, “eyebrow raising”, etc.). The optical flow vectors, which were used as input to the system, were computed at the high intensity gradient points of the primary features of the face (mouth, eyes, eyebrows, and nose). The system however, made the limiting assumption that the head was not undergoing any rigid motion during the display of facial expressions.

Indeed, recognizing facial expressions in sequences of images with head motion is a difficult problem due to the occurrence of both rigid and non-rigid motion. This problem has been addressed by a model of rigid and non-rigid motion using a collection of parametric models (Black and Yacoob 1995b). The image motion of the face, mouth, eyebrows, and eyes are modeled using image flow models with only a few parameters which correspond to various facial expressions. The database generated comprised 70 image sequences of 40 subjects, each sequence including 1-3 expressions of the six basic expressions (Ekman and Friesen 1975), and a total of 128 expressions. The expressions of *fear* and *sadness* were found to be difficult to elicit compared to the other four. The subjects were also asked to move their head while avoiding profile views.

The system achieved a successful recognition rate from 87% to 100% depending on which emotion was recognized (*anger* and *disgust* ranking the highest). It is not clear however how the recognition rate was arrived at. It seems that the system results were compared to what expression was expected, given that it had been elicited. There seems to be no comparison, however, between the system results and what a human subject would have rated the input image sequences. We have pointed out earlier the need to address carefully this issue, in light of the various procedures for measuring facial actions described above.

3.2.2. The Connectionist Approach to Facial Expression Recognition

Connectionist models are networks of simple interconnected processing units that operate in parallel. The term neural network is also sometimes used to refer to the same approach because the processing units of the network are neuron-like, and they function metaphorically as neurons in the human body. Each unit receives inputs from other units, sums them, and calculates the output to be send to the other units that this unit is connected to. Learning or the acquiring of knowledge, results from the modification of the strength of the connection (the *weight*, akin to a synaptic connection) between interconnected units: the higher the weight value, the stronger the connection, the higher propensity of one neuron to cause its neighboring

neurons to become active or to fire, the more learning happens, i.e. the more knowledge is retained in the strength of the connection. Neural networks can learn patterns for object recognition by *unsupervised clustering* (i.e. no specific target is provided as the target label/response to be learned), or by *supervised learning/training by example* (i.e. the input and the corresponding desired target response are both given). Networks learn by successively modifying the strengths of the connections between units, in a direction to reduce the error at the output.

Neural networks offer a promising potential for face processing because they are notoriously known for dealing well with noisy, partial, potentially conflicting data. They appear to be most powerful when explicit *a priori* knowledge about the data to be categorized is unknown, as with the human face and all its possible variations in motions and individuals. Most importantly, with their ability to learn by example, neural networks can apply their learning acquired during training, and generalize well to newly encountered data. They can accurately approximate unknown systems based on sparse sets of noisy data. Temporal neural networks, which are specifically designed to deal with temporal signals, further expand the application domain in facial information processing, and multimedia processing in general (Kung and Hwang 1998).

Connectionist or neural network models of face processing typically operate directly on *image-based representations* of faces in the form of two-dimensional (2D) pixel intensity arrays, by contrast with non-connectionist approaches which usually use feature-based representations of the face. In a survey of connectionist models of face processing (Valentin et al. 1994), categorizes neural networks in two categories: (1) the linear autoassociators which implement principal component analysis, and which have been mostly used in tasks including recognition and categorization of faces by sex and race; (2) nonlinear three-layer autoassociative networks with a back-propagation algorithm, which have been used as compression networks.

Principal Component Analysis (PCA) finds a set of component dimensions that account for the differences in the data set. Such a PCA representation of a facial image – also known as *eigenface* (Turk and Pentland 1991) or *holon* (Fleming and Cottrell 1990) and which looks like a ghost face – is a set of templates where similarities of a facial image to each component image is measured. The principal components are the *eigenvectors* extracted from the cross-product matrix of a set of facial images, in which each image is treated as a vector. The eigenvectors represent a set of “features” characterizing the face, which are different from the ones assumed in the feature-based models described above. These features, also referred to as *macro-features*, are not defined *a priori*, but they are generated *a posteriori* on a statistical basis (hence the flexibility and advantage of such representations). For example, from a database of facial images made up of a majority of faces of one race and a minority of faces of another race, the eigenvector representation will represent featural information of the majority-race faces, but will be less likely to discriminate between minority-race faces.

Another learning algorithm associated with neural networks is *Back Propagation* (BP) (Rumelhart, Hinton, and William 1986; Chauvin and Rumelhart 1995). Backpropagation networks (in contrast to linear associative networks in which the input units are directly connected to the output units), include a layer of nonlinear hidden units between the input and the output units. This algorithm involves two phases: (1) in the propagation phase, a forward flow of activation is generated from the input layer to the output layer via the hidden layer; each hidden unit computes a weighted sum of its inputs and transforms it via an activation function (most often a logistic function); (2) in the backpropagation phase, the difference between the actual output and the desired target response, i.e. the *error*, is calculated, and it is backpropagated through the network, layer by layer; the weight between connected units are adjusted in order to minimize the error between the network output and the desired target.

Holons or eigenfaces have been used as input to a backpropagation network for performing face, emotion and gender recognition (Cottrell and Metcalfe 1991). Tested on different facial expressions, the network was found to be more confused than human subjects on distant emotions in the emotion-space described earlier (Russell 1980). More recently, three different representation schemes of facial images using a single neural network to classify six facial expressions were compared (Padgett and Cottrell 1997). The three representations were: eigenfaces, eigenfeatures (or macro-features) of the eye and mouth areas, and a

projection of the eye and mouth areas onto the eigenvectors obtained from random image patches. The latter system performed with 86% accuracy in its generalization on novel images.

It should be noted that, while the connectionist model has the advantage of preserving the relationship between features and texture, it has a very high sensitivity to variations in lighting conditions, head orientation, and size of the picture (Valentin, Abdi, O'Toole, and Cottrell 1994). These problems justify a large amount of work in preprocessing the images. Normalization for size and position is necessary and can be performed automatically with algorithms for locating the face in the image and rescaling it (Turk and Pentland 1991). Region or feature tracking can also be implemented. Such an algorithm for range data has been developed (Yacoob and Davis 1993) and a similar algorithm for intensity images could (and needs to) be developed.

Another approach, involving learning correlation of facial feature motion patterns and human emotions (Rosenblum et al. 1996), uses a hierarchy of networks approach based on emotion decomposition. With this hierarchical approach, the complexity of recognizing facial expressions is divided into three layers of decomposition. The first layer identifies expression of emotion, and occurs at the network level (six separate networks in total, one for each of the six "basic" emotions): during training the network is only exposed to one expression of emotion for multiple subjects. The second layer identifies motion of three facial features (the right and left eyebrows, and the mouth), and it is at the sub-network level: each emotion network is broken into 3 sub-networks, where each sub-network specializes in a particular facial feature. The third layer recovers motion directions, and it is at the sub-sub-network level: each one is sensitive to one direction of motion (up, down, right, or left) for a specific pre-assigned facial feature for a specific emotion.

Individual networks were trained to recognize the *smile* and *surprise* expressions. When tested for its ability to perform successfully on familiar sequences, i.e. *retention*, the success rate was 88%. When tested on its ability to perform successfully on sequences of unfamiliar faces, i.e. *generalization*, the success rate was 73%. When tested on its ability to reject a sequence that did not express the emotion that the network was tuned for, i.e. *rejection*, the success rate was 79%.

Another neural network, using the backpropagation algorithm, was built to recognize facial expressions with various degrees of expressiveness (Lisetti and Rumelhart 1998) as opposed to the more common automated process of recognizing extreme expressions only. The network design (shown in figure 1) was inspired by the notion that there are three areas of the face capable of independent muscular movement: the brow/forehead; the eyes/lids and root of the nose; and the lower face including the cheeks, mouth, most of the nose and the chin (Ekman 1975).

Furthermore, not every expression is shown in the same area of the face. For example, *surprise* is often shown in the upper part of the face with wrinkles in the forehead and raised eyebrows, while *smile* is mostly shown in the lower face (and in the eye muscles as discussed above). In our database, we only included two different poses per person: namely one full face with a *neutral* expression, and the other full face with a *smile*. Not every one of the pictures had the same degree of neutrality, or the same degree of "smilingness". We have designed various approaches to test this scalability among images.

The network includes one layer of input units, one layer of hidden units and one layer of output units. The input units are 2D pixel intensity arrays of the cropped images. The output units express the value of expressiveness of a particular expression ranging from 1 to 6. The hidden layer is connected to each portion of the face and to each output unit in order to simulate "holistic" face perception. Alternatively, the connections can be loosened, to model more local pattern recognition algorithms. Finally, because, in the future, we also want to be able to refer to the recognized expression with a label such as *happy*, or *angry*, we provide an ordered input mask of binary values for each of the emotions to be recognized. These can be the six basic emotions, plus *neutral*. These binary output values are to be turned on if the expression is identified, while the others remain turned off. The output also includes the level of expressiveness of the facial image.

