Automated facial expression measurement: Recent applications to basic research in human behavior, learning, and education

Marian Stewart Bartlett and Jacob Whitehill,

Institute for Neural Computation, University of California, San Diego

Introduction

Automatic facial expression measurement systems have been under development since the early 1990's. The early attempts worked well in highly controlled conditions, but failed for spontaneous expressions, and real application environments. Automatic facial expression recognition has now advanced to the point that we are able to apply it to spontaneous expressions. These systems extract a sufficiently reliable signal that we can employ them in behavioral studies and to begin to develop applications that respond to spontaneous expressions in real time.

Tools for automatic expression measurement will bring about paradigmatic shifts in a number of fields by making facial expression more accessible as a behavioral measure. Previous behavioral studies employed objective coding of facial expression by hand, which required extensive training and could take hours to code each minute of video. The automated tools will enable new research activity not only in psychology, but also in cognitive neuroscience, psychiatry, education, human-machine communication, and human social dynamics. Statistical pattern recognition on large quantities of video data can reveal emergent behavioral patterns that previously would have required hundreds of coding hours by human experts, and would be unattainable by the non-expert. The explosion of research in these fields will also provide critical information for computer science and engineering efforts to make computers and robots that interact effectively with humans and understand nonverbal behavior. Moreover, automated facial expression analysis will enable investigations into facial expression dynamics that were previously intractable by human coding because of the time required to code intensity changes.

This chapter first overviews the state of the art in computer vision approaches to facial expression recognition, including methods for characterizing expression dynamics. The chapter then reviews behavioral studies that have employed automatic facial expression recognition to learn new information about the relationships of facial expression to internal state, and reviews the first generation of applications in learning and education that take advantage of the real-time expression signal.

State-of-the-art in computer vision approaches to automatic facial expression measurement

Automated facial expression recognition systems have been under development since the early 1990's (e.g. Mase 1991; Cottrell and Metcalfe, 1991). While the early systems worked well for the face video on which they were developed, generalization to new individuals, and new camera conditions, even when approximately frontal and well lit, remained a major challenge. Performance on spontaneous expressions tumbled to near chance. Much of the current research has focused on achieving robustness through machine learning, or statistical models, where the system parameters are estimated from large data samples. Advances have also included more informative image features and robust motion tracking.

INSERT FIGURE 1 ABOUT HERE

Automatic expression recognition systems share a similar overall architecture, shown in Figure 1, (1) The face is first localized in the image; (2) Information about the image is extracted from the face region ("feature extraction"); and (3) this information used to make a decision about facial expression ("classification"). The most important ways in which expression recognizers differ is the type of features extracted, the method of classification, and the integration of information over time. Below we overview these steps, and review some of the strengths and weaknesses of current approaches. For a more thorough analysis and comparison, the reviewer is referred to the survey papers (Tian et al., 2003; Fasel and Luettin, 2003; Pantic and Rothkrantz, 2000; Zeng et al., 2009).

Feature Types

Image features for expression recognition fall into three main categories: geometric features, motion features, and appearance-based features. Geometric features include the shape of the mouth or eye opening, relative distances between fiducial points such as the inner eyebrows, or relative positions of many points on a face mesh; (see Tian et al., 2001, for an example).

Motion features consist of displacements estimated by tracking individual feature points or tracking more complex shapes. Motion tracking continues to be a highly challenging area of computer vision research, in which even state-of-the-art tracking algorithms are subject to drift after a couple of seconds of tracking, and require re-initialization. Drift refers to an accumulation of position error over time. A particular challenge is that most tracking algorithms depend on a brightness constraint equation which assumes that brightness has neither been added nor subtracted from the image, but rather it has just moved. Facial expressions include numerous violations of this constraint, including lips parting to show the teeth, and wrinkling.

Appearance-based features attempt to extract information about the spatial patterns of light and dark in the face image. This information is typically extracted by applying banks of image filters on the pixel intensities. An image filter is like a template match. The more similar the image window is to the spatial pattern in the 2-D filter, the higher the output value. Commonly used features include Gabor filters (e.g. Bartlett et al., 2006), eigenfaces (e.g. Cottrell and Metcalfe, 1991), independent component filters (e.g. Donato et al., 1999), integral image filters (Wang et al., 2004; Whitehill et al., 2009), and histograms of edges at different spatial orientations (Levi and Weiss, 2004). Gabor filters, for example, are akin to a local Fourier analysis on the image obtained by applying templates of sine wave grating patches at multiple spatial scales, orientations, and positions on the image. Since the dimensionality of appearance-based feature vectors is high, on the order of tens of thousands of features, practical expression recognition systems typically require thousands of training images to achieve robust performance. Machine learning systems taking large sets of appearance-features as input, and trained on a large database of examples, are emerging as some of the most robust systems in computer vision for tasks such as face detection (Viola and Jones, 2004; Fasel et al., 2005), feature detection (Vukadinovic and Pantic, 2005; Fasel, 2006), identity recognition (Phillips et al., 2006), and expression recognition (Littlewort et al., 2006). Appearance-based features also don't suffer from drift, which is a major challenge for motion tracking.

Active Appearance Models (AAMs) integrate elements from geometric, tracking, and appearance-based approaches (Cootes et al., 2001; Lucey et al., 2006; Huang et al., 2008). AAM's are essentially a method for robust tracking of a set of facial landmarks by fitting the face image to a flexible face model. This flexible face model contains information not only about how landmark positions change with expression, but also how the image graylevels near the landmark points change in appearance. Robustness is achieved by constraining the motion tracking to fit statistical models of both the shape and appearance of the face. The approach requires a set of training images, usually of the individual face to be tracked, in which multiple landmark positions have been labeled as the face undergoes a range of facial expression changes that spans the tracking requirements in the run-time system. AAM's have performed well on individuals in which models have been trained, but have difficulty generalizing to novel individuals. Approaches to improve generalization of AAM's to novel faces is an area of active research. It is of note that AAM approaches to expression recognition typically use just the final landmark displacement information for expression recognition.

It is an open question which feature class, appearance-based, motion-based, or geometric, are best for expression recognition. Several studies that suggest that appearance-based features may contain more information about facial expression than displacements of a set of points (Zhang et al., 1998; Donato et al., 1999), although findings are mixed (e.g., Pantic and Patras, 2006). One compelling finding is that an upper-bound on expression recognition performance using hand-labeled feature positions (e.g. Michel and El Kaliouby, 2003) was found to be lower than the performance of a fully automated system using appearance-based features, tested on the same dataset (Littlewort et al., 2006). However there is no question that motion is an important signal, and may be crucial for detecting low intensity facial expressions. Ultimately, combining appearance-based and motion-based representations may be the most powerful, and there is some experimental evidence that this is indeed the case (e.g., Bartlett et al., 1999).

All three classes of features are highly affected by out-of-plane head rotations. Most systems require head pose to remain within about 10-20 degrees of a fully frontal view. Methods of handling out-of-plane rotation are an area of active research, and include learning view specific expression detectors for profile views (Pantic and Patras, 2006), or mapping onto a 3D head model, rotating to frontal, and re-projecting (e.g. Bartlett et al., 2002). As cameras and data storage become less expensive, multi-camera approaches are emerging as one of the more effective ways to address this problem.

Classification Methods

After image features are extracted from the face, a decision is made about facial expression based on the set of image measures. A few systems have used rule-based classifiers in which the mapping from feature values to facial expression is defined manually (e.g. Moriyama et al., 2002). For the most part, however, facial expression recognition systems use machine learning-based classifiers, such as neural networks (Tian et al., 2001), and more recent variants on neural networks such as support vector machines (Bartlett et al., 2006) and Adaboost (Tong et al., 2007; Whitehill et al., 2009), to make a decision about the facial expression in the image. Generally, the machine learning based classifiers, when trained on large amounts of data, give rise to more robust systems.

Methods of Integrating Over Time

There is a large body of research showing that the dynamics of facial expressions (i.e., the timing and the duration of muscle movements) are crucial for interpretation of human facial behavior (Russell and Fernandez-Dols, 1997; Ekman and Rosenberg, 2005; Frank et al., 1993; Valstar et al. 2006). In recognition of this, several computational methods have been explored for integrating information over time, and this is an active area of current research. Approaches include employing spatio-temporal image features, such as spatio-temporal Gabor filters or motion energy filters (e.g. Yang et al., 2007). Another method is to compute expression estimates, using only spatial features for each video frame and then combining these with a dynamic time series model such as a Hidden Markov Model (e.g. Zhang et al., 2008; el Kalioubi and Robson, 2005; De La Torre et all, 2007, Chang et al., 2006).

Levels of Description

Another consideration in facial expression recognition is the level of categorization of the stimulus. One approach is to recognize facial expressions according to a categorical model, for example, at the level of basic emotions such as happy, sad, afraid, etc. However there may be many states of interest, such as stress, interest, and fatigue, or variants of an emotion, such as frustration, annoyance, and rage, and the facial configurations of these states may be unknown.

A second, more flexible approach is to recognize individual facial actions (Facial Action Units (AUs), Ekman and Friesen, 1978), which can then be combined if categorical representations are required. An advantage of facial action systems is that they provide a more objective description of the facial expression, and enable discovery of new associations between facial movement and an emotional or cognitive state. The facial action coding system (FACS) (Ekman and Friesen, 1978) is a widely used method for coding facial expressions in the behavioral sciences. The system describes facial expressions in terms of 46 component movements, which roughly correspond to the individual facial muscle movements. An example is shown in Figure 2. Because it is comprehensive, FACS has proven useful for discovering new associations between facial movements and affective or cognitive states (see Ekman and Rosenberg, 2005) for a review of facial expression studies using FACS). The primary limitation to the widespread use of FACS is the time required to manually code the individual facial actions by human experts. It takes over 100 hours of training to become proficient in FACS, and it takes approximately 2 hours for human experts to code a single minute of video. An automated system for facial action coding would enable a large range of new research in behavioral science.

INSERT FIGURE 2 ABOUT HERE

Another parameterized expression coding system is the Facial Animation Parameters (FAPS) in the MPEG4 video standard (Pandzic and Forchheimer, 2002). FAPS codes the movements of a set of 66 facial feature points. This coding standard is an important advance in terms of compatibility of multiple systems. A drawback is that it was developed by engineers with experience in speech animation, not facial expression. The set of feature points provide sparse information on many behaviorally relevant movements other than those immediately around the mouth, and it also encourages systems to ignore appearance changes, resulting in the plastic looking animations in computer generated films such as Polar Express. However, FACS is not well suited for coding speech movements. Some combination of the two systems may avoid the problems inherent in each.

