20 Using automated facial expression analysis for emotion and behavior prediction

Sun Joo Ahn, Jeremy Bailenson, Jesse Fox, and Maria Jabon

The expression of emotion is achieved through intricate combinations of verbal and nonverbal information produced from various sources of the body and the brain. Nonverbal information encompasses any message that is not expressed in words (DePaulo and Friedman 1998) including gestures, postures, and vocal tones. Though people often do not invest much thought to the nonverbal aspect of communication due to its automaticity, nonverbal cues are the primary vehicles for expressing emotions and managing emotional experiences (Burgoon and Hoobler 2002). Among the communicative channels that express emotions, the face is often considered to be the richest source of nonverbal information (e.g., Collier 1985; Ekman 2003).

In the past, much of the research on facial expression has been based on Ekman's proposed method of measurement by dividing expressions into categories of predefined emotion labels (Ekman and Friesen 1978). Not only is this process labor-intensive but it is also restrictive in the sense that the interpretation of facial expressions must be based on this predefined number of categories regardless of the intensity, duration, and variety of feature movements. Automated facial expression analysis offers many benefits when compared to these methods; computers enable researchers to process incredible volumes of data in a short amount of time in a more systematic manner (Colmenarez et al. 2004). However, most of the work applying automatic facial feature detection until now is limited in that it still depends heavily on the conventional category-based measurements in the data analysis phase.

In this chapter, we propose a model of approaching facial expression detection and analysis that goes beyond category-based measurements by incorporating automated technologies that make create and automatically improve prediction models based on raw facial feature movements. Using a computer equipped with a small camera, tracking software, and machine learning (an automated form of computer-generated self-enhancement), we are able to select the most relevant facial features out of the massive collection of raw data and improve the prediction model in a time- and cost-efficient way. This allows us to utilize the raw data without fitting them into predefined categories, giving us greater analytical power than conventional category-based predictions. Moreover, our suggested methodology is notably unobtrusive when compared to other behavioral measures such as physiological measures, which usually require numerous sensors to be attached to the body.

We begin the discussion with a brief survey of facial expressions as a source of emotion signaling and conventional measurement methodologies used in prior studies of facial expression. We then introduce some advantages of using computers to classify emotions and to predict behavior based on raw facial feature data. The following section presents our proposed methodology of coupling automated facial expression analysis and machine learning in detail. In order to facilitate the understanding of this rather technically dense section, we also introduce several empirical studies applying the proposed methodology. Finally, we discuss the implications of this technology and suggest that automated facial expression analysis may be a valid new methodological tool for social scientists across diverse fields.

Facial expression of emotion

People have long believed that facial expressions are indicative of mental states but the scientific community has yet to reach a consensus regarding the relationship between facial expressions and emotions (for review, see Bente et al. 2008; Manstead and Fischer 2002; Russell and Fernandez-Dols 1997). The assumptions underlying our proposed methodology are that automatically gauged facial expressions will be valid reflections of mental states (see Izard 2009 for a review) which will allow us to predict future behaviors. Thus, the debates most relevant to our proposal would be those regarding the automaticity (vs. conscious control) and universality (vs. specificity) of facial expressions across individuals.

The bulk of the debate surrounding the emotion–expression link is based on dual-processing theories that give way to models with differing arguments for the interplay between automatic and controlled processing following stimuli (e.g., Gilbert 1991; Trope 1986). These models reflect contesting views on the degree of influence exercised by conscious control over the emotional response, determining the degree of expressed emotion. At the same time, the models agree that the initial emotional response following the stimulus is automatic (Barrett et al. 2007). Furthermore, as conscious control of emotion expression tends to be driven by social context (Jakobs et al. 1996) or particular goals (Zaalberg et al. 2004), it may be inferred that in situations devoid of particular social interactions, the emotional responses following stimuli are generated automatically.

In contrast to automatic facial responses, Ekman and Friesen (1969a) coined the term *display rules*, which are socially learned and often culturally distinct rules about the management of expression. Emotional responses are initially automatic and not controlled by conscious will (LeDoux 1996), but it appears that social rules can regulate the appropriate manifestation of emotion expressions. Ekman (1972) provided empirical evidence of these display rules by demonstrating in a series of studies that upon viewing films of surgery and accidents, Japanese people tend to mask negative expressions with a smile when in the presence of another person. When left alone, however, the Japanese participants displayed the same facial expressions demonstrated by American participants in a separate group. Thus, in situations where individuals need not incorporate such conscious control of emotion expression, they will allow the initial automatic responses to dominate, yielding veridical reflections of the internal state of their minds.

The issue of universality of facial expression across cultures has also gained some persuasive evidence. Recently, Matsumoto and Willingham (2009) tested the universality of emotion expression via facial expressions by using blind participants. Their results indicated that people who were born blind, people who became blind later on in their lives, and normally sighted people all shared common expressions of happy smiles after winning an athletic event. Further evidence is provided by neuroscience such as Panksepp's (1998) neurobiological model (see also Panksepp et al. 2000) which focuses on the emotions behind expressions. Similar to Ekman's work (1972) on universality of the expression for a set of 'basic' or 'core' emotions (e.g., anger, fear, joy, sadness, playfulness) across cultures, Panksepp also argues for basic emotion systems which are hardwired at birth by using positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) technologies. Although all emotion cannot be categorized into discrete, predetermined categories, these researchers maintain that the emotional range is created from subtle variations of the basic set of emotions.