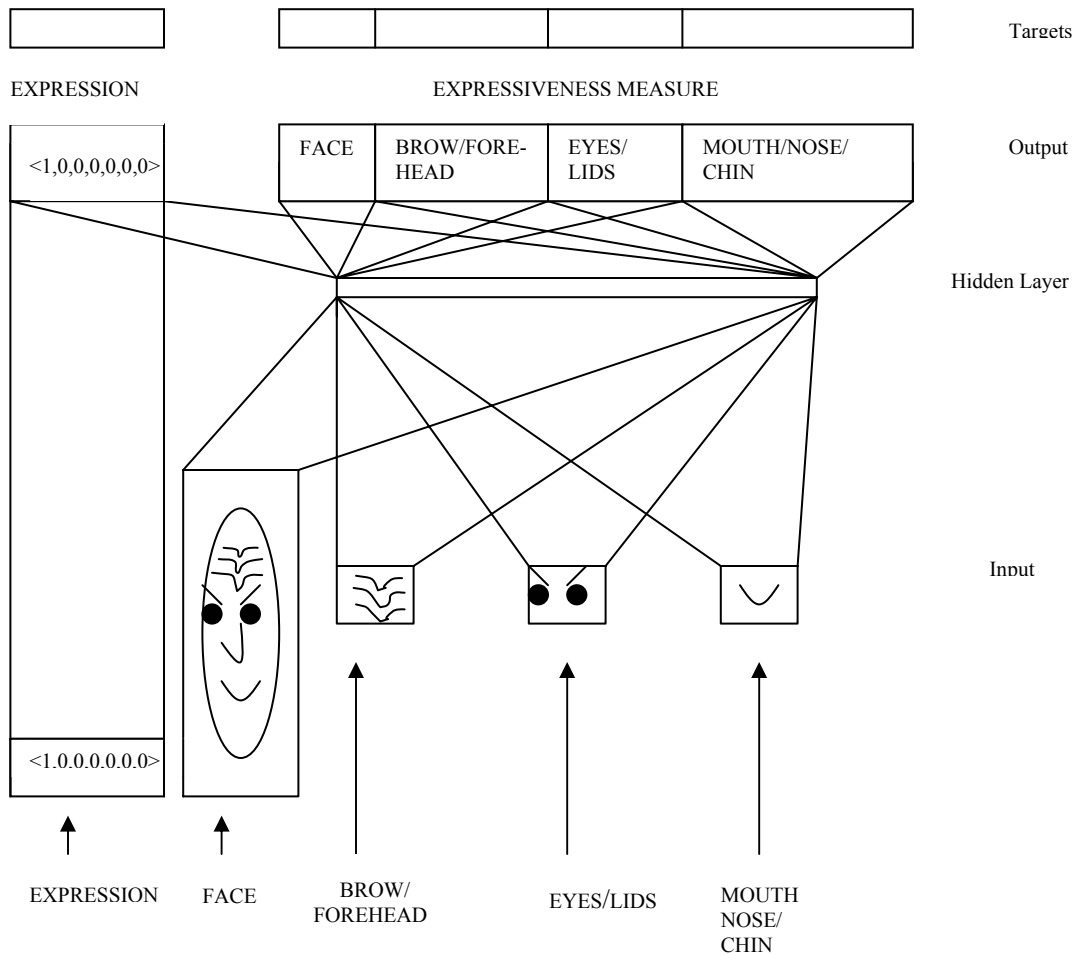


Figure 1: Network Design

In order to test each portion of the network, however, we built sub-networks each dealing with different portions of the face which are described in (Lisetti and Rumelhart 1998). We designed two procedures to test how well the network could generalize on new images that it had not been trained with: (1) testing whether the network could generalize if it had been trained on the person with one expression but not on the other, (2) testing whether the network could generalize on people which it had never been exposed to (i.e. with neither expression).

Our results for retention and generalization suggest that zooming in particular areas of the face for expression interpretation offers better results than processing the full face as a whole. Similar results were arrived at by a different study (Padgett and Cottrell 1997) for expressions without various degrees of expressiveness.

Recently more promising results have been found with a neural network (Padgett and Cottrell 1998) which models categorical perception of facial expressions (Young et al. 1997) described above. The network model consists feed-forward neural networks trained with the backpropagation learning algorithm, using

principal components of patches randomly selected from the face image. The results show very close correlation between the human study and the network performance, with some exception.

Finally, from a study that compared three approaches to the problem of automatically detecting facial action, (Bartlett et al. 1999), a hybrid system was developed which combined the three approaches: holistic spatial principal component analysis, measurement of local image features, and motion flow fields estimation. This hybrid system was shown to classify six upper facial actions with 92% accuracy. In this task, the system performed better than human non-experts, and as well as highly trained experts at facial expression recognition.

The recent results from the various neural network approaches that have been applied toward automated facial expression recognition show that a lot of progress has been made since the conception of the early neural network, a perceptron, that could classify smiles from frowns (Stonham 1986). These recent attempts suggest that facial expression recognition using a computational approach appears quite feasible. Given that facial expressions have been associated with inner emotional states, we now explain how an automatic facial expression interpreter could be integrated in a larger system that recognizes affect from multiple modalities.

4. An Overview of Multimodal Affect Processing in Human-Computer Interaction

4.1. The Interface between Affect and Cognition and its relevance to HCI

The field of human-computer interaction has recently witnessed an explosion of adaptive and customizable human-computer interfaces which use cognitive user modeling, for example, to extract and represent a student's knowledge, skills, and goals, to help users find information in hypermedia applications, or to tailor information presentation to the user (Jameson, Paris, and Tasso 1997). Computer interfaces can also adapt to a specific user, choose suitable teaching exercises or interventions (Selker 1994), give the user feedback about the user's knowledge, and predict the user's future behavior such as answers, goals, preferences, and actions.

It occurred to us that researchers in HCI and AI could benefit from learning more about the unsuspected *strong interface between affect and cognition*. Affective states play an important role in many aspects of the activities we find ourselves involved in, not excluding tasks performed in front of a computer. Being aware of how the user is receiving a piece of information provided would be very valuable. Is the user *satisfied, more confused, frustrated, or simply sleepy?* Being able to know when the user needs more feedback, by not only keeping track of the user's actions (Selker 1994), but also by observing cues about the user's emotional experience, also has advantages.

Indeed, as a result of recent findings, emotions are now considered to be associated with adaptive, organizing, and energizing processes. We mention a few already identified phenomena of interaction between affect and cognition, which we expect will be further studied and manipulated by building intelligent interfaces which acknowledge such an interaction:

- *organization of memory and learning*: we recall an event better when we are in the same mood as when the learning occurred (Bower 1981). A computer learning environment could aim at eliciting the same affective states as previously experienced (and recorded by the environment) during the learning process in order to reduce the learner's cognitive overload.
- *perception*: when we are happy, our perception is biased at selecting happy events, likewise for negative emotions (Bower 1981). While making decisions, users are often influenced by their affective states: for example, reading a text while experiencing a negatively valenced emotional state often leads to a very different interpretation than reading the same text in a positive state. A computer interface aware of the user's current emotional state can tailor the textual information to maximize the user's understanding of the intended meaning of the text.

- *goal generation, evaluation, and decision-making*: patients who have damage in their frontal lobes (cortex communication with limbic system is altered) become unable to feel, which results in their complete dysfunctionality in real-life settings where they are unable to decide what is the next action they need to perform (Damasio 1994). Normal emotional arousal, on the other hand, is intertwined with goal generation and decision-making. Emotions are also very influential for prioritizing activities by levels of importance and for determining values.
- *cognitive style changing*: when people are in a positive mood, they experience more creative, expansive, and divergent thinking. When they are in a negative mood, on the contrary, they tend to be more conservative, linear, and sequential in their thinking process and they experience having less options.
- *strategic planning*: when time constraints are such that quick action is needed (as in fear of a rattlesnake), neurological shortcut pathways for deciding upon the next appropriate action are preferred over more optimal but slower ones (Ledoux 1992). Furthermore people with different personalities can have very different preference models (Myers-Briggs Type Indicator) and user models of personality (Paranagama et al. 1997) can be further refined with the user's affective profile.
- *focus and attention*: emotions restrict the range of cue utilization such that fewer cues are attended to (Derryberry and Tucker 1992);
- *motivation and performance*: an increase in emotional intensity causes an increase in performance, up to an optimal point (inverted U-curve Yerkes-Dodson Law). User models which provide qualitative and quantitative feedback to help students think about and reflect on the feedback they have received (Bull, 1997), need to include affective feedback about cognition-emotion states experienced by the user and recorded during the task.

Currently of particular interest to us are the next three points because they are directly related to the information conveyed in the face.