Training images and spontaneous expressions

For the task of learning categorical models, often a set of undergraduates or professional actors will be used to pose the desired facial expression. However, these posed expressions often differ from their spontaneous counterparts. Spontaneous and posed expressions have different structural and temporal properties. Part of the reason for these differences is physiological. It is well known that there are two distinct neural pathways for posed and spontaneous facial expressions, each one originating in different areas of the brain (see Rinn, 1984, for a review). Subcortically initiated facial expressions (the spontaneous group) are characterized by synchronized, smooth, symmetrical, and ballistic muscle movements whereas cortically initiated facial expressions (posed expressions) tend to be less smooth, with more variable dynamics,

and less synchrony among different muscles (Rinn, 1984; Frank et al., 1993; Schmidt et al., 2003; Cohn and Schmidt, 2004). Because of these differences, it is important to employ databases of spontaneous expressions, such as the drowsiness example described later in this chapter. While this is recognized by many research groups, elicitation and verification of the desired states is a major research challenge. (See Cowie 2008.)

Approaches that focus on recognition of elemental facial movements such as facial actions have another set of challenges for development of training data Such approaches require expert coding of face video, which is expensive and time consuming.

For both approaches, large numbers of training examples are required in order to recognize facial behavior with robustness. Moderate performance can be attained with tens of examples (Bartlett et al., 2003), and asymptotes with tens of thousands of examples (Whitehill et al., 2009). Experience in our lab has suggested that automated detection of facial actions, where the AU detectors were developed from a large number of training samples, provides a good foundation for subsequent recognition of subject states for which there may be a much smaller number of samples.

The precision of automated facial expression recognition systems depends on the richness of the training set. It is essential to not only have a good range of positive examples (images containing the target expression) but also to have a good range of negative examples (images to which your system should respond 'not happy'). The negative set is often overlooked, at great peril, as this can lead to false positives when the system is applied to real behavior. (See Whitehill et al., 2009).

The Computer Expression Recognition Toolbox

A number of the system features described above have been combined into a end-to-end system for fully automated facial expression recognition, called the Computer Expression Recognition Toolbox (CERT). CERT was developed at developed at University of California, San Diego, originating from a collaboration between Ekman and Sejnowski (Bartlett et al., 1996, 1999, 2006; Donato et al., 1999; Littlewort et al., 2006). The current system automatically detects frontal faces in the video stream and codes each frame with respect to 40 continuous dimensions, including basic expressions of anger, disgust, fear, joy, sadness, surprise, contempt, a continuous measure of head pose (yaw, pitch, and roll), as well as 30 facial action units (AU's) from the Facial Action Coding System. See Figure 3.

The technical approach to CERT is an appearance-based, discriminative approach. As described above, these approaches have proven highly robust and fast for face detection and tracking (e.g. Viola and Jones, 2001), do not suffer from initialization and drift, which presents challenges for state of the art tracking algorithms, and take advantage of the rich appearance-based information in facial expression images. Face detection, as well as detection of internal

facial features, is first performed on each frame using a generalization of the Viola and Jones face detector (Fasel et al. 2005). The automatically located faces are then aligned using a fast least squares fit on the detected features, and finally passed through a bank of Gabor filters at 8 orientations and 9 spatial frequencies (2:32 pixels per cycle at 1/2 octave steps). Output magnitudes are then normalized and passed to facial action classifiers.

Facial action detectors were developed by training separate support vector machines to detect the presence or absence of each facial action. The training set consisted of over 10,000 images that were coded for facial actions from the Facial Action Coding System, including over 5000 examples of spontaneous expressions. Tests on a benchmark dataset (Cohn-Kanade) show state of the art performance for recognition of basic emotions (98% correct detection for 1 vs. all, and 93% correct for 7 alternative forced choice of the six basic emotions plus neutral), and for recognizing facial actions from the Facial Action Coding System (mean .93 area under the ROC[1] curve for posed facial actions, .84 for spontaneous facial actions with speech). The system outputs a continuous value for each emotion and each facial action. These outputs are significantly correlated with the intensity of the facial action (Bartlett et al., 2006; Whitehill et al, 2009). More information about the facial expression detection system can be found in Bartlett et al., 2006.

INSERT FIGURE 3 ABOUT HERE

This system was employed in some of the earliest experiments in which spontaneous behavior was analyzed with automated expression recognition (Bartlett et al., 2008). These experiments addressed automated discrimination of posed from genuine expressions of pain, automated detection of driver drowsiness, adaptive tutoring systems, and an intervention for children with autism. The analysis revealed information about facial behavior that were previously unknown, including the coupling of movements. These experiments are described in the next section, along with landmark studies from other research labs, which were among the first to employ computer vision for basic research into facial behavior.

Applications to basic research in human behavior, education, and medicine

Pain, Fatigue, and Stress

[1] The Receiver Operator Characteristic curve (ROC) plots hits against false alarms as the decision threshold shifts from one extreme to the other. The area under the ROC is 0.5 for a system at chance and 1 for perfect detection. It is equivalent to percent correct on a 2-alternative forced choice in which a target and non-target are randomly selected and the system must choose which is the target (Green & Swets, 1966).

Automated Discrimination of Real From Faked Expressions of Pain. Given the two different neural pathways for facial expressions, one may expect to find differences between genuine and posed expressions of states such as pain. The ultimate goal of this work is not the detection of malingering per se, but rather to demonstrate the ability of an automated system to detect facial behavior that the untrained eye might fail to interpret, and to differentiate types of neural control of the face. It holds out the prospect of illuminating basic questions pertaining to the behavioral fingerprint of neural control systems, and thus opens many future lines of inquiry.

In a study by Littlewort and colleagues (2009), the computer expression recognition toolbox (CERT) was applied to spontaneous and posed facial expressions of pain (Figure 4). In this study, 26 participants were videotaped under three experimental conditions: baseline, posed pain, and real pain. The real pain condition consisted of cold pressor pain induced by submerging the arm in ice water. The study assessed whether the automated measurements were consistent with expression measurements obtained by human experts, and developed a classifier to automatically differentiate real from faked pain in a subject-independent manner from the automated measurements. A machine learning approach was employed in a two-stage system. In the first stage, a set of 20 detectors for facial actions from the Facial Action Coding System operated on the continuous video stream. This data was then passed to a second machine learning stage, in which a nonlinear support vector machine (SVM) was trained to detect the difference between expressions of real pain and fake pain. Measures of AU dynamics were extracted from the CERT outputs and passed to the real pain / faked pain classifier.

Naïve human subjects tested on the same videos were at chance for differentiating faked from real pain expressions, obtaining only 49% accuracy, where chance is 50%. The automated system was successfully able to differentiate faked from real pain. In an analysis of 26 subjects with faked pain before real pain, the system obtained 88% correct for subject independent discrimination of real versus fake pain on a 2-alternative forced choice. Moreover, the most discriminative facial actions in the automated system were consistent with findings using human expert FACS codes. In particular, in the faked pain condition the automated system output showed exaggerated activity of the brow lowering action (corrugator, as well as inner brow raise (central frontalis), and eyelid tightening, which were consistent with a previous study on faked versus real cold pressor pain that employed manual FACS coding (LaRochette et al., 2006).

The temporal event analysis performed significantly better than a SVM trained just on individual frames, suggesting that the real versus faked expression discrimination depends not only on which subset of AU's are present at which intensity, but also on the duration and number of AU events.

INSERT FIGURE 4 ABOUT HERE

Pain or no pain? In a related study, Ashraf et al (2007) measured the ability of an automated facial expression recognition system to estimate pain intensity using a system developed at Carnegie Mellon University (CMU). Pain is typically assessed by patient self-report. Self-reported pain, however, is difficult to interpret and may be impaired or not even possible, as in

young children or the severely ill. Behavioral scientists have identified reliable and valid facial indicators of pain. Until now they required manual measurement by highly skilled observers. Ashraf et al. developed an approach that automatically recognizes acute pain. Adult patients with rotator cuff injury were video-recorded while a physiotherapist manipulated their affected and unaffected shoulder. Skilled observers rated pain expression from the video on a 5-point Likert-type scale. From these ratings, sequences were categorized as no-pain (rating of 0), or pain (rating of 3, 4, or 5). Ratings of 1 or 2 were discarded as indeterminate. They explored machine learning approaches for pain-no pain classification using Active Appearance Models (AAMs). Keyframes within each video sequence were manually labeled for the positions of internal facial landmarks, while the remaining frames were automatiacally tracked the positions of the landmarks with the AAM. A set of appearance and shape features were then derived from the AAM and passed to a support vector machine for the pain/no-pain classification. The system achieved a hit rate of 81% for detecting pain versus no-pain. These results support the feasibility of automatic pain detection from video.

Automated Detection of Driver Fatigue. It is estimated that driver drowsiness causes more fatal crashes in the United States than drunk driving (Department of Transportation, 2001). Hence an automated system that could detect drowsiness and alert the driver or truck dispatcher could save many lives. Previous approaches to drowsiness detection by computer make presumptions about the relevant behavior, focusing on blink rate, eye closure, yawning, and head nods (Gu and Ji, 2004; Zhang and Zhang, 2006). While there is considerable empirical evidence that blink rate can predict falling asleep, it was unknown whether there were other facial behaviors that could predict sleep episodes. The work described here employed machine learning methods to real human behavior during drowsiness episodes. The objective of this study was to discover what facial configurations are predictors of fatigue. In this study, facial motion was analyzed automatically using the Computer Recognition Toolbox (CERT). In addition, we also collected head motion data using an accelerometer placed on the subject's head, as well as steering wheel data. (The automated yaw pitch and roll detectors had not been developed at the time of this study).

In this study, 4 subjects participated in a driving simulation task over a 3 hour period between midnight and 3AM. Video of the subjects' faces and time-locked crash events were recorded (Figure 5). The subjects' data were partitioned into drowsy and alert states as follows. The one minute preceding a crash was labeled as a drowsy state. A set of 'alert' video segments were identified from the first 20 minutes of the task in which there were no crashes by any subject. This resulted in a mean of 14 alert segments and 24 crash segments per subject.

INSERT FIGURE 5 ABOUT HERE

In order to understand how each action unit is associated with drowsiness across different subjects, Multinomial Logistic Ridge Regression (MLR) was trained on each facial action

individually. The five facial actions that were the most predictive of drowsiness by increasing in drowsy states were blink, outer brow raise, frown, chin raise, and nose wrinkle. The five actions that were the most predictive of drowsiness by decreasing in drowsy states were smile, lid tighten, nostril compress, brow lower, and jaw drop. The high predictive ability of the blink/eye closure measure was expected. However the predictability of the outer brow raise was previously unknown. We observed during this study that many subjects raised their eyebrows in an attempt to keep their eyes open. Also of note is that action 26, jaw drop, which occurs during yawning, actually occurred less often in the critical 60 seconds prior to a crash.

A fatigue detector that combines multiple AU's was then developed. An MLR classifier was trained using contingent feature selection, starting with the most discriminative feature (blink), and then iteratively adding the next most discriminative feature given the features already selected. MLR outputs were then temporally integrated over a 12 second window. Best performance of .98 area under the ROC was obtained with five features.

We also observed changes in the coupling of behaviours with drowsiness. For some of the subjects coupling between brow raise and eye openness increased in the drowsy state (Figure 6a,b). Subjects appear to have pulled up their eyebrows in an attempt to keep their eyes open. Head motion was next examined. Head motion increased as the driver became drowsy, with large roll motion coupled with the steering motion as the driver became drowsy. Just before falling asleep, the head would become still. See Figure 6c,d.

INSERT FIGURE 6 ABOUT HERE

This is the first work to our knowledge to reveal significant associations between facial expression and fatigue beyond eyeblinks. The project also revealed a potential association between head roll and driver drowsiness, and the coupling of head roll with steering motion during drowsiness. Of note is that a behavior that is often assumed to be predictive of drowsiness, yawn, was in fact a negative predictor of the 60-second window prior to a crash. It appears that in the moments just before falling asleep, drivers may yawn less, not more, often. This highlights the importance of designing a system around real, not posed, examples of examples of fatigue and drowsiness.

Automatic Detection of Stress. Dinges, et al. (2005) prototyped an automated system to discriminate between high and low levels of stress as expressed by a subject's face. The particular application that the research targeted was detection of stress in astronauts during space flight, but in fact the methods were quite general and were not tailored to this specific domain. In their experiment, 60 subjects completed a battery of computerized neurobehavioral tests, and the test sessions were video recorded. Tests were presented to each subject in both easy and difficult versions to induce stress of low and high levels, respectively. The high-stress version contained more difficult questions and allowed less time for answers.

The approach employed motion tracking followed by a dynamical model. The system tracks the face using a 3-D deformable face mesh that models both translation and rotation of the head (rigid motion) and deformations of the face itself (non-rigid motion). Feature vectors are then extracted from the face, including not only deformation parameters from the face mesh but also grayscale information from the eyes. These feature vectors are input to two dynamical models (Hidden Markov Models, HMMs), one trained on high-stress and the other trained on low-stress sequences. The two HMM's output a probability estimate that the sequence of input features were generated under high or low stress, respectively. Dinges, et al., post-processed these probabilities with an additional discriminative classifier (support vector machine), for deciding high versus low stress. The system was tested for subject dependent recognition. From each subject video, four sequences were extracted of 5-10 seconds duration. Two sequences of each subject (one low and one high-stress) were employed to train the system, the other two sequences were used for testing performance. The overall accuracy was reported at 70% for a high versus low stress decision. This was above chance but below the 85% accuracy of human judges. This is moderate performance, but provides the first support to our knowledge for detection of stress states using automated expression measurement.

The Science and Technology of Educating

Automated Feedback for Intelligent Tutoring Systems. There has been a growing thrust to develop tutoring systems and agents that respond to the students' emotional and cognitive state and interact with them in a social manner (e.g. Kapoor et al. 2007,D'Mello et al., 2007). Whitehill, et al. (2008) investigated the utility of integrating automatic facial expression recognition into an automated teaching system. This work used expression to estimate the student's preferred viewing speed of the videos, and the level of difficulty, as perceived by the individual student, of the lecture at each moment of time. This pilot study took first steps towards developing methods for closed loop teaching policies, i.e., systems that have access to real time estimates of cognitive and emotional states of the students and act accordingly.

In this study, 8 subjects separately watched a video lecture composed of several short clips on mathematics, physics, psychology, and other topics. The playback speed of the video was controlled by the subject using a keypress. The subjects were instructed to watch the video as quickly as possible (so as to be efficient with their time) while still retaining accurate knowledge of the video's content, since they would be quizzed afterwards.

While watching the lecture, the student's facial expressions were measured in real-time by the CERT system (Bartlett et al., 2006). After watching the video and taking the quiz, each subject then watched the lecture video again at a fixed speed of 1.0. During this second viewing, subjects specified how easy or difficult they found the lecture to be at each moment in time using the keyboard.

For each subject, a regression analysis was performed to predict perceived difficulty and preferred viewing speed from the facial expression measures. The expression intensities

themselves, as well as their first temporal derivatives, measuring the instantaneous change in intensity, were the independent variables in a standard linear regression. An example of such predictions is shown in Figure 7c for one subject.

INSERT FIGURE 7 ABOUT HERE

The facial expression measures were significantly predictive of both perceived difficulty (r=.75) and preferred viewing speed (r=.51). The correlations on validation data were 0.42 and 0.29, respectively. The specific facial expressions that were correlated with difficulty and speed varied highly from subject to subject. The most consistently correlated expression was AU 45 (``blink''), where subjects blinked less during the more difficult sections of video. This is consistent with previous work associating decreases in blink rate with increases in cognitive load (Holland and Tarlow, 1972; Tada 1986).

Overall, this study provided proof of principle, that fully automated facial expression recognition at the present state of the art can be used to provide real-time feedback in automated tutoring systems. The recognition system was able to extract a signal from the face video in real-time that provided information about internal states relevant to teaching and learning.

A related project that attempts to approximate the benefits of face-to-face tutoring interaction is a collaboration between the MIT media lab and the developers of AutoTutor (D'Mello et al., 2007). AutoTutor is an intelligent tutoring system that interacts with students using natural language to teach physics, computer literacy, and critical thinking skills. The current system adapts to the cognitive states of the learner as inferred from dialogue and performance. A new affect sensitive version is presently under development (D'Mello, et al., 2008) which detects four emotions (boredom, flow/engagement, confusion, frustration) by monitoring conversational cues, gross body language, and facial expressions. Towards this end, they have developed a database of spontaneous expressions while interacting with the automated tutor, which will significantly advance the field.

Applications in Neuropsychology and Medicine

Facial Expression Perception and Production in Children with Autism. Children with Autism Spectrum Disorders (ASD) are impaired in their ability to produce and perceive dynamic facial expressions (Adolphs et al., 2001). Automated facial expression recognition systems can now be leveraged in the investigation of issues such as the facial expression recognition and production deficits common to children with autism spectrum disorder (ASD). Not only can these technologies assist in quantifying these deficits, but they can also be used as part of interventions aimed at reducing deficit severity.

The Let's Face It! training program (LFI!) (Cockburn et al., 2008) is an intervention for children with autism spectrum disorder (ASD) that has been shown to significantly improve their face processing abilities. However, it is only capable of improving their receptive ability. In this project, we introduced CERT into LFI! in order to provide the children with immediate feedback on their facial expression production. In a prototype game, called SmileMaze, the system responds to the subject's smiles (Figure 8a). Such facial expression tasks engage children with autism and may aid in learning nonverbal behaviors essential for social functioning. Moreover, training in facial expression production may improve recognition, as perception and production have been shown to be linked in many areas of development.

This convergence of expertise from computer and behavioral science provides additional scientific opportunities beyond development of intervention games. For example, it enables us to more readily explore questions such as the effect of familiarity on recognition and generalization. Using CERT, we can capture and quantify training stimuli from the participant's environment, including parents, teachers, siblings and friends. The use of familiar faces in the program may not only provide a more engaging environment for the participant, but may also facilitate generalization of learned skills from familiar faces to novel faces.

Facial expression recognition technology also enables us to develop an "Emotion Mirror" application (Figure 8b) in which players control the expressions of a computer-generated avatar and/or images and short video clips of real faces. Here, participants can explore the same expression on different faces. This aids in training a generalized understanding of facial expressions. It also knits expressive production and perception as it is the participant's own face that drives the expressions shown on the avatar and/or image.

INSERT FIGURE 8 ABOUT HERE

A related project is underway at the MIT Media Lab (Madsen et al., 2008; Picard and Goodwin, 2008). They are developing new technology to help individuals with autism to capture, analyze, and reflect on a set of social-emotional signals communicated by facial and head movements in live social interaction. The system employs an ultramobile PC and miniature camera, which enables them to capture and analyze facial behavior of their own everyday social companions. The system then presents interpretations of the face and head movements, such as agreeing or confused. This approach with wearable technologies offers, for the first time, the ability to conduct *just-in-time in situ* assistance to help individuals with high functioning autism to learn facial expressions and underlying emotions in their own specific natural environments. A novel output display, called 'emotion bubbles' was developed, in which each mental state was represented by a different color, and bubble size indicated the magnitude of that state.

The facial expression analysis employs a system developed by el Kaliouby and Robson (2005).

This framework employs a commercial feature point tracker to obtain real-time measures of the locations of 24 features on the face. It uses the motion, shape, and color deformations of these features to classify 20 movement primitives including action units from the Facial Action Coding System, as well as 11 communicative gestures such as head nod or eyebrow flash. These measures are then passed to a dynamic Bayesian network to interpret the meaning of head and facial signals over time. The system was trained on a database of actors who displayed a range of cognitive and mental states. Rather than recognizing basic emotions, recognition is performed for the following six mental states: *agreeing, concentrating, disagreeing, interested, thinking* and *confused*. This database, called the Mind-Reading database (Baron-Cohen et al., 2004) was collected with the objective of providing children with autism examples of facial expressions that are relevant in every day life.