- *intention*: not only are there positive consequences to positive emotions, but there are also positive consequences to negative emotions -- they signal the need for an action to take place in order to maintain or change a given kind of situation or interaction with the environment (Frijda 1986). Facial expressions associated with negative emotional states such as feeling *frustrated* (Klein 1999) and *overwhelmed* are often observed on the faces of computer users at many levels of expertise.
- *categorization and preference*: familiar objects become preferred objects (Zajonc 1984). User models which aim at discovering the user's preferences (Linden et al., 1997), also need to acknowledge and make use of the knowledge that people prefer objects that they have been exposed to, even if they were shown these objects subliminally. Preference can be associated with a facial *smile* and an expression of *content*.
- *communication*: important information in a conversational exchange comes from body language (Birdwhistle 1970), voice prosody, facial expression revealing emotional content (Ekman 1975), and facial displays connected with various aspects of discourse (Chovil 1991).
- *learning*: people are more or less receptive to the information to be learned depending upon their liking of the instructor, or of the visual presentation, or of how feedback is given. For example, given feedback while performing a task, subjects seem to be more receptive to the feedback if an image of their own face is used to give the feedback than if it is someone else's face who gives the feedback (Nass 1998). Emotional intelligence is, furthermore, learnable (Goleman 1995). In particular, people can learn to recognize facial expressions with increased accuracy given appropriate training.

With such a strong interface between affect and cognition on the one hand, and with the increasing versatility of computer agents on the other hand (Maes 1990) (Bradshaw 1997), the attempt to enable our computer tools to acknowledge affective phenomena rather than to remain blind to them appears desirable.

Because emotions are very complex phenomena, however, if we want to attempt to recognize them automatically, it becomes important to acknowledge all the components associated with them. When it comes to pattern recognition, furthermore, the human brain/body/mind complex is a wonder of parallel processing mechanisms (Rumelhart and McClelland 1986; McClelland and Rumelhart 1986). It relies simultaneously on a mix of sensory impressions as well as on previously stored schemata-like scenarios to recognize patterns.

Facial information recognition is no exception to this parallelism. Computer systems successful at recognizing affect from facial information are likely to combine patterns received from different modalities as well. In the following section, we give a brief review of the various components associated with affective phenomena

4.2. Some Relevant Background on Affect Representation

It has been traditionally thought that the interface between cognition and affect happened principally at the level of internal mental representation. If affect were to influence information processing, there would have to be some affective representation (i.e. the subjective experience of emotion) which intertwines with the cognitive representation being processed. Thus the interaction of affect and cognition has typically been studied by focussing on the associative structures that represent both types of elements. Alternatively, it has been considered (Zajonc and Markus, 1984b) that affect and cognition can *both* be represented in multiple ways, *including* in the motor system. This is very observable in the case of affect. While moods and emotions eventually result in cognitive states, they can easily be identified by responses of the motor and visceral system: smiling or frowning faces, embarrassed grins, tapping fingers, queasy stomachs, pounding hearts, or tense shoulders.

Cognition, like affect, may also be represented within the organism's activity. While involved in thinking, problem solving, or recalling, people are often found looking for the answers in the ceiling, scratching their heads for inspiration, stroking their chins, or biting their lips. The suggestion here is that *both* cognition *and* affect can be represented in the motor system. By building tools which can observe and measure the motor systems, we expect to build a rich database of revealing affective and cognitive phenomena, as well as to enrich our interaction with these tools.

As is illustrated in figure 1, emotions are elicited by a combination of events: sensory, cognitive, and biological (not drawn). The actual generation of the basic emotional state – including its autonomic arousal, visceral and muscular activity -- depends upon a number of gating processes, such as attention, existing conflicting emotional states which might interfere with the new one, competing muscular engagement, or cognitive conscious or unconscious suppression.

Emotion generation is associated with three phenomena: autonomic nervous system (ANS) arousal, expression, and subjective experience (Zajonc and Markus 1984b). More recent views emphasize the plasticity of the brain and include mechanisms previously considered as results of emotional arousal (e.g. facial actions, ANS activity and breathing patterns) as sources of arousal as well (Zajonc 1989) (Levenson 1992).

4.3. An Architecture for Affect Processing in Human-Computer Interaction

We propose an architecture for a system which can take as input both mental and physiological components associated with a particular emotion. It is illustrated in figure 2. Physiological components are to be identified and collected from observing the user via receiving sensors with different modalities: **V**isual, **K**inesthetic, and **A**uditory (**V, K, A**). The system is also intended to receive input from Linguistic tools

(L) in the form of linguistic terms for emotion concepts, which describe the subjective experience associated with a particular emotion (Lisetti 1997).

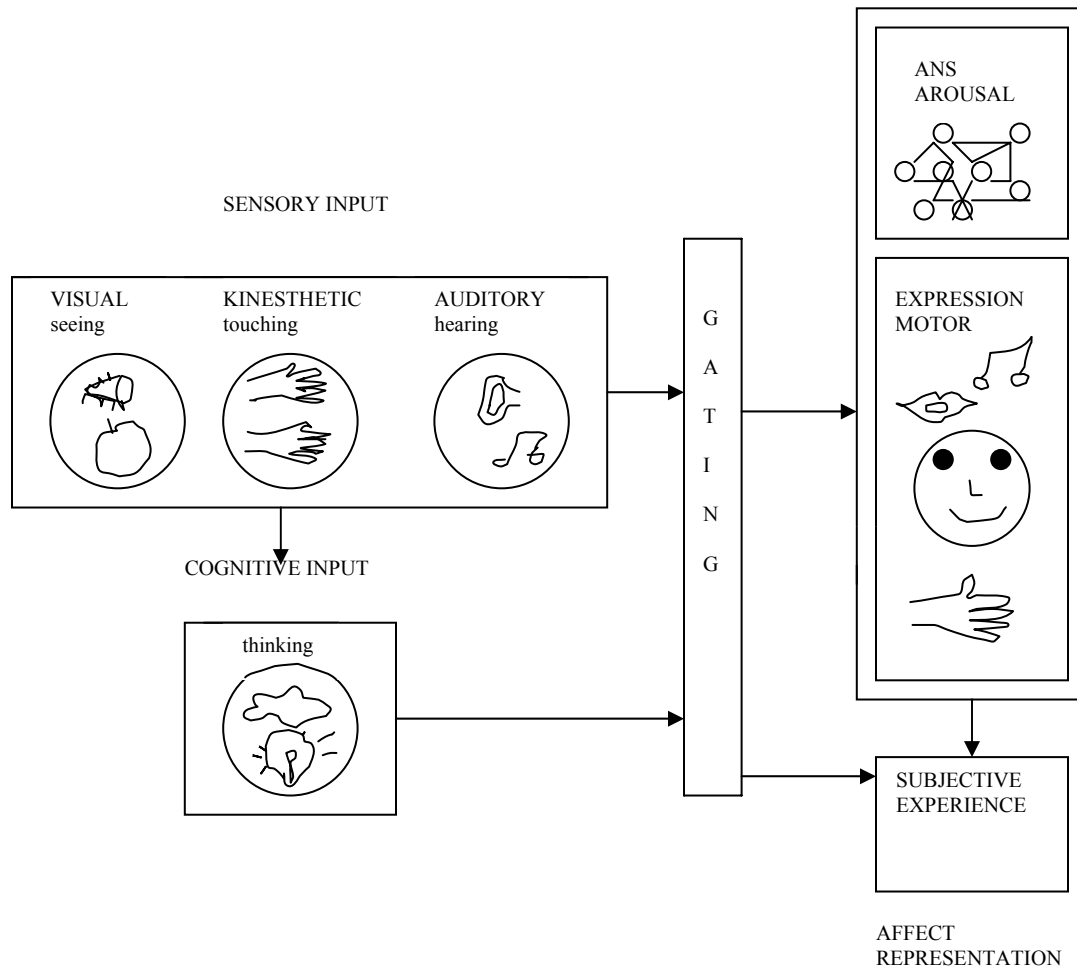


Figure2: Affect Representation

The output of the system is given in the form of a synthesis for the most likely emotion concept corresponding to the sensory observations. This synthesis constitutes a descriptive feed-back to the user about his and her current state, including suggestions as to what next action might be possible to change state. As discussed in section 3, the system is designed to be extended by providing appropriate multimodal feedback to the user depending upon his/her current state. Examples of these adjustments are: changing the interface agent's voice intonation, slowing down the pace of a tutoring session, and selecting the facial expression of an animated agent.