Pilot studies were conducted with adolescents diagnosed as high functioning autism. Subjects watched the bubble display respond to their own facial expressions, and also attempted to elicit specific bubbles in their own conversations with their friends. Experimenters witnessed multiple instances of subjects adjusting their own conversation flow to try to elicit the desired bubble, thereby providing practice for both eliciting and understanding mental states of others.

Automated Facial Expression Analysis of Psychiatric Disorders. Wang et al (2008) were the first to apply video-based automated facial expression analysis in neuropsychiatric research. They conducted case studies on two patients: One with schizophrenia and one with Asperger's syndrome. While it is well known that patients with Schizophrenia exhibit impairments in facial expression including flattened affect, and 'abnormal affect,' there is little objective data on their facial behavior due to the time required for manual coding. Similarly, little objective data exists characterizing facial expression production in Autism Spectrum Disorders. Studies such as this one will provide important information on facial expression production in these populations for relating to underlying neural pathology and social interaction deficits.

They employed a system for recognition of basic emotions that was trained on a dataset of 32 actors with evoked facial expressions. The evoked expressions were obtained by asking participants to describe a situation in their life pertaining to each emotion. These situations were recounted back to them by a psychiatrist and video recordings were taken during the recounting session. It was trained on four expressions plus neutral, and 3 intensities. This made a relatively small training set in machine learning terms (384 images, or 96 per class), but it is one of the very few systems to be trained on spontaneous expressions of basic emotions.

The evoked expression paradigm was then repeated for the patients. The automated system was used to measure dimensions such as the frequency of occurance of the target facial expression, and the probability of the subjects expression given the model trained on the control subjects. This study obtained some general findings, such as reduced occurances for sadness, anger, and fear for the patient with schizophrenia, and reduced occurances of fear for the patient with Asperger's, as well as a poor match to the controls for fear.

This paper is a first step towards a larger study to compare the facial behavior of these patients to the distribution of facial behavior in the healthy population. Automated expression measurement facilitates such studies, and provides a consistent measurement tool for comparing populations, something typically not possible with manual coding studies due to inter-coder variability. Approaches such as the one in this paper that are trained on full face expressions from healthy controls can indicate the degree to which expressions match the healthy population, but are less well suited to illustrating *how* they differ. Also, depending on the composition of the training set, they may be unable to differentiate some movements such as the zygomatic from the risorious, which move the lip corners obliquely versus laterally.[2] Systems that perform facial action coding may be better suited to making such discriminations.

Basic Research in Dynamic Facial Behavior

The Dynamics of Infant-Mother Smiles. Messinger et al (2008) conducted the first application of automated measurement to study facial expression coupling in infant-parent interaction. They studied the facial behavior of mothers and infants during natural play sessions for two mother-infant dyads. The face analysis software was the system developed at CMU by Cohn, Kanade and colleagues based on Active Appearance Models (AAM), described above. As is common with AAM-based methods, manual initialization of the face mesh, as well as intermittent re-initialization was necessary.

The facial behaviors that were analyzed with the automated system included smile (AU 12), eye constriction (AU 6), and mouth opening (25, 26). First, analysis of expression dynamics within subjects revealed that synchrony of smile-related movements differed for infants than for adults, For infants, correlations between mouth opening and smile strength, and mouth opening and eye constriction, were moderate to high, whereas for mothers, these correlations were lower and more variable.

Perhaps most demonstrative of the utility of automatic face measurement was the study of correlations between smile activity in one partner of the dyad and subsequent smile activity in the other partner. They investigated interaction between mother and infant by computing windowed cross-correlations of smile activity over time. The infant-mother smile activity exhibited changing (nonstationary) local patterns of association, providing a glimpse into turn-taking and the formation and dissolution of states of affective synchrony.

In this study, the automated system enabled analysis of expression dynamics that was previously not possible with manual FACS coding. While some important studies of dynamics exist (e.g. Frank, Ekman, and Friesen 1993), the coding of dynamics in these studies is coarse

[2] The schizophrenic patient in a figure from the paper, for example, is smiling with the risorius, yet receives high scores for 'happy.'

due to the time required for manual FACS coding of intensity, consisting of measures such as time to apex, duration of apex, and time to offset. Automated systems provided measurements of intensity on a frame-by-frame basis which facilitate new experiments on expression dynamics and coupling.

All Smiles Are Not Created Equal. Ambadar, et al. (2009) used automatic face analysis to study the morphological and dynamic characteristics of different types of spontaneous smiles, and more precisely, how these characteristics affect how smiles are perceived by other humans. All dynamics information except duration were measured automatically using an earlier version of the facial expression analysis software developed by Cohn and Kanade's group at CMU (Cohn and Kanade, 2007).

In Ambadar's experiment, 101 observers evaluated 122 different video sequences containing smiles. Each video sequence contained a single human subject spontaneously smiling while interacting with another human. The observers judged each video sequence to be either amused, embarrassed, nervous, polite, or other. Facial dynamics were then analyzed for three categories: amused, polite, and embarrassed/nervous.

Using the manual coding of the morphological characteristics (which AU was present) and automatic coding of the dynamics, the authors assessed which smile characteristics best distinguished each smile type from the others. Relative to perceived polite smiles, perceived amused smiles had larger amplitude, longer duration, more abrupt onset and offset, and more often included AU 6, open mouth, and smile controls. Relative to those perceived as embarrassed/nervous, perceived amused smiles were more likely to include AU 6 and have less downward head movement. Relative to those perceived as polite, perceived embarrassed/nervous smiles had greater amplitude, longer duration, more downward head movement, and were more likely to include open mouth.

Contrary to Ambadar, et al.'s hypothesis, asymmetry between left and right side of the face was not significantly different across the three smile types. The researchers speculated that the facial expression analysis software, in its current version of development, may not have been sufficient to detect subtle facial asymmetries, especially when the faces analyzed are non-frontal, as often occurs in practice.

Human Dynamic Facial Expression Perception. Automated facial expression measurement also provides a way to test the *perception* of dynamic faces by providing a means for developing dynamic stimuli. Several recent studies have emerged which measure facial movements of a human subject, and then map them onto an avatar, which enables aspects of the face such as appearance features or dynamic features to be manipulated. Curio et al (2008) employed this technology to enable new experiments on adaptation to dynamic facial expressions.

This study showed for the first time an after-effect for dynamic facial expressions. Subjects adapted to anti-happy or anti-disgust expressions, where an anti-expression is a morph in the

opposite direction to that of the original expression, relative to neutral. They were then tested for discrimination of reduced expressions, where a reduced expression is a morph in the same direction as the original expression, but attenuated. The task was a 2-alternative forced choice of happy or disgust.

Adaptation to anti-happy facial motions increased the recognition rates for happy and decreased recognition rates for disgust, and vice versa. The aftereffect was much stronger for dynamic vs. static adapting stimuli. (In both cases the test stimulus was dynamic. It would be interesting to look at cross-over to static test stimuli as well.) The study also showed that the aftereffect depended on identity. Both dynamic and static adaptation aftereffects were stronger when the identity of the adapting stimulus matched the identity of the test stimulus. This result contrasts other data from cognitive neuroscience suggesting separate encoding of expression and identity, although is consistent with some studies of static adaptation aftereffects (e.g. Ellamil, Anderson, and Susskind, 2008; Fox and Barton, 2007). Please see the chapter by Calder for a more thorough discussion of whether identity and expression are processed by separate visual routes. Interestingly, there was no significant difference in forward versus reverse dynamic display on the adaptation aftereffect.

Another study by Boker and colleagues (in press) used automated facial expression measurement and synthesis to explore gender effects on head nods. It was previously shown that women tend to nod their heads more than men, and individuals of either gender nod more when speaking with a woman than a man. This study attempted to differentiate whether the increase in nodding when speaking with a woman was due to facial mimicry, or awareness of the gender of the conversant. In this study, subjects conversed with an avatar that was driven by another person, and the effects of changing the appearance or head motion of the avatar on the subject's nonverbal behavior was examined. An Active Appearance Model (AAM) was employed to drive the avatar, where identity was encoded by mean shape and appearance of the model, and changes in expression were encoded by the coefficients on the basis vectors that span shape and appearance for that person. Apparent gender was manipulated by changing mean shape and mean appearance to another confederate of the same or opposite gender. Voice pitch was altered to match the apparent gender. Head motion in both individuals was measured using motion capture.

They found that changing appearance of the avatar from one gender to another did not affect head nods in the subject, but that the motion dynamics of the avatar did. Thus the gender effect on head nods was related to dynamic aspects of the stimulus, and not to static appearance parameters related to gender. This finding supports the role of facial mimicry in the head nods. Of course, dynamics can influence perceived gender, as can the audio signal, pitch manipulations not withstanding. The facial dynamics of the avatar nevertheless accounted for more of the subjects head nod behavior than the gender of the avatar as indicated by appearance and voice pitch.

This team used a similar approach to investigate the effect of depression on face-to-face interaction (Boker at al. 2009). They specifically looked at the effects of dampened facial expressions and head movements. They found that attenuated head movements led to

increased head nods and lateral head turns, and attenuated facial expressions also led to increased head nodding. These results are consistent with a hypothesis that the dynamics of head movements in dyadic conversation include a shared equilibrium, and contribute a new perspective on the effect of dampened affect on dyadic interaction.

A related line of research by Jonathan Gratch's group at USC has investigated the role of facial mimicry in eliciting social rapport. In these studies, two subjects interact through a computer monitor, where each subject views an avatar rendition of the other subject. Head pose is automatically tracked, and can either be rendered with fidelity, or with a manipulation such as displaying the head motion from the previous conversation. Such studies have shown a strong relationship between mimicry and ratings of rapport (Gratch et al., 2006; 2007).

Overall, this is a promising technique that will enable investigations of dynamic nonverbal behavior that were previously impossible.

Summary and conclusions

Automatic facial expression recognition has advanced to the point that we are now able to apply it to spontaneous expressions with some success. While the accuracy of automated facial expression recognition systems is still below that of human experts, automated systems already bring strengths to the table that enable new experiments into facial behavior that were previously infeasible. Automated systems can be applied to much larger quantities of video data than human coding. Statistical pattern recognition on this large quantity of data can reveal emergent behavioral patterns that previously would have required hundreds of coding hours by human experts, and would be unattainable by the non-expert. Moreover, automated facial expression analysis is enabling investigations into facial expression dynamics that were previously intractable by human coding because of the time required to code intensity changes. Automated facial expression technology such as CERT can be used in order to objectively characterize the distribution of facial expression productions in a large set of typically developing children and adults. Indeed, such a project is underway (Kang et al., 2008). This will provide a way to measure the degree to which facial expressions of patient populations diverge from norms, and describe the dimensions on which they diverge.