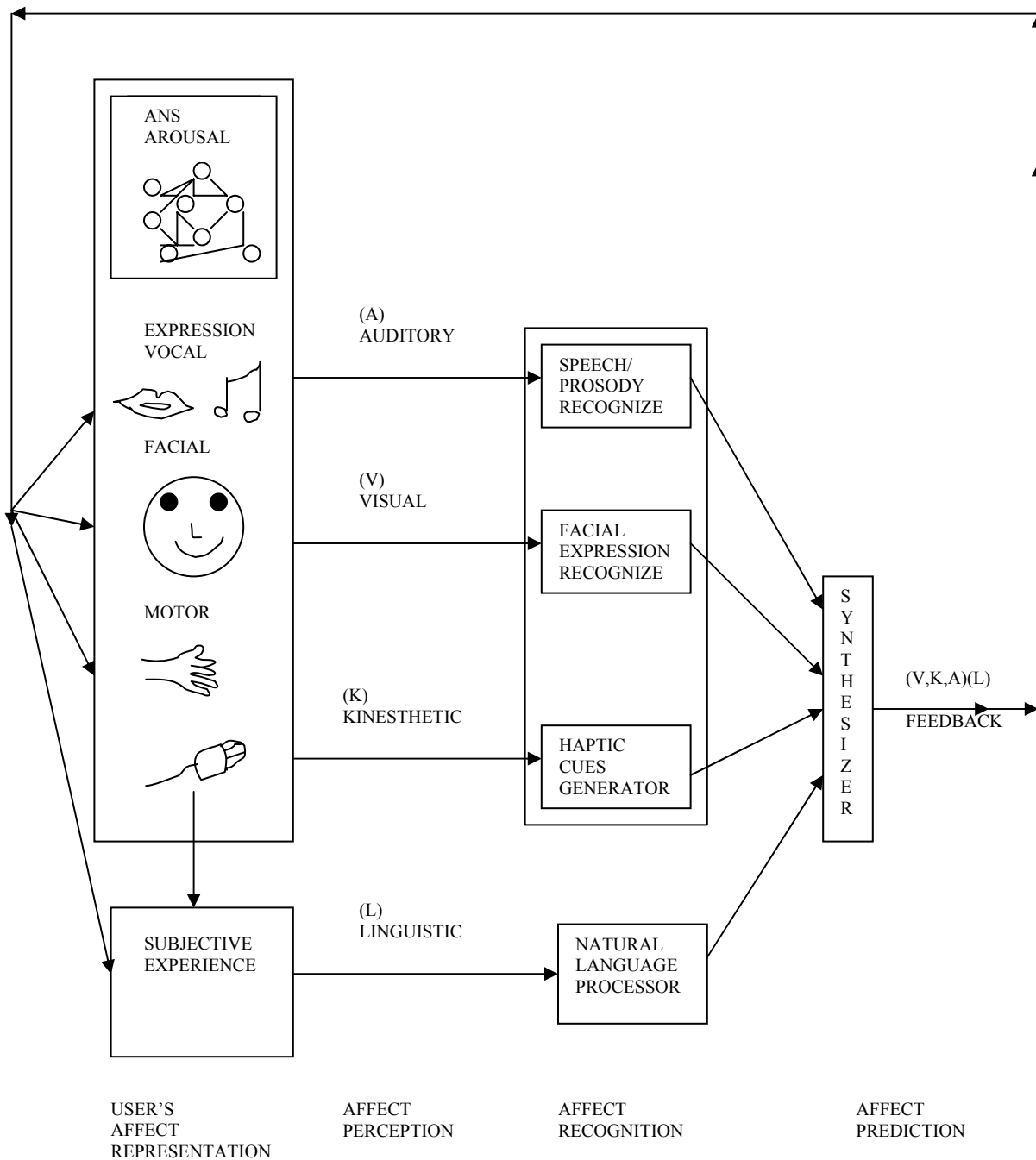


Figure 3: Affect, AI, and HCI

Recent progress in such AI domains as machine vision, speech recognition, and haptic output, indeed make possible the integration of techniques for building intelligent interfaces that reflect the importance of emotion in human intelligence (Picard, 1997). Using the three main sensory systems of machine perception *visual* (**V**), *kinesthetic* (**K**), *auditory* (**A**), and via *natural language processing* (**L**), it has become possible to process computationally:

- *affect perception via:*
 1. (Visual) facial expression (Essa and Pentland 1997), (Essa, Darrell, and Pentland, 1994), (Darrell, Essa and Pentland 1994), (Black and Yacoob 1995, 1995b), (Mase 1991), (Rosenblum, Yacoob,

- and Davis 1994), (Terzopoulos and Waters 1993), and (Kearney and McKenzie 1993), (Yacoob, Lam and Davis 1995), (Yacoob and Davis 1996), (Padgett and Cottrell 1997, 1998);
2. **(Auditory)** vocal emotion (Murray and Arnott 1993) (Slaney and McRoberts 1998); prosody analysis (Gardner and Essa 1997);
 3. **(Kinesthetic)** advanced bio-feedback via wearable computers capable of categorizing low-level signals such as electromyogram, galvanic skin response, and blood pressure (Picard 1997);
 4. **(Linguistic)** spoken or written natural language (O'Rourke and Ortony 1994).

Perception here needs to be understood in terms of measuring observable behaviors of the motor system which correspond with high *probabilities* to one emotion experienced subjectively. Interpretation of the measured percepts using contextual and possibly multimodal information is treated separately in our architecture.

- *affect generation and production via:*
 1. cognitive generation (Elliott 1994) (Frijda and Swagerman 1987), (Dyer, 1987);
 2. a combination of cognitive and sensory generation with schemata-like representations (Lisetti 1997).

Generation is needed in order to enable a computer system to have some level of understanding of what is it like for the user to have emotional states. Generating states for a computer system with similar properties and/or functions to human emotions, is here considered to be one of the first steps in building such an understanding.

- *affect expression via:*
 1. **(A)** vocal prosody (Cahn 1990), (Fleming 1997);
 2. **(V)** expressive believable agents (Bates 1994);
 3. **(L)** semantic descriptions as in computational scripts (Lisetti 1997); speech dialog with facial displays (Nagao and Takeuchi 1994).

The agent can adapt its interface appropriately by adjusting its output (expressions, gestures, voice inflection) depending upon the recognized user's state and the personality of the user. In our architecture, the agent does so after the synthesizer, described in a forthcoming article (Lisetti 1999), has generated an appropriate computer state associated with the observed user's state.

In the future, any serious effort to work with emotions, affect and computer environments will also need to include:

- *affect interpretation* (of the measured percepts) via:
 1. inclusion and integration of interpretive contextual information through person-based models (Winograd 1998);
 2. multimodal integration – for example integrating acoustic and visual information – which sometimes offers promising improvements (Stork and Henneke 1996). There is furthermore increasing research in the area of multimodal integration (Blattner and Glinert 1996), (Sharma, Pavlovic, and Huang 1998). Multimodal integration is highly relevant to affect recognition given the multi-component nature of emotion (described earlier). It is known, for example, that emotions are closely related (if not associated) with cognition (Leventhal and Scherer 1987), that facial expressions can generate emotion physiology (Ekman 1993), and that facial expressions are associated with emotions (Ekman 1975).

Interpretation will need to be carried out carefully. Human perception of visual emotional cues is performed by different brain mechanisms compared with the perception of auditory emotional cues (Etcoff 1989). Furthermore within the visual modality for example, it is likely that facial cues are mediated differently than nonfacial visual cues (Rolls 1992). Whether computer perception will benefit from such differentiated subsystems or whether it will need multimodal integration for best performance needs to be studied.

Finally, it is also important to note, that while healthy humans have all these different emotional components (perception, interpretation, expression, generation), computers can be implemented with one or more of them -- for example, with recognition capabilities but without generation capabilities -- and as a result exhibit differing performance. Emotion is a construct that subsumes heterogeneous group of processes, and there are many ways that emotion can be parsed at the human level (Davidson 1994). Human perception of emotional cues appears to be implemented in different circuitry from the production of emotion responses. The same distinction might need to be true for computer perception and production of emotions, and may even be desirable if only certain system functionalities are sought for.

In general, an interface to *recognize* and *express* affective states can be attractive and can improve the communication from the computer to the user. It involves building interfaces which can:

- render the computer more trustable, believable, and human-like in its interaction; “emotion is one of the primary means to achieve this believability, this *illusion* of life, because it helps us know that characters really care about what happens in the world, that they truly have desires” (Bates 1994);
- adapt itself to induce various emotions which might be more desirable given the current activity;
- record and remember the user's states during an interaction;
- change the pace of a tutoring session based upon the monitored user's cognitive and emotional states (i.e. bored, overwhelmed, frustrated, etc.) in a similar manner as when COACH adjusts its feedback depending on the user's level of expertise (Selker 1994);
- guide the user to avoid cognitive and or emotional paths where [s/he] gets blocked during a task; whereas humans often mostly rely on intuitive, sometimes unconscious, processing to perceive and make sense of others' subjective experience, there might be ways to analyze and recognize cues to people's subjective experience in a more formal and automatic manner (Pasztor, 1998);
- implicitly adapt its interface using multimodal devices (expression, posture, vocal inflection) to provide adaptive feedback.

From the user to the computer, it might be desirable to:

- give computer agents awareness of what emotional state the user might be feeling so that inferences can be drawn about the motivations implied in them (Nagao and Takeuchi 1994);
- motivate agents to initiate actions (i.e. search and retrieve items) using explicitly-set agent competence levels. The competence level can also be evaluated and updated in terms of the accuracy of its predictions made from observations of the user (Maes 1994);
- explicitly change some aspects of the agent's interface depending on user's state; this option is available for people who prefer direct manipulation (i.e. the user always remains in control of any changes in the interface and directly alters any aspect of the interface) to the more recent interface agents who are there to assist and reduce the user's information overload (Maes 1994) (Shneiderman and Maes 1997).