This chapter reviewed the state of the art in automated expression recognition technology, and outlined its capabilities and limitations. It then described a new generation of experiments that used this technology to study facial behavior and to develop applications in learning and education that take advantage of the real-time expression signal. The chapter also reviewed new experiments in dynamic face perception that have been enabled by this technology. Recent developments in expression tracking and animation have provided a way to parameterize and explore dynamic face space.

Tools for automatic expression measurement are beginning to bring about paradigmatic shifts in a number of fields by making facial expression more accessible as a behavioral measure. This

chapter described how these tools are beginning to enable new research activity not only in psychology, but also in cognitive neuroscience, psychiatry, education, human-machine communication, and human social dynamics.

References

Adolphs, R., Sears, L., and Piven, J., (2001). Abnormal Processing of Social Information from Faces in Autism, *J. Cogn. Neurosci.*, vol. 13, pp. 232-240, 2001.

Ambadar, Z., Cohn, J.F., and Reed, L.I. (2009). All Smiles are Not Created Equal: Morphology and Timing of Amused, Polite, and Embarrassed Smiles as Perceived by Observers. Journal of Nonverbal Behavior 33(1) p. 17-34.

Ashraf, A.B., Lucey, S. Chen, T., Prkachin, K., Solomon, P., Ambadar, Z., & Cohn, J.F. (2007). The painful face: Pain expression recognition using active appearance models. *Proceedings of the ACM International Conference on Multimodal Interfaces (ICMI'07),* Nagoya, Japan, 9-14.

Baron-Cohen, S., O. Golan, S. Wheelwright, and J. J. Hill. *Mind Reading: The Interactive Guide to Emotions*. London: Jessica Kingsley Publishers, 2004.

Bartlett, M. Stewart, Viola, P.A., Sejnowski, T.J., Golomb, B.A., Larsen, J., Hager, J.C., and Ekman, P. (1996). Classifying facial action, *Advances in Neural Information Processing Systems 8*, MIT Press, Cambridge, MA. p. 823-829.

Bartlett, M.S., Hager, J.C., Ekman, P., and Sejnowski, T.J. (1999). Measuring facial expressions by computer image analysis. *Psychophysiology*, *36*, 253-263.

Bartlett, M.S., Littlewort, G., Braathen, B., Sejnowski, T.J., & Movellan, J.R. (2003). A prototype for automatic recognition of spontaneous facial actions. In S. Becker & S. Thrun & K. Obermayer, (Eds.) Advances in Neural Information Processing Systems, Vol 15, p. 1271-1278, MIT Press.

Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2006). Automatic recognition of facial actions in spontaneous expressions. Journal of Multimedia, 1 (6), 22-35.

Bartlett, M., Littlewort, G., Whitehill, J., Vural, E., Wu, T., Lee, K., Ercil, A., Cetin, M. Movellan, J.

(2010). Insights on spontaneous facial expressions from automatic expression measurement. In Giese, M. Curio, C., Bulthoff, H. (Eds.) Dynamic Faces: Insights from Experiments and Computation. MIT Press.

Boker, S. M., Cohn, J. F., Theobald, B.-J., Matthews, I., Mangini, M., Spies, J. R., et al. (in press). Something in the way we move: Motion dynamics, not perceived sex, influence head movements in conversation. Journal of Experimental Psychology: Human Perception and Performance

Boker, S. M., Cohn, J. F., Theobald, B.-J., Matthews, I., Brick, T., Spies, J. R., (2009). Effects of Damping Head Movement and Facial Expression in Dyadic Conversation Using Real-Time Facial Expression Tracking and Synthesized Avatars. Proc. Philosophical Transactions of the Royal Society B v. 364 p. 3485-3495.

Chang, Y., Hu, C., Feris, R. & Turk, M. (2006). Manifold based analysis of facial expression. *J. Image & Vision Computing,* Vol. 24, No. 6, pp. 605-614.

Cockburn, J., Bartlett, M., Tanaka, J., Movellan, J., Pierce, M., and Schultz, R. (2008). SmileMaze: A Tutoring System in Real-Time Facial Expression Perception and Production for Children with Autism Spectrum Disorder. Intl Conference on Automatic Face and Gesture Recognition, Workshop on Facial and Bodily expressions for Control and Adaptation of Games.

Cohen, I., Sebe, N., Chen, L., Garg, A., & Huang, T. S. (2003). Facial expression recognition from video sequences: Temporal and static modelling. CVIU Special Issue on Face Recognition, 91, 160-187.

Cohn, J., & Kanade, T.(2007). Automated facial image analysis for measurement of emotion expression. In J. A. Coan & J. B. Allen (Eds.), The handbook of emotion elicitation and assessment. New York: Oxford University Press Series in Affective Science.

Cohn, J.F. & Schmidt, K.L. (2004). The timing of facial motion in posed and spontaneous smiles. *J. Wavelets, Multi-resolution & Information Processing*, Vol. 2, No. 2, pp. 121-132.

Cootes, T., Edwards, G., & Taylor, C. (2001). Active appearance models. Transactions on Pattern Analysis and Machine Intelligence, 23 (6), 681-685.

Cottrell G. & Metcalfe, J. (1991). Face, gender, and expression recognition using holons. In D. Touretzky (Ed.) Advances in Neural Information Processing Systems 3 p. 564-571.

Curio, C., Giese, M., Breidt, M., Kleiner, M., B¨ulthoff, H.(2008). Exploring human dynamic facial expression recognition with animation. Intl Conference on Cognitive Systems, Karlsruhe.

De la Torre, F., J. Campoy, Z. Ambadar and J. F. Cohn (2007). Temporal Segmentation of Facial Behavior. Proc. International Conference on Computer Vision, October 2007.

Dinges, D. F., Rider, R. L., Dorrian, J., McGlinchey, E. L., Rogers, N. L., Cizman, Z., et al. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. Aviation, Space, and Environmental Medicine, 76 (1), b172-b182.

Dinges, D.F., Venkataraman, S., McGlinchey, E.L., Metaxas, D.N.: Monitoring of facial stress during space flight: Optical computer recognition combining discriminative and generative methods. Acta Astronautica, 60:341-350, 2007.

Cowie, R. (2008). Building the databases needed to understand rich, spontaneous human behavior. Keynote talk, IEEE Conference on Automatic Face & Gesture Recognition.

D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White,H., Person, N., Kort, B., el Kaliouby, R., Picard., R.W. and Graesser, A., AutoTutor Detects and Responds to Learners Affective and Cognitive States, Workshop on Emotional and Cognitive Issues at the International Conference of Intelligent Tutoring Systems, June 23-27, 2008, Montreal, Canada.

D'Mello, S., Picard, R., & Graesser, A.(2007). Towards an affect-sensitive autotutor. IEEE Intelligent Systems, Special issue on Intelligent Educational Systems, 22 (4), 53-61.

Department of Transportation (2001). Saving lives through advanced vehicle safety technology. USA Department of Transportation. http://www.its.dot.gov/ivi/docs/AR2001.pdf.

Donato, G., Bartlett, M.S., Hager, J.C., Ekman, P. & Sejnowski, T.J. (1999). Classifying facial actions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 21, No. 10, pp. 974-989.

Ekman, P. (2001). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. W.W. Norton, New York, USA.

Ekman, P. (2003). Darwin, deception, and facial expression. *Annals New York Academy of sciences,* Vol. 1000, pp. 205-221.

Ekman, P., T. Huang, T. Sejnowski and J. Hager, eds., Final Report to NSF of the Planning Workshop on Facial Expression Understanding, technical report, Nat'l Science Foundation, Human Interaction Lab., Univ. of California, San Francisco, 1993.

Ekman, P. & Friesen, W.V. (1978). *Facial Action Coding System*, Consulting Psychologists Press, Palo Alto, USA.

Ekman, P. & Rosenberg, E.L., (Eds.), (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the FACS*, Oxford University Press, Oxford, UK.

el Kaliouby, R. and P. Robinson, Real-time Inference of Complex Mental States from Facial Expressions and Head Gestures, in Real-Time Vision for Human-Computer Interaction. 2005, Springer-Verlag. p. 181-200.

Ellamil, M., Susskind, J.M. and Anderson, A.K. (2008) Examinations of identity invariance in facial expression adaptation. *Cognitive, Affective, & Behavioral Neuroscience*, **8**, 273-81.

Essa IA, Pentland AP. Facial expression recognition using a dynamic model and motion energy. ICCV 1995:360–7.

Fasel I., Fortenberry B., Movellan J.R. “A generative framework for real-time object detection and classification.,” *Computer Vision and Image Understanding* 98, 2005.

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: Survey. Pattern Recognition, 36, 259-275.

Fox, C.J. and Barton, J.J.S. (2007) What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research*, **1127**, 80-89.

Frank MG, Ekman P, Friesen WV. (1993). Behavioral markers and recognizability of the smile of enjoyment. J Pers Soc Psychol. 64(1):83-93.

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., R. J. van der Werf, R., and Morency, L-P. (2006). Virtual Rapport. in *6th International Conference on Intelligent Virtual Agents*. 2006. Marina del Rey, CA: Springer.

Gratch, J., Wang, N., Okhmatovskaia, A., Lamothe, F., Morales, M and Morency, L-P (2007). Can virtual humans be more engaging than real ones? In *12th International Conference on*

Human-Computer Interaction. 2007. Beijing, China.

Gratch, Wang, Gerten, Fast, & Duffy (2007). Creating rapport with virtual agents. Int. Conf. on Intelligent Virtual Agents, Paris, France.

Green, D. M. and Swets, J.A. (1966). Signal Detection Theory and Psychophysics. New York, Wiley.

Gu, H., Ji, Q. (2004). An automated face reader for fatigue detection. In: Proc. Int. Conference on Automated Face and Gesture Recognition (2004) p. 111–116.

Holland, M.K. and G. Tarlow (1972). Blinking and mental load. Psychological Reports, 31(1).

Huang, X., and Metaxas, D. (2008), "Metamorphs: Deformable Shape and Appearance Models," In IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 30, No. 8, pp. 1444-1459, Aug. 2008.

Kang, G., Littlewort-Ford, G., Bartlett, M., Movellan, M., and Reilly, J. (2008). Facial expression production and temporal integration during development. Poster, UCSD Temporal Dynamics of Learning Center, NSF Site Visit.

Kaliouby, R., Robinson, P.: Real-time Inference of Complex Mental States from Facial Expressions and Head Gestures. In: Real-Time Vision for HCI, pp. 181–200. Spring-Verlag (2005)

Kapoor, A. and Picard, R. E. (2005), Multimodal Affect Recognition in Learning Envi-ronments, ACM MM'05, November 6-11, 2005, Singapore.

Kapoor, A., Burleson, W., & Picard, R. (2007). Automatic prediction of frustration. International Journal of Human-Computer Studies, 65(8):724-736.

Koelstra, S., & Pantic, M. (2008). Non-rigid registration using free-form deformations for recognition of facial actions and their temporal dynamics, Proceedings of IEEE Int'l Conf. Automatic Face and Gesture Recognition (FG'08), Amsterdam.

Larochette AC, Chambers CT, Craig KD (2006). Genuine, suppressed and faked facial expressions of pain in children. *Pain*. 2006 126(1-3):64-71.

Levi K and Weiss Y (2004). Learning Object Detection from a Small Number of Examples: The Importance of Good Features. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004.

Littlewort, Bartlett, & Lee (2009). Automatic coding of Facial Expressions displayed during Posed and Genuine Pain. Image and Vision Computing 27(12) p. 1797-1803.

Littlewort, G., Bartlett, M., Fasel, I., Susskind, J., and Movellan, J. (2004). Dynamics of facial expression extracted automatically from video. In IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Face Processing in Video.

Lucey, S., I. Matthews, C. Hu, Z. Ambadar, F. De la Torre and J. F. Cohn (2006). AMM Derived Face Representations for Robust Facial Action Recognition. Proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp. 155-160, Southampton, April 2006.

Madsen, M., el Kaliouby, R., Goodwin, M., and Picard, R.W., Technology for Just-In-Time In-Situ Learning of Facial Affect for Persons Diagnosed with an Autism Spectrum Disorder. Proceedings of the 10th ACM Conference on Computers and Accessibility (ASSETS), October 13-15, 2008, Halifax, Canada.

Mase, K. (1991). Recognition of facial expression from optical flow, IEICE Trans. 74 (1991), pp. 3474–3483.

Messinger, D.S., Cassel, T.D., & Cohn, J.F. (2008). The dynamics of infant smiling and perceived positive emotion. *Journal of Nonverbal Behavior* 32 (3), 133-155.

Michel, P. & el Kaliouby, R. (2003). Real time facial expression recognition in video using support vector machines. Proceedings of the 5th international conference on Multimodal interfaces.

Morecraft RJ, Louie JL, Herrick JL, Stilwell-Morecraft KS. (2001). Cortical innervation of the facial nucleus in the non-human primate: a new interpretation of the effects of stroke and related subtotal brain trauma on the muscles of facial expression. Brain 124(Pt 1):176-208.

Morency, L.-P., Rahimi, A., & Darrell, T. (2003). Adaptive view-based appearance model. In Proceedings of the 2003 IEEE conference on computer vision and pattern recognition (Vol. 1, pp. 803-810).

Moriyama, T., Kanade, T., Cohn, J., Xiao, J., Ambadar, Z., Gao, J., et al. (2002). Automatic recognition of eye blinking in spontaneously occurring behavior. In Proceedings of the 16th international conference on pattern recognition (pp. 78-81).

Pandzic, I. & Forchheimer, R. (Eds.) MPEG-4 Facial Animation: The Standard, Implementation and Applications. Wiley, 2002.

Pantic, M. and and I. Patras, I. (2006). Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments from Face Profile Image Sequences, IEEE Transactions on Systems, Man and Cybernetics - Part B, vol. 36, no. 2, pp. 433-449.

Pantic, M., & Rothkrantz, L. (2000). Automatic analysis of facial expressions: the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 1424-1445.

Phillips, J., Flynn, P., Scruggs T., (2006). Preliminary Face Recognition Grand Challenge Results, Proc. Int Conf on Automatic Face & Gesture Recognition.

Picard, R.W. and Goodwin, M., "Developing Innovative Technology for Future Personalized Autism Research and Treatment," Autism Advocate, First Edition 2008, Volume 50, No. 1, pages 32-39.

Rinn WE. The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expression. Psychol Bull 95:52–77.

Rogers, C.R., Schmidt, K.L., Van Swearingen, J.M., Cohn, J.F., Wachtman, G.S., Manders, E.K., Deleyiannis, F. W.-B. (2007). Automated facial image analysis: Detecting improvement in abnormal facial movement after treatment with Botulinum toxin A. *Annals of Plastic Surgery, 58,* 39-47.

Russell, J.A. & Fernandez-Dols, J.M., (Eds.), (1997). *The Psychology of Facial Expression*, Cambridge University Press, New York, USA.

Schmidt KL, Cohn JF, Tian Y. (2003). Signal characteristics of spontaneous facial expressions: automatic movement in solitary and social smiles. Biol Psychol. 65(1):49-66.

Tada, H. (1986). Eyeblink rates as a function of the interest value of video stimuli. Tohoku Psychologica Folica, 45.

Tian, Y.-L., Kanade, T., & Cohn, J. (2001). Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (2), 97-115.

Tian, Y.-L., Kanade, T., & Cohn, J. (2003, October). Facial expression analysis. In S. L. . A. Jain (Ed.), Handbook of face recognition. New York: Springer.

Tong, Y., Liao, W., & Ji, Q. (2007). Facial action unit recognition by exploiting their dynamic and semantic relationships. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29 (10), 1683-1699.

Valstar, M., H. Gunes and M. Pantic, 'How to Distinguish Posed from Spontaneous Smiles using Geometric Features', in Proceedings of ACM Int'l Conf. Multimodal Interfaces (ICMI'07), pp. 38-45, Nagoya, Japan, November 2007.

Valstar, M.F., Pantic, M., Ambadar, Z. & Cohn, J.F. (2006). Spontaneous vs. posed facial behavior: Automatic analysis of brow actions, *Proc. ACM Int'l Conf. Multimodal Interfaces*, pp. 162-170.

Viola, P., & Jones, M.(2004). Robust real-time face detection. International Journal of Computer Vision, 57 (2), 137-154.

Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., and Movellan, J. (2007). Drowsy driver detection through facial movement analysis. ICCV Workshop on Human Computer Interaction.

Wang, P, Barrett, F, Martin, E, Milonova, M, Gurd, R, Gurb, R, Kohler, C, Verma, R, (2008). Automated video-based facial expression analysis of neuropsychiatric disorders. Journal of Neuroscience Methods 168 (2008) 224–238.

Wang, Y., Ai, H., Wu, B., & Huang, C.(2004). Real time facial expression recognition with adaboost. In Proceedings of the 17th international conference on pattern recognition (ICPR 2004) (Vol. 3, pp. 926-929).

Whitehill, J., Bartlett, M., and Movellan, J. (2008). Automated teacher feedback using facial expression recognition. Workshop on CVPR for Human Communicative Behavior Analysis, IEEE Conference on Computer Vision and Pattern Recognition.

Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M., & Movellan, J. (2009). Toward practical smile detection. Transactions on Pattern Analysis and Machine Intelligence, 31(11) p. 2106-2111 2009.

Yacoob, Y., Davis, L., (1994). Computer spatio-temporal represatation of human faces. In: Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition'94, pp. 70–75.

Yacoob Y, Davis LS. Recognizing human facial expressions from long image sequeces using optical flow. IEEE Trans PAMI 1996;18(6):636–42.

Yang, P., Liu, Q., & Metaxas, D. N.(2007). Boosting coded dynamic features for facial action units and facial expression recognition. In Proceedings of the 2004 IEEE conference on computer vision and pattern recognition (pp. 1-6).

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (1), 39-58.

Zhang, Y, Ji, Q., Zhu, Z., and Beifang Yi, B. (2008). Dynamic Facial Expression Analysis and Synthesis with MPEG-4 Facial Animation Parameters. IEEE Transactions on Circuits and Systems for Video Technology 18(10) p. 1383-1396.

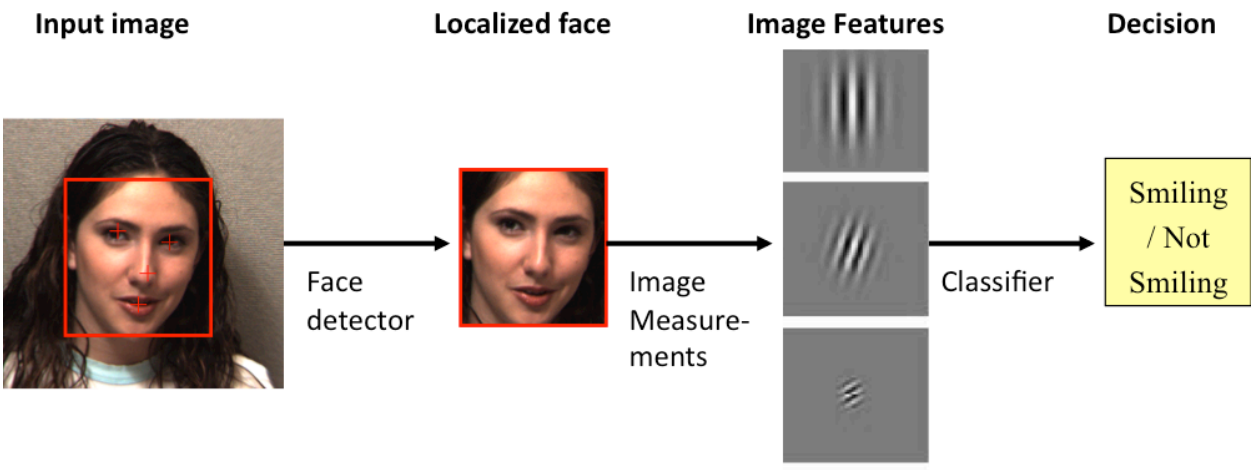


Figure 1. Schematic of automatic facial expression recognition.