From a broader AI perspective, *affect simulation and generation* might lead to the development of computational models of emotion in order to:

- test emotion theories by providing an artificial environment for exploring the nature and development of emotional intelligence;
- learn from naive psychology: explain, understand, predict behaviors of others, and build user models;
- improve AI process-control: control of cognition, attention, and action;
- choose various planning algorithms under different time pressures signaled by intensity of artificial motivational state;
- develop pro-active intelligent agents: self-motivated software agents or robots with motivational states (Sloman 1990);
- self-adjust the agent's commitment to an ongoing activity based upon valence of its current state (negative: slow down waste of energy and reevaluate context, positive: continue in the same direction).

5. Multidisciplinary Issues and Research Questions

We now address some of the current questions that need to be addressed from an interdisciplinary approach, for real progress to be achieved in automatic facial expression recognition so that it can find some beneficial uses in human-computer interaction and psychological research. Even though they are somewhat separated out, all these issues and questions are intended for cognitive scientists, psychologists, HCI and AI researchers interested in facial expression recognition.

5.1. Cognitive Science and Psychological Issues

- *Individual differences:* There are indications that there exist differences in terms of the speed, magnitude, and duration of facial expressions across individuals. There are also variations in which expression of emotion occurs in response to a particular event. The question that needs to be answered, both for psychology and artificial intelligence, is whether these differences are consistent across emotions or situations or over time. If psychologists can answer this question, it would benefit AI researchers in computer vision because automatic pattern recognition has often given exceptional results recognizing patterns in a user-dependent manner. This has been the case for speech recognition and is expected to be the case for facial expressions. In addition, the answer to this question would also provide valuable cues to begin to address contextual and situational information, such that it will become easier to include in a person-based model, which situations and/or stimuli elicit which expression of emotion during a human-computer interaction.
- *Emotion without facial action:* The question of whether facial activity is a necessary part of any emotional experience also is of interest for human-computer environment aiming at recognizing their user's affective states. It would be important to know, for example, under what circumstances and with what kind of people might there be physiological changes relevant to emotion and the subjective experience of emotion, with no evidence of visible expression or non-visible electromyographic facial activity.
- *Facial Display without emotion:* Even if we consider facial displays as only affecting our intention, as opposed to reflecting our emotional inner life, by signaling intentions to others so that they can respond appropriately to us (Chovil 1991b; Fridlund 1994), the controversial assumption that people may relate to computers as if they were social actors (Reeves and Nass 1996), would suggest *a fortiori* the importance of having computer tools that recognize and respond appropriately to these signals of intent.
- *Personality trait and expression:* One question that would benefit human-computer interaction research is that of finding out whether personality traits, moods and psychopathology have facial markers or whether they are second-order inferences drawn from occurrences of facial expression of emotion. For example, depression is often revealed by frequent expressions of *sadness*, and Alzheimer's patient often portray a certain constant *tormented* look on their faces. Similarly, it might be possible for personality traits to hold a specific set of favored expressions, which would enable interface agent to adapt their interface to the personality of the user.
- *Volitional and non-volitional expression:* There is evidence from neuroscience that different neural pathways for volitional and non-volitional facial expressions are activated (Buck 1982). An important question for HCI is to know whether there are differences in the expression itself? For example, it seems that Duchenne's smile, hypothesized to occur with spontaneously occurring enjoyment (as opposed to other voluntary smiles) (Ekman et. al 1990), involves not only muscle movements pulling the lip corners up (as in all smiles), but also the activity in the muscle that orbits the eye. Being able to differentiate between these different smiles with reasonable accuracy would prove valuable for HCI (as well as for psychologists collecting data on this question). By focussing the computer vision algorithm on the eye area of the digital image, it might become possible to reach such accuracy.

- *Different types of smiles*: Do different positive emotions such as *amusement*, *contentment*, and *relief* have distinctive forms of smiling, or do all positive emotions share one signal? The latter would imply that specific distinct positive emotions would need to be inferred only from other behavioral or contextual cues rather than from the facial information contained in a smile (Ekman 1992). These issues are again highly relevant for building interface agents capable of recognizing affective phenomena. The question of knowing whether there are distinctive forms of non-enjoyment smiles is also important for interface agents. Are there smiles associated with *compliance*, *embarrassment*, *grin-and-bear-it* kind of smiles?

5.2. Artificial Intelligence and Computer Vision Issues

The previous section has pointed at the most recent progress in the field of automatic facial expression recognition, and demonstrated that placing facial information within human-computer communication is possible. A few more questions need to be answered before it becomes a reality.

- There is a strong need to generate the equivalent of the FERET database of facial images for face recognition specifically for facial expression recognition. This will enable researchers to train, test, and compare their systems as well as to compare them with human performance. Psychologists, HCI and AI researchers need to merge efforts and share results.
- Real-time processing necessitates the expression recognizer to be integrated with a visual front-end responsible for tracking the face, capturing the facial image and segmenting it.
- While facial expressions are good pointers to inner emotional states, studies confirm the existence of distinct neurological systems in humans for dealing with *volitional as opposed to spontaneous* expression of emotions: apparently, like other primates, our ability to convey emotions deliberately to others proceeds along a separate track from our spontaneous and involuntary experience and expression of emotions (Buck 1982). These results need to be kept in mind when we speak about automatic recognition of facial expressions: the recognized expressions could be volitional in some cases, and spontaneous in others!
- Different *hair styles* in recognition of unfamiliar faces renders recognition more difficult (Bruce, Burton, and Hancock 1995). For automatic recognition, hair around the face (not facial like mustache and beards) are usually removed by cropping the images so as to reduce the “noise” introduced by different hair styles for a particular person.
- Most facial expressions are very *brief*. To complicate matters, there exist micro expressions, extremely rapid (lasting only a fraction of a second), and typically missed by observers. These micro expressions can usually reveal emotions a person is trying to conceal. Macro expressions, on the other hand, typically last between 2 to 3 seconds, and rarely more than 10 seconds. Automatic recognition has been concerned principally thus far with macro expressions.
- Expressions have an *offset, a peak, and a decline*. A series of consecutive picture frames would need to be taken, trained and tested in order to recognize what the entire expression really was. It is important to note here that the system would not be asked at this point to perform better than humans do in their own facial expression recognition process.
- The issue of the system’s performance brings about the need to develop methods to compare automatic with human performance. Any system would need to be validated by comparing its output with the judgements of human experts in recognizing expressions, as well as with non-expert humans.
- *User-dependent recognition*: It is also very feasible to work with user-specific expressions corresponding to some of the user's most frequent affective-cognitive states experienced while interacting with the environment. People typically experience reoccurring patterns of states. While the variety of emotional states that one can experience is infinite, during everyday life there is usually a

limited number of states that reoccur frequently. It is then possible to identify these regularities and patterns by “looking over the user’s shoulder” while [s/]he is involved in a task, in a similar manner as Maes suggests (Maes 1994). Such identification will enable the fine-tuning of facial expression recognition not only by analyzing expressions but also the contexts in which they occurred.

- Integrating static image recognition and motion recognition seems to be promising.
- Real time recognition of facial expressions will need to rely on a good face tracker, a difficult vision problem.
- Good interpretation of facial expressions will need to tune in to:
 1. different users
 2. different contexts
 3. different cultures
- To integrate different modalities (V, K, A) to double check and tune-in prediction in a similar way to Stork and Henneke (1996) seems needed.

5.3. Human-Computer Interaction and Ethical Issues

Some of these issues have been addressed at the Panel on Affect and Emotion in the User Interface at the 1998 International Conference on Intelligent Interfaces (Hayes-Roth et. al 1998). We have included here some of the suggestions and comments discussed by Lisetti during the panel discussion.

- *Anthropomorphism: Is anthropomorphism of human-computer interfaces to be desired?*

Humanizing computer interfaces has two aspects: making interfaces easier to use and making interfaces more human-like. As mentioned earlier, there has been an effort to introduce communicative signals with speech dialogue interfaces (Nagao and Takeuchi 1994). The aim of such a system is to improve human-computer dialogue by introducing human-like behavior into a dialogue system. The benefits are intended to reduce the stress experienced by users and to lower the complexity associated with understanding the system status.

As pointed out by other studies (Walker, Sproull, and Subramani 1994), however, using a human face in an interface need not *necessarily* give the expected result. The goal of HCI with synthetic faces may not be to give a computer a human face, but rather to establish *when* a face-like interface is appropriate. Specific attention to gender effects, personality types and age groups need to be addressed.

The question of *whose* face should appear has also began to be studied (Nass et. al. 1998). There seems to be indication, for example, that the effects of receiving negative evaluation from audio-visual image of oneself on a computer screen are clearly different from that of receiving someone else’s. Receiving the self-image seems to lead to a higher sense of responsibility for the evaluation for the user, and to an increase in the perception of the validity of the negative feedback.

Additional questions such as *which expressions* should the face portray and in *what contexts* and *applications* should the face appear will also need to be answered.

- *Emoting: Do people actually emote towards computers? If so who does it, how do they do it and when do they do it?*

Social psychologists, sociologists, and communication researchers are invited to join human-computer interaction researchers to set up studies to answer the many facets of these questions. By taking a more comprehensive perspective on the role of emotions and their interaction with cognition (as discussed earlier in this current article), it is anticipated that the answer to the first question will be ‘yes’. Furthermore, by expanding the current human-computer interaction with the “personal” computer, to the emerging

“interspace that is inhabited by multiple people, workstations, servers, and other devices in a complex web of interactions” (Winograd 1997), to the future ubiquitous computing that is likely to add many more possible contexts to human-computer interactions, the answer to the first question again seems to be ‘yes’.

It will take some effort to establish the relations involved in the second question. . People can be thought to emote to computers if they experience feelings and emotions toward computers, or while they are interacting with computers. Emotions such as fear, and anxiety often keeps novices away from computing technology, which they perceive as intimidating. Techno-phobic people can also experience fear and anxiety, or even disgust and hate, toward computing technology. The majority of users, on the other hand, often experience emotions such as confusion, frustration, understanding, or satisfaction, at one time or the other while interacting with computing technology. In order to document these cases better, for example, videos of students interacting with a computer tutor could be recorded and analyzed in terms of facial expressions to draw statistics about the tutor’s effectiveness. Interactive kiosks could be equipped with facial expression recognizers to determine *who*, *when*, and *how* their users emote. Difference in personalities and different contexts will need to be studied again in terms of how they affect the interaction.

- *Privacy: Will people feel exposed under the scrutiny of machines that monitor and interpret their emotional states?*

Different personality types matter. For people who remain mostly neutral, who do not experience nor express emotions while using computing technology (in many cases computing experts), the issue of feeling scrutinized by machines almost vanishes. For people who do, however, experience and express emotions while interacting with computer technology, scrutiny becomes a potential concern. That is why emotional information raises *privacy issues*, in the same way as many other kinds of information, such as medical records and the like, raise privacy questions. What really matters is how new technology is interpreted and by whom: interpretation of users’ emotional states could potentially be used to mind-read and manipulate people; or it can be used to assist users in their daily tasks; or maybe even to increase and enhance one's emotional awareness. If employed properly, these new technologies that deal with affective information could potentially enhance quality of life (as discussed in the application section of this article).

For example new technology on facial expression recognition could improve human-human communication by teaching and helping people to become more aware of their own facial expressions and of those of others: whether they are a facial withholder (not expressing emotions), an unwitting expressor (expressing emotions without knowing it), or a substitute expressor (expressing one emotion but while experiencing another). Another example is found with girls interacting with computer games who have been found to enjoy entering secrets in diaries, and to interact with software characters which exhibit strong social and emotional skills. Ultimately, if users have control about when and with whom they can share this private information, the notion of scrutiny vanishes. The issue that remains then, is whether or not the individual user values the capability of this technology to be aware of his/her emotional states and what it does with it.

- *Manipulation: Will people feel emotionally manipulated by machines that feign empathic responses?*

It is possible that people will feel emotionally manipulated by machines that are capable of feigning empathic responses. Many people, however, already feel emotionally manipulated by the unnatural non-empathetic responses of the increasingly numerous machines they have to interact with (often not by choice), as studied in Norman's book *Things that Make us Smart* (Norman 1993). We need to move from a machine-centered orientation to life toward a person-centered point of view (Winograd 1998). Technology should serve us (not vice and versa), and having more humane technology might become important since information technology will continue to affect increasingly more aspects of life.

There also exists the need to consider the potential risk of finding people’s natural abilities to feel and express emotions eroded, due to prolonged exposure to computer interfaces which cannot be aware – not to mention to respond to – humans’ rich affective and social world. In some odd way, many people today consider existing computer interfaces as cold, rude, stubborn unfriendly, and unforgiving, and feel

negatively emotionally manipulated by a machine-centered orientation to life which emphasizes the needs of technology over those of people.

- *Empathic machines: Will people tolerate and appreciate empathic machines in a greater number and variety of roles than are currently filled by their affectless predecessors?*

One of the important challenges for the user interface designer is to consider the "third person" perspective, i.e. the user's perspective. In order to acquire this knowledge, designers will need to develop skills to understand and predict the user's experience in order to guide their design appropriately (Winograd 1996). From that perspective, designers will need to acknowledge social factors such as differences in personalities, gender, or social context. Incidentally, while personality types do matter in general, politeness seems to be well received by all in most contexts. As studied by Reeves and Nass, politeness even seems to be expected from computing technology (Reeves and Nass 1996).

We also need to realize the strong interaction between affect and cognition: memory and retrieval may be most important, but also attention, decision-making, preferences, etc. This means that many of the tasks involved in a work setting that are considered "high level" cognition are actually very linked with affect and various emotional phenomena. Exploring the range of these could very well lead to intelligent agents and interfaces with capabilities for affective "awareness" and affective expression, which could enhance the user's recollection, understanding, and overall satisfaction by subtly responding to the user's current state.

Given what Reeves and Nass' results seem to imply (Reeves and Nass 1996), some people might be very receptive to computer and media responses, and these people interpret computer responses socially. Designing interfaces that are polite, that can adapt to various types of user's personalities, thus appear the more appealing. As of today, too many technological products arouse states of frustration and dissatisfaction in the user, sometimes even way above their threshold level of frustration tolerance (Norman 1993). This seems to reflect a certain lack of consideration for user's feeling and a lack of responsibility at a societal level. While enabling computer systems to recognize and respond to user's emotions is not a panacea, a little compassion, apology or understanding might sometimes alleviate the impact on the user's experience of some undesired flows in functionality or design.

- *Responsibility: Is there danger in creating very believable, affective characters and interfaces, and their creators must be careful and responsible?*

In human-human communication, *how* something is said is often more important than *what* is being said. In particular, believability of the speaker is often derived from the congruency of signals coming across the various communication channels (mostly kinesthetic, visual, and auditory). If the message perceived is congruent over all channels (for example if the facial expression, body posture, and vocal intonation all convey the same modal, and sometimes unconscious, information) then the message is most likely to be believed. Hence, if computer interfaces, characters, and avatars are endowed with communication capabilities across multiple modalities, they could potentially become more believable. It then becomes conceivable that they could be used to deceive the user, in a similar manner that humans can deceive others better once they are believed and trusted by those they are trying to fool. The responsibility of the designer and conceiver of such applications would therefore be crucial.

Another danger in creating computer interfaces with good believability and affective capabilities is associated with how these could possibly raise the user's expectations concerning the remaining capabilities of the system. If, for example, a computer can express in a very natural, articulate, and compassionate manner that they did not understand the user's request, the user might start to "project" a human intelligence onto the system, and expect a set of system behaviors equivalent to that humans. If however, only the interface of the system has been enhanced, but its performance, functionality and overall "intelligence" have not been improved accordingly, then the user might become confused and/or disappointed by the system. The responsibility of the designer would remain to indicate to the user that the overall intelligence of the system is still limited, even though it can converse and interact in a very natural manner.

Not only is overall system performance a related issue, but so is just the interface performance. If for example, the system has poor performance on facial expression recognition, then, depending upon the application, it can lead to major miscommunication. But if its recognition rate is good however, and because once again *how* something is said is often more important than *what* is being said, it could enhance communication tremendously by adding affective modal communication, which is often the only source of miscommunication in plain non-modal textual communication.

So, yes, creators of affective interfaces must be careful and responsible, in a similar manner that creators of highly functional but very unaware interfaces must be responsible in keeping computer users sane and satisfied.

6. Conclusion

In this paper, we have emphasized the primordial role of emotions on 'high-level' cognitive processes. We discussed a possible architecture for a system able to acknowledge the interface between affect and cognition and to provide multimodal intelligent feedback. We also pointed out some of the issues involved in automatic facial expression recognition from an interdisciplinary perspective between HCI, AI, and Cognitive Science. We hope that this discussion will motivate further research in all these areas, both individually and from an interdisciplinary perspective.