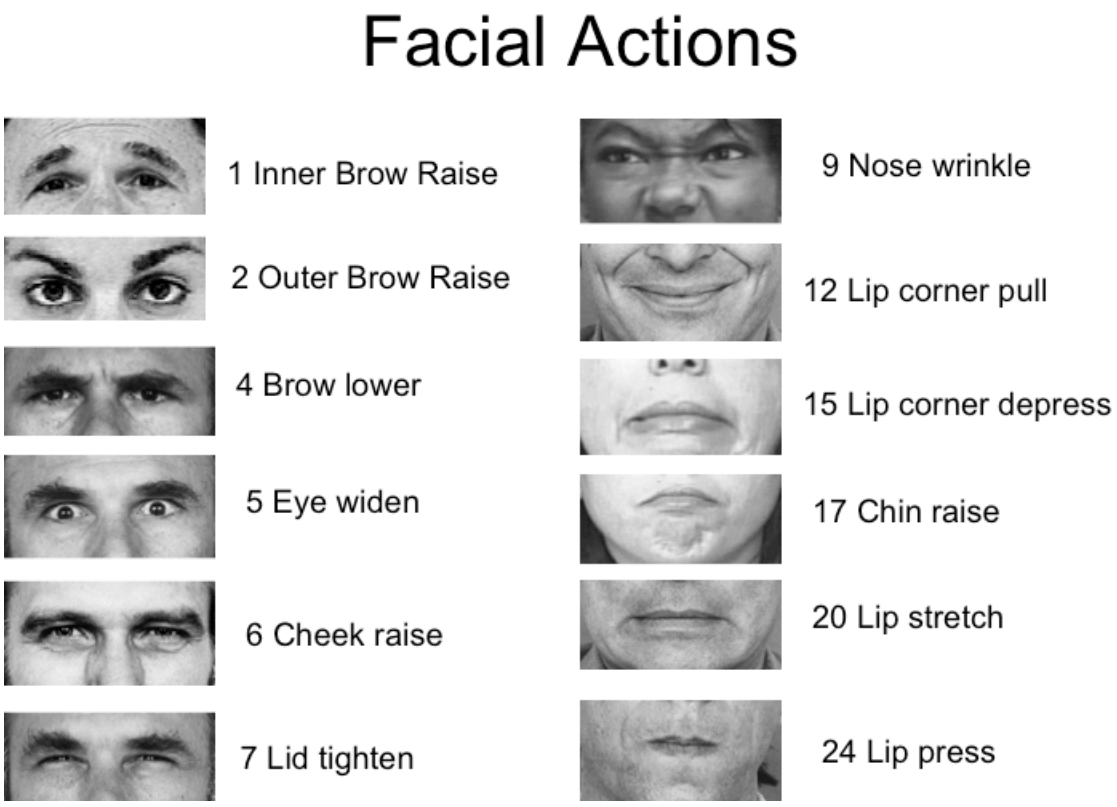


Figure 2. Sample facial actions from the facial action coding system. The system defines 46 distinct facial movements. (Reprinted from Bartlett et al., 2010, © MIT Press 2010.)

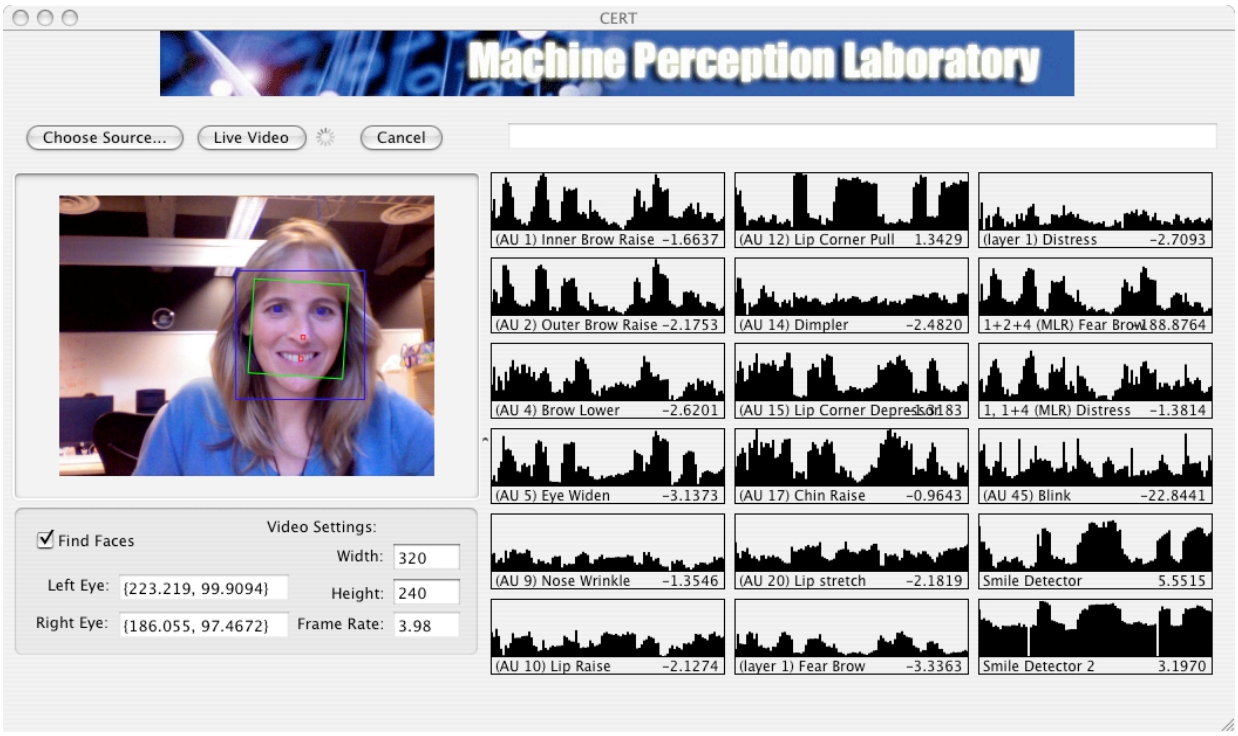
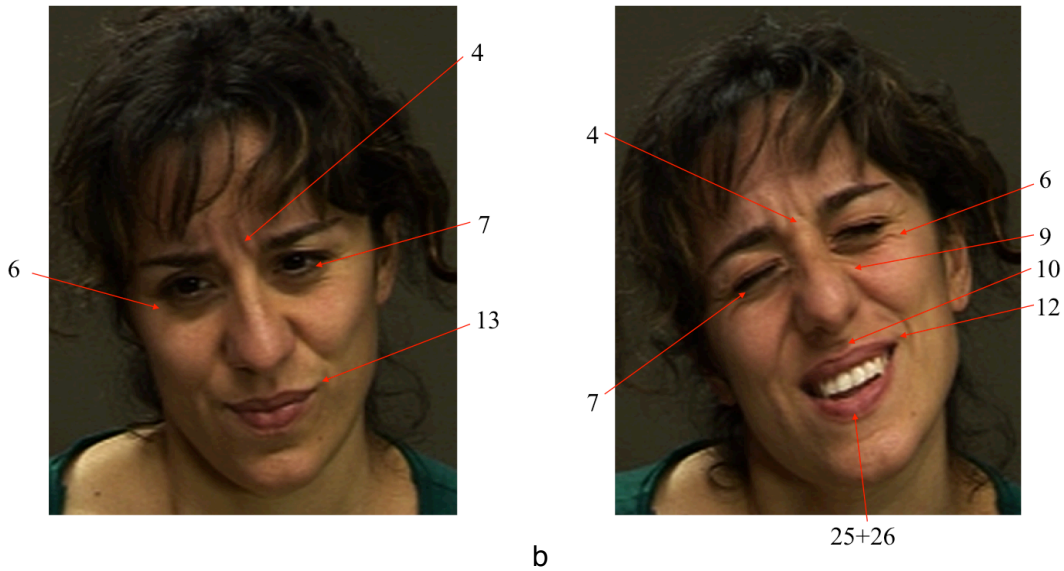


Figure 3. Example of CERT running on live video. In each subplot, the horizontal axis is time and the vertical axis indicates the intensity of a particular facial movement. (Reprinted from Bartlett et al., 2008. © 2008 IEEE).



a b

Figure 4. Facial expression of faked pain (a) and real pain (b), with corresponding FACS codes.

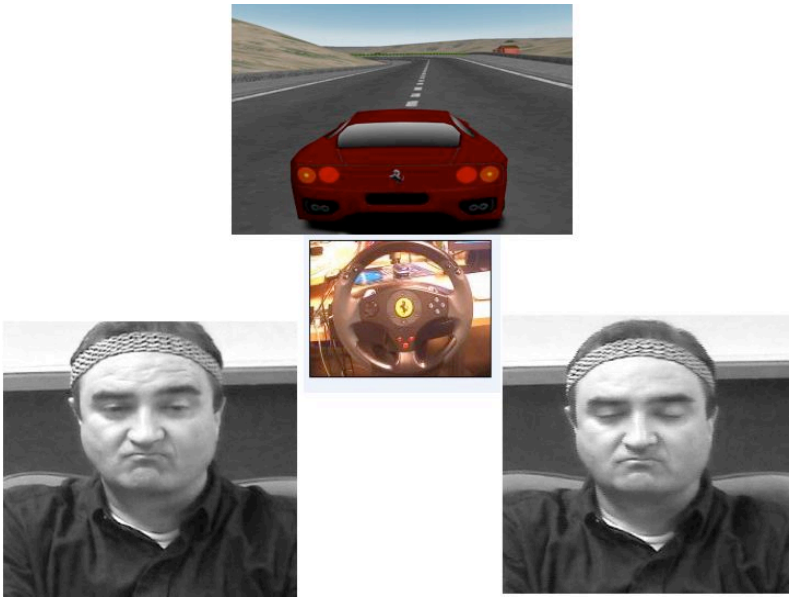


Figure 5. Driving Simulation Task. (Reprinted from Vural et al., 2007. © 2007 IEEE).