References

- Bartlett, M.S. and Hager, J.C., and Ekman, P. and Sejnowski, T. J. 1999. "Measuring facial expressions by computer image analysis". *Psychophysiology* 36(2):253-263.
- Bates, J. 1994. "The role of emotions in believable agents". *Communications of the ACM* 37(7): 122-125.
- Birdwhistle. 1970. *Kinesics and Context: Essays on Body Motion and Communication*. Philadelphia, PA:University of Pennsylvania Press.
- Black, M. J. and Yacoob, Y. 1995. "Recognizing facial expressions under rigid and non-rigid facial motions". In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, IEEE Press, 12-17.
- Black, M. and Yacoob, Y. 1997. "Recognizing facial expressions in image sequences using local parametric models of image motion". *International Journal of Computer Vision* 25(1): 23-48.
- Blattner, M. and Glinert, E. 1996. "Multimodal integration". In *IEEE MultiMedia*, Winter 1996, 14-24.
- Bower, G. 1981. "Mood and memory". *American Psychologist* 36(2): 129-148.
- Bradshaw, J. 1997. "An Introduction to Software Agents". In J. Bradshaw (ed), *Software Agents*. Cambridge, MA: AAAI Press/MIT Press, 3-46.
- Bruce, V., Burton, M., and Hancock, P. 1995. "Missing dimensions of distinctiveness". In Valente, T. (ed), *Cognitive and Computational Aspects of Face Recognition*. New York: Routledge, 138-158.
- Buck, R. 1982. "A theory of spontaneous and symbolic expression: Implications for facial laterization". Presented at the Meeting of the International Neuropsychological Society, Pittsburgh.
- Bull, S. 1997. "See yourself write: A simple student model to make students think". In *User-Modeling: Proceedings of the Sixth International Conference (UM'97)*. New-York: Springer, 315-326.
- Cacioppo, J.T., Tassinary, L.G., and Fridlund, A.F. 1990. "The skeletomotor system". In J.T. Cacioppo and L.G. Tassinary (eds), *Principles of Psychophysiology: Physical, Social, and Inferential Elements*. New-York: Cambridge University Press, 325-384.
- Cahn, J. 1990. "The generation of affect in synthesized speech". *Journal of the American Voice I/O Society* 8: 1-19.
- Camras, L. 1992. "Early development of emotional expression". In K.T. Strongman (ed), *International Review of Studies on Emotion*, Volume 1. New York: Wiley, 16-36.
- Chauvin, Y. and Rumelhart, D.E. 1995. *Backpropagation: Theory, Architectures, and Applications*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Chovil, N. 1991a. "Discourse-oriented facial displays in conversation". *Research on Language and Social Interaction* 25: 163-194.
- Chovil, N. 1991b. "Social determinants of facial displays". *Journal of Nonverbal Behavior* 15: 141-154.
- Cottrell, G. and Metcalfe, J. 1991. "EMPATH: Face, emotion, and gender recognition using holons". In R. P. Lippman, J. Moody, and D. S. Touretzky (eds.), *Advances in Neural Information Processing Systems*, volume 3. San Mateo, CA: Morgan Kaufmann Publishers, 564-571.
- Damasio, A. 1994. *Descartes' Error*. New York: Avon Books.
- Darrell, T., Essa, I., and Pentland, A. 1994. "Task-specific gesture analysis in real-time using interpolated views". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(12): 1236-1242.
- Darwin, C. 1872. *The Expression of Emotions in Man and Animals*. London: Murray [Reprinted Chicago: University of Chicago Press, 1965].
- Davidson, R. 1994. "Honoring biology in the study of affective style". In P. Ekman and R. Davidson (eds), *The Nature of Emotion: Fundamental Questions*. New York: Oxford University Press, 321-328.
- Derryberry, D. and Tucker, D. 1992. "Neural mechanisms of emotion". *Journal of Consulting and Clinical Psychology* 60(3): 329-337.
- Dror, I., Florer, F., Rios, D., and Zagaeski, M. 1996. "Using artificial bat sonar neural networks for complex pattern recognition: Recognizing faces and the speed of a moving target". *Biological Cybernetics* 74: 331-338.
- Dyer, M. 1987. "Emotions and their computations: Three computer models". *Cognition and Emotion* 1(3): 323-347.
- Ehrlich, S., Schiano, D., Sheridan, K., Beck, D., and Pinto, J. 1998. "Facing the issues: Methods matter". *Abstracts of the Psychonomic Society, 39th Annual Meeting, Volume 3*, 397.
- Ekman, P. 1982. "Methods for measuring facial action". In K. Scherer and P. Ekman (eds), *Handbook of Methods in Nonverbal Behavior Research*. Cambridge, MA: Cambridge University Press, 45-135.
- Ekman, P. 1992. "Facial expressions of emotion: New findings, new questions". *Psychological Science* 3: 34-38.
- Ekman, P. 1993. "Facial expression and emotion". *American Psychologist* 48: 384-392.
- Ekman, P. and Davidson, R. 1993. "Voluntary smiling changes regional brain activity". *Psychological Science* 4: 342-345.
- Ekman, P. and Davidson, R. 1994. *The Nature of Emotion: Fundamental Questions*. New York: Oxford University Press.
- Ekman, P., Davidson, R., and Friesen, W. 1990. "The Duchenne smile: Emotional expression and brain physiology II". *Journal of Personality and Social Psychology* 58(2): 342-353.
- Ekman, P. and Friesen, W. 1975. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions*. Englewood Cliffs, NJ: Prentice Hall.
- Ekman, P. and Friesen, W. 1976. "Measuring facial movement". *Journal of Environmental Psychology and Nonverbal Behavior* 1(1): 56-75.
- Ekman, P., Friesen, W., and O'Sullivan, M. 1988. "Smiles when lying". *Journal of Personality and Social Psychology* 54: 414-420.
- Ekman, P., Friesen, W., and Tomkins, S. 1971. "Facial affect scoring technique: A first validity study". *Semiotica* 3: 37-58.
- Ekman, P., Huang, T., Sejnowski, T., and Hager, J. 1993. *Final Report to NSF of the Planning Workshop on Facial Expression Understanding*.
- Ekman, P. and Rosenberg, E. 1997. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System*. New York: Oxford University Press.
- Elliott, C. 1994. "Components of two-way emotion communication between humans and computers using broad, rudimentary model of affect and personality". *Cognitive Studies: Bulletin of the Japanese Cognitive Society* 1(2): 16-30.
- Essa, I., Darrell, T., and Pentland, A. 1994. "Tracking facial motion". In *Proceedings of the IEEE Workshop on Nonrigid and Articulate Motion*, 36-42.
- Essa, I. and Pentland, A. 1997. "Coding, analysis, interpretation and recognition of facial expressions". *IEEE Pattern Analysis and Machine Intelligence* 19(7): 757-763.
- Etcoff, N. 1989. "Asymmetries in recognition of emotion". In F. Boller and J. Grafman (eds), *Handbook of Neuropsychology* 3: 363-382.
- Etcoff, N. and Magee, J. 1992. "Categorical perception of facial expressions". *Cognition* 44: 227-240.

- Fleming, M.A. 1997. *Neural Network Model of Micro and Macroprosody*. Ph.D. dissertation., Dept. of Psychology, Stanford University.
- Fleming, M.A. and Cottrell, G. 1990. "Categorization of faces using unsupervised feature extraction". In *Proceedings of the International Joint Conference on Neural Networks*, 65-70.
- Fridlund, A.J. 1994. *Human Facial Expression: An Evolutionary View*. San Diego: Academic Press.
- Fried, R. 1987. "Relaxation with biofeedback-assisted guided imagery: The importance of breathing rate as an index of hypoarousal". *Biofeedback and Self-Regulation* 12: 273-279.
- Fried, R. 1990. *The Breath Connection*. New-York: Plenum.
- Frijda, N. 1986. *The Emotions*. New-York: Cambridge University Press.
- Frijda, N. and Swagerman J. 1987. "Can computers feel? Theory and design of an emotional system". *Cognition and Emotion* 1(3): 235-257.
- Gardner, A. and Essa, I. 1997. "Prosody analysis for speaker affect determination". In *Proceedings of the Workshop on Perceptual User Interface*, 45-46.
- Goleman, D. 1995. *Emotional Intelligence*. New-York: Bantam Books.
- Gross, J. J., John, O. P., and J. M. Richards, (in press). "The dissociation of emotion expression from emotion experience: A personality perspective". *Personality and Social Psychology Bulletin*.
- Hager, J. C. 1985. "A comparison of units for visually measuring facial action". *Behavior Research Methods, Instruments and Computers* 17: 450-468.
- Hayes-Roth, B., Ball, G., Lisetti, C., Picard, R., and Stern, A. 1998. "Panel on affect and emotion in the user interface". In *Proceedings of the 1998 International Conference on Intelligent User Interfaces (IUI'98)*. New York: ACM Press, 91-94.
- Hiroshi, I. and Ullmer, B. 1997. "Tangible bits: Towards seamless interfaces between people, bits, and atoms". In *Proceedings of the International Conference on Human Factors in Computing Systems (CHI'97)*. New-York: ACM Press, 234-241.
- Izard, C.E. 1979. "The maximally discriminative facial movement coding system (MAX)". Unpublished manuscript [Available from Instructional Resource Center, University of Delaware].
- Jameson, A., Paris, C., and Tasso, C. 1997. *User Modeling: Proceedings of the Sixth International Conference (UM'97)*. New-York: Springer Wien.
- Kearney, G. and McKenzie, S. 1993. "Machine interpretation of emotion: Design of a memory-based expert system for interpreting facial expressions in terms of signaled emotions". *Cognitive Science* 17: 589-622.
- Klein, J. 1999. *Computer Response to User Frustration*. Ph.D. Dissertation, MIT Media Arts and Sciences.
- Kung, S. and Hwang, J. 1998. "Neural networks for intelligent multimedia processing". *Proceedings of the IEEE* 86(6): 1244-1272.
- Ledoux, J. 1992. "Brain mechanisms of emotion and emotional learning". *Current Opinion in Neurobiology* 2: 191-197.
- Levenson, R. 1992. "Autonomic nervous system differences among emotions". *Psychological Science* 3(1): 23-27.
- Levenson, R., Ekman, P., and Friesen, W. 1990. "Voluntary facial action generates emotion-specific autonomic nervous system activity". *Psychophysiology*, 27(4): 363-383.
- Leventhal, H. and Scherer, K. 1987. "The relationship of emotion to cognition: A functional approach to a semantic controversy". *Cognition and Emotion* 1 (1): 3-28.
- Linden, G., Hanks, S., and Lesh, N. 1997. "Interactive assessment of user preference models". In Jameson et al. (eds), *User-Modeling: Proceedings of the Sixth International Conference (UM'97)*. New-York, NY: Springer Wien, 67-78.
- Lisetti, C. L. 1997. "Motives for intelligent agents: Computational scripts for emotion concepts". In Grahne, G. (ed), *Proceedings of the Sixth Scandinavian Conference on Artificial Intelligence (SCAI'97)*. Amsterdam, Netherlands: IOS Press Frontiers in Artificial Intelligence and Applications, 59-70.
- Lisetti, C. L. 1999. "A user-model of cognition-emotion". In *Proceedings of the User Modeling (UM'99) Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*, 25-35.
- Lisetti, C. L. and Rumelhart, D. 1998. "Facial expression recognition using a neural network". In *Proceedings of the 1998 International Florida Artificial Intelligence Research Symposium Conference (FLAIRS'98)*. Menlo Park, CA: AAAI Press, 328-332.
- McClelland and Rumelhart, D. 1986. *Parallel Distributed Processing Explorations in the Microstructures of Cognition Volume 2: Psychological and Biological Models*. Cambridge, MA: MIT Press.
- Maes, P. 1990. *Designing Autonomous Agents*. Cambridge, MA: MIT Press.

Maes, P. 1994. "Agents that reduce work and information overload". *Communications of the ACM* 37(7): 31-40.

Mase, K. 1991. "Recognition of facial expressions from optical flow". *IEICE Transactions (Special Issue on Computer Vision and its Applications)* 74(10): 3474-3483.

Mozer, M. 1998. "The neural network house: An environment that adapts to its inhabitants". In *Working Notes of the 1998 AAAI Spring Symposium Series on Intelligent Environments* (= Technical Report SS-98-02). Menlo Park, CA: AAAI Press, 110-114.

Murray, I. and Arnott, J. 1993. "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion". *Journal of the Acoustical Society of America* 93(2): 1097-1108.

Nagao, K. and Takeuchi, A. 1994. "Speech dialogue with facial displays: Multimodal human-computer conversation". In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, 102-109.

Nass, C., Kim, E., and Lee, E. 1998. "When my face is the interface: Experimental comparison of interacting with one's own face or someone else's face". In *Proceedings of the International Conference on Human Factors in Computing Systems (CHI'98)*. New York, NY: ACM Press, 148-154.

Norman, D. 1993. *Things That Make Us Smart: Defending the Human Attributes in the Age of the Machine*. Reading, MA: Addison-Wesley.

Norman, D. 1998. *The Invisible Computer: Why Good Products Fail, Why the Personal Computer Is So Complex, and How To Do It Right*. Cambridge, MA: MIT Press.

O'Rourke, P. and Ortony, A. 1994. "Explaining emotions". *Cognitive Science* 18(2): 283-323.

Padgett C. and Cottrell, G. 1997. "Representing face images for classifying emotions". In Jordan, M. I.; Mozer, M. C.; Petsche, T. (eds), *Advances in Neural Information Processing Systems 9*. Cambridge, MA: MIT Press, 894-900.

Padgett C. and Cottrell, G. 1998. "A simple neural network models categorical perception of facial expressions". In *Proceedings of the Twenty First Annual Cognitive Conference of the Cognitive Science Society*, 806-811.

Paranagama, P., Burstein, F., and Arnott, D. 1997. "Modeling personality of decision makers for active decision support". In Jameson et al. (eds), *User-Modeling: Proceedings of the Sixth International Conference (UM'97)*. New-York, NY: Springer Wien, 79-82.

Pasztor, A. 1998. "Subjective experience divided and conquered". *Communication and Cognition* 31(1): 73-102.

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

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

Rinn, W.E. 1984. "The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions". *Psychological Bulletin* 95: 52-77.

Rolls, E. 1992. "Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas". In Bruce, V. and Cowey, A. Ellis, A. and Perrett, D. (eds), *Processing the Facial Image*. Oxford: Oxford University Press, 11-21.

Rosenblum, M., Yacoob, Y., and Davis, L. 1996. "Human expression recognition from motion using a radial basis function network architecture". *IEEE Transactions on Neural Networks* 7(5): 1121-1138.

Rumelhart, D.E., Hinton, G.E., and William, R.J. 1986. "Learning internal representation by error propagation". In Rumelhart and McClelland (eds), *Parallel Distributed Processing Explorations in the Microstructures of Cognition Volume 1: Foundations*. Cambridge, MA: MIT Press, 318-362.

Rumelhart, D.E. and McClelland, D. 1986. *Parallel Distributed Processing - Explorations in the Microstructures of Cognition Volume 1: Foundations*. Cambridge, MA: MIT Press.

Russell, J. 1980. "A circumplex model of affect". *Journal of Personality and Social Psychology* 39: 1161-1178.

Russell, J. 1994. "Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies". *Psychological Bulletin* 115(1): 102-141.

Samal, A. and Iyengar, P.A. 1992. "Automatic recognition and analysis of human faces and facial expressions: A survey". *Pattern Recognition* 25: 65-77.

Selker, T. 1994. "COACH: A teaching agent that learns". *Communications of the ACM* 37(7): 92-99.

Sharma, R., Pavlovic, V., and Huang, T. 1998. "Toward multimodal human-computer interface". *Proceedings of the IEEE* 86(5): 853-869.

Shneiderman, B. and Maes, P. 1997. "Direct manipulation vs. interface agents (Excerpts from debates at UII 97 and CHI 97)". *Interactions* 4(6): 42-61.

- Slaney, M. and McRoberts, G. 1998. "Baby ears: A recognition system for affective vocalizations". In *Proceedings of 1998 International Conference on Acoustics, Speech, and Signal Processing*. Los Alamitos, CA: IEEE Computer Society Press, 985-988.
- Sloman, A. 1990. "Motives, mechanisms, and emotions". In M. Boden (ed), *The Philosophy of Artificial Intelligence*. New York: Oxford University Press, 231-247.
- Stonham, T.J. 1986. "Practical face recognition and verification with Wisard". In H. Ellis and M.A. Jeeves (eds), *Aspects of Face Processing*. Lancaster: Martinus Nijhoff, 426-441.
- Stork, D. and Henneke, M. 1996. "Speechreading: An overview of image processing, feature extraction, sensory integration and pattern recognition techniques". In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*. Los Alamitos, CA: IEEE Computer Society Press, xvi-xxvi.
- Terzopoulos, D. and Waters, K. 1993. "Analysis and synthesis of facial image sequences using physical and anatomical models". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(6): 569-579.
- Turk, M. and Pentland, A. 1991. "Eigenfaces for recognition". *Journal of Cognitive Neuroscience* 3(1): 71-86.
- Valentin, T., Abdi, H., O'Toole, A., and Cottrell, G. 1994. "Connectionist models of face processing: A survey". *Pattern Recognition* 27: 1209-1230.
- Walker, J., Sproull, L., and Subramani, R. 1994. "Using a human face in an interface". In *Proceedings of the International Conference on Human Factors in Computing Systems*. New York: ACM Press, 185-191..
- Waynbaum, I. 1907. *La Physionomie Humaine: Son Mechanisme et son Rôle Social*. Paris: Alcan.
- Wierzbicka, A. 1992. "Defining emotion concepts". *Cognitive Science* 16: 539-581.
- Winograd, T. 1996. *Bringing Design Into Software*. Reading, MA: Addison-Wesley.
- Winograd, T. 1997. "The design of interaction". In P. Denning and R. Metcalfe (eds), *Beyond Calculation: The Next 50 Years of Computing*. Springer-Verlag, 149-162.
- Winograd, T. 1998. "A human-centered interaction architecture". Working Paper of The People, Computer, and Design Project.
- Yacoob, Y., and Davis, L. 1993. "Labeling of human face components from range data". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'93)*, 592-593.
- Yacoob, Y., and Davis, L. 1996. "Recognizing facial expressions by spatio-temporal analysis". In *Proceedings of the Twelfth International Conference of Pattern Recognition*, Jerusalem, Volume 1, 747-749.
- Yacoob, Y., Lam, H., and Davis, L. 1995. "Recognizing faces showing expressions". In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, 278-283.
- Young, A., Rowland, D., Calder, A., Etcoff, N., Seth, A., and Perrett, D. "Facial expression megamix: Tests of dimensional and category accounts of emotion recognition". *Cognition* 63: 271-313.
- Zajonc, R. 1984. "On the primacy of affect". *American Psychologist* 39: 117-124.
- Zajonc, R. 1989. "Feeling and facial efference: Implications of the vascular theory of emotion". *Psychological Review* 39: 117-124.
- Zajonc, R. 1994. "Emotional expression and temperature modulation". In S. Van Goosen, N. Van de Poll, and J. Sergeant (eds), *Emotions: Essays on Emotion Theory*. Hillsdale, NJ: Lawrence Erlbaum, 3-27.
- Zajonc, R. and Markus, H. 1984. "Affect and cognition: The hard interface". In C. Izard, J. Kagan, and R. Zajonc (eds), *Emotion, Cognition and Behavior*. Cambridge, MA: Cambridge University Press, 117-124.