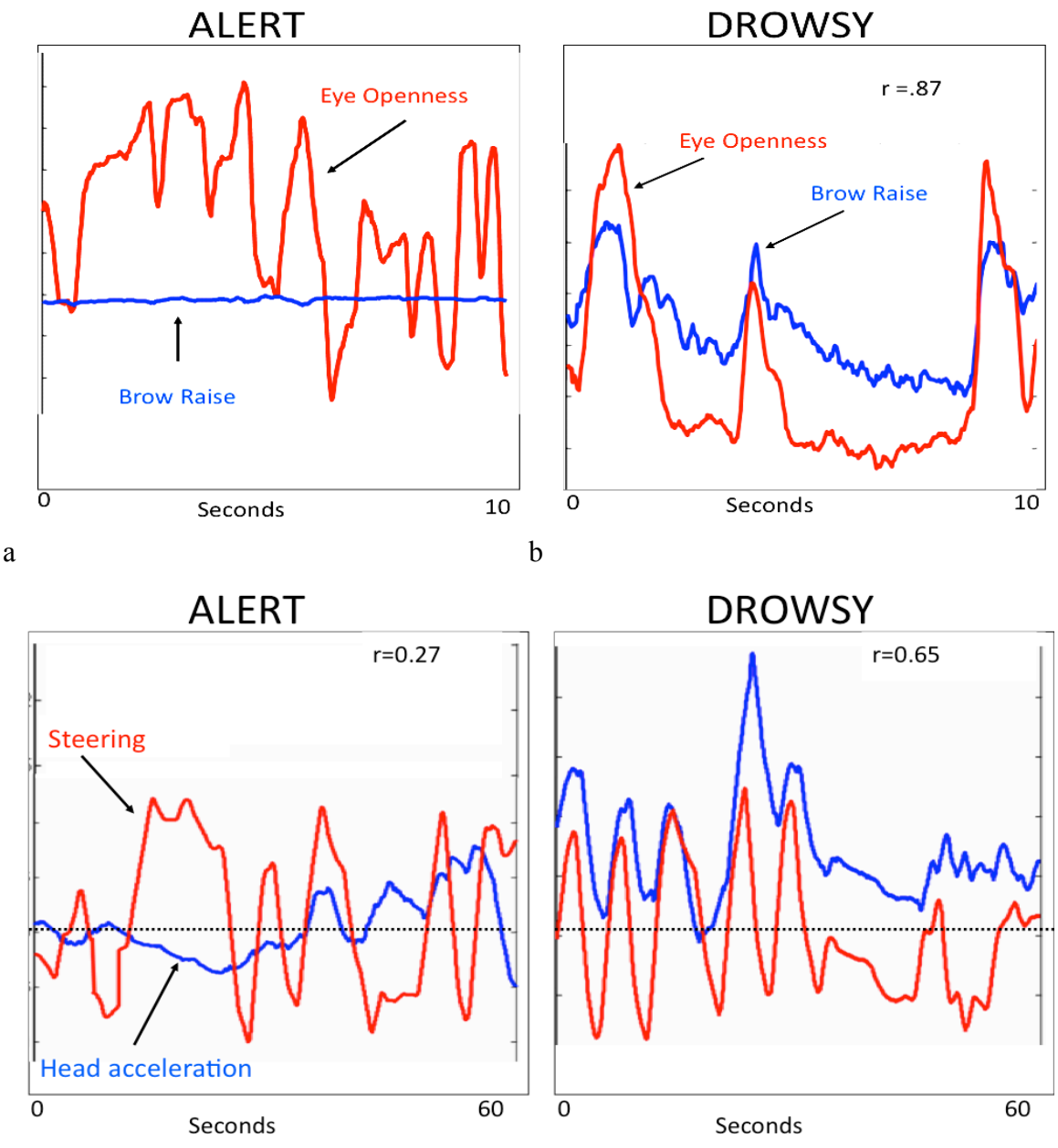


Figure 6. Changes in movement coupling with drowsiniess. a,b: Eye Openness (red) and Eye Brow Raise (AU2) (Blue) for 10 seconds in an alert state (a) and 10 seconds prior to a crash (b), for one subject. c,d: Head motion (blue) and steering position (red) for 60 seconds in an alert state (c) and 60 seconds prior to a crash (d) for one subject. Head motion is the output of the roll dimension of the accelerometer. (In grayscale, gray=blue, red=black.) (Reprinted from Bartlett et al., 2008, © 2008 Springer.)

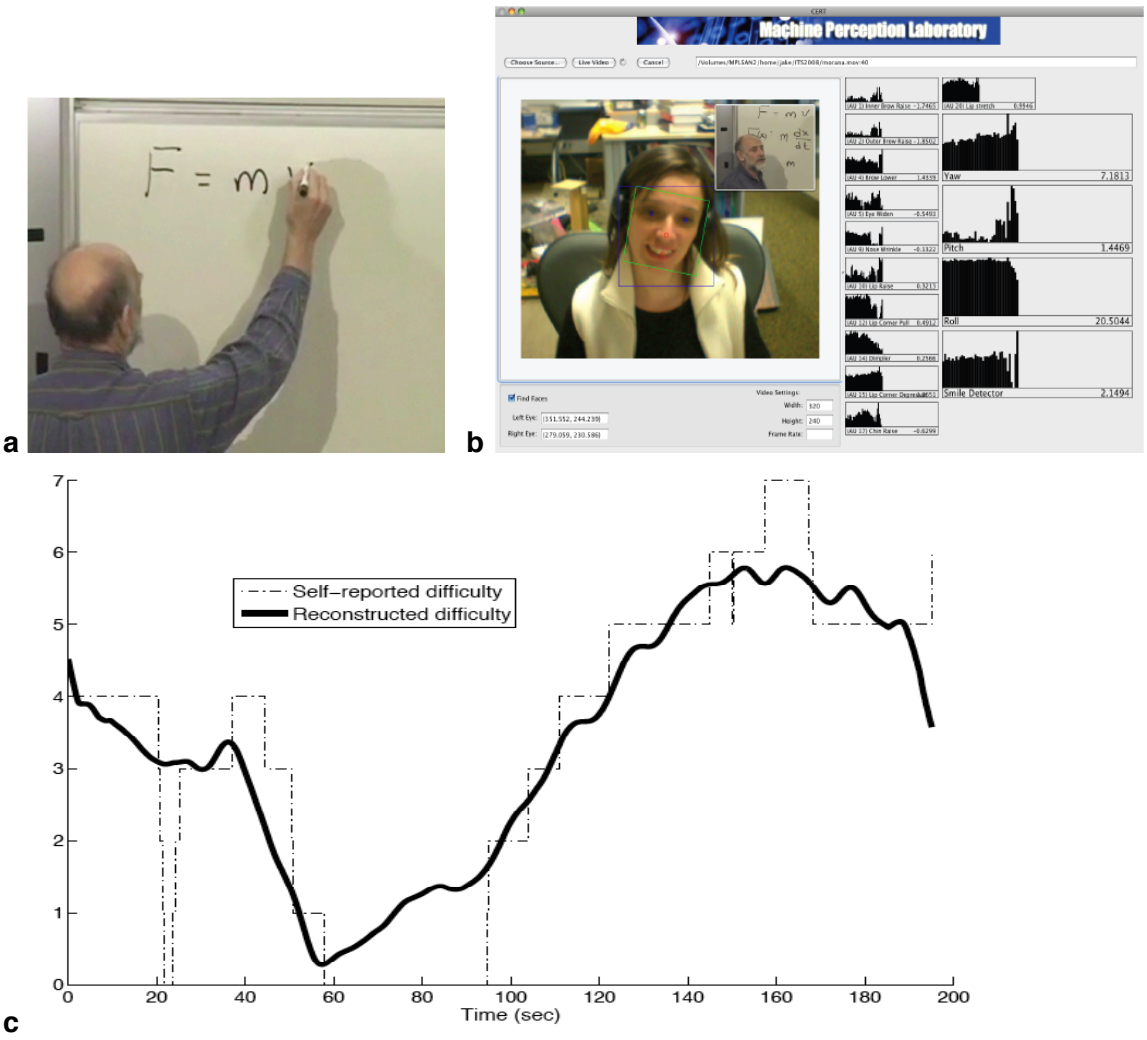


Figure 7 a. Sample video lecture. b. Automated facial expression recognition is performed on subjects face as she watches the lecture. c. Self-reported difficulty values (dashed), and the reconstructed difficulty values (solid) computed using linear regression over facial expression movements for one subject. (Reprinted from Whitehill et al., 2008. © 2008 IEEE).

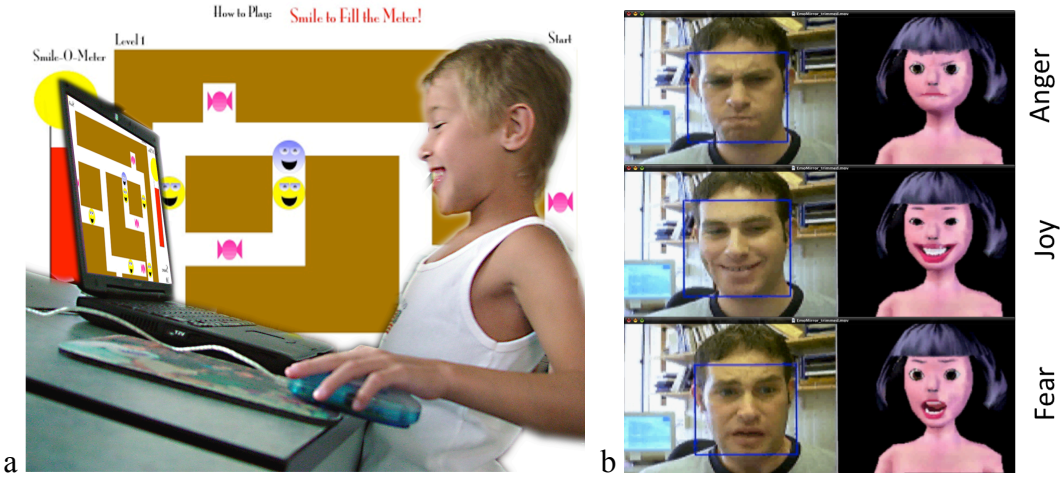


Figure 8. Prototype intervention tasks for children with ASD. a: Smile maze. The smile maze responds in real time to smiles by the subject. (Reprinted from Cockburn et al., 2008. © 2008 IEEE.) b. Emotion Mirror: An avatar responds to facial expressions of the subject in real-time. (Reprinted from Littlewort et al., 2004, © 2004 IEEE.)