Thus, according to the theories of automatic and universal expressions of emotion, detecting and assessing facial expressions will yield insights to internal emotional states and these results may even be generalized across cultures. Therefore, the discussion of various methodologies to scientifically measure and analyze facial movements is important in our attempt to capture the expressions of the face. Wagner (1997) divides these methods into two groups: measurement studies, which answer questions regarding facial behavior by the objective measurement of facial movements, and judgment studies, which answer research questions by gauging responses from observers. This chapter focuses on the former, as it applies to the prediction of emotion and behavior. One of the two most widely used methods in measurement studies of the face is electromyography (EMG), which measures the contraction of the muscles underlying facial skin (Cacioppo et al. 1990). However, the EMG tends to be intrusive due to the use of electrodes, the equipment is costly, and the output data can be unreliable (Wagner 1997).

The other form of measurement uses objective coding schemes based on visible units of facial behavior. This is usually done by post-processing videos recorded during an experiment, using slow-motion replay and frame-by-frame analysis of facial movements by human coders who systematically follow descriptive rules of judgment. Systematic measurements, such as the Facial Action Coding System (FACS) created by Ekman and Friesen (1978), the Maximally Discriminative Facial Movement Coding System (MAX) (Izard 1979), and emotional FACS (EMFACS) (Ekman and Friesen 1982) have been developed to objectively study the relationship between facial expressions and emotions. However, use of these methodologies requires extensive prior training of the coders and a frame-by-frame analysis of videotaped participants, which is labor intensive and raises problems regarding inter-observer reliability (Sayette et al. 2001). Furthermore, this process often leaves room for measurement variance due to human error, as there is no way to determine the precise onset, progression, and termination of each expression.

A factor that makes the measurement of facial expression difficult is that manifestation of emotion is fleeting and transitory, typically lasting from half a second up to four seconds (Ekman and Friesen 1969a, 1969b; Ekman 1984; Izard 1997). Despite conscious efforts to conceal true emotions following display rules, these involuntary micro-expressions tend to slip through and 'leak' actual internal states (Ekman 1997). For instance, in deception, a person may be able to consciously maintain an expressionless face (i.e., 'poker face') but may unconsciously shake his or her leg at the same time, leaking signs of nervousness.

In order to detect patterns in these veridical leakages of emotion on the face (i.e., by modeling facial feature movements), extensive observation is required, in which case using a computer to replace the human coder would yield optimal outcome. Frank (1997) comments that automatic facial tracking 'has the potential to save countless hours in scoring the dynamics of facial actions, thus it may make viable the discovery and exploratory work that is needed on dynamic markers of emotion in the face' (p. 240). As such, more researchers are embracing computers to collect, classify, and interpret various forms of nonverbal cues such as facial expressions, voice, or gestures as a more effective means of investigating emotion expressions compared to the human-coder. For instance, Picard's (1997) recent work includes using sensory inputs from multiple sources (e.g., facial expressions, head movement, and posture) to predict frustration in learning environments (Kapoor et al. 2007), monitoring dialogue, posture, and facial features to detect and respond to a learner's emotions and cognitive states (D'Mello et al. 2008), and developing portable aids which track, capture, and interpret facial and head movements of other people to assist individuals diagnosed with autism spectrum disorders in social interaction (Madsen et al. 2008).

Advantages of automatic facial expression analysis

Computer systems open up new horizons for emotion detection by recognizing and detecting nonverbal cues via automated devices. Although these systems are still far from achieving the capacity of human perception, they are able to classify and assess user emotions through predetermined mathematical models with limited human intervention (for a review, see Konijn and Van Vugt 2008; cf. Gratch, this volume; Prendinger and Ishizuka, this volume). Among the various modes of nonverbal communication, we focus on facial expressions which are captured by small cameras and later analyzed with computer software. Before introducing our proposed method in detail, we highlight some advantages of using technology to replace the more traditional forms of measurement discussed earlier.

First of all, automatic facial expression analysis performs with a higher accuracy than human coders who have large margins of error and may overlook important information. As Osgood (1953) noted, 'From the total splurge of sounds made by an actively vocal infant, only the small sample that happens to strike the observer is recorded at all' (p. 684). In contrast, computers are able to detect micro-expressions, even those that last for only a few seconds at a time, that human coders may miss. In addition, the computer does not fatigue and shows relatively little inconsistency in performance. As Webb et al. (2000) point out, 'People are low-fidelity observational instruments … recording and interpretation may be erratic over time, as the observer learns and responds to the research phenomena he [or she] observes' (pp. 143–4).

Recent studies applied automated feature extraction and classification to extract macro features such as the head and hand position and angle from video cameras (but not changes in facial features) taken during an experiment where a mock theft took place (Meservy et al. 2005). Computer models obtained up to 71 percent correct classification of innocent or guilty participants based on the macro features extracted from the video camera. Furthermore, in an overview of deception detection research, Meservy et al. (2008) noted that the accuracy of humans coding behavioral indicators only falls around 50 percent, but that computers trained to automatically extract and identify relevant behavioral cues detect deception with significantly higher accuracy. Furthermore, computers operate without the invasiveness of other methods (e.g., physiological measures such as polygraph machines or lie detectors) and the cost of extensively trained human interviewers.

Yet another advantage of using automated facial detection technology coupled with computational models is that once the system secures the parameters for a model, *prediction* of behavior (vs. simple detection and classification) can be made using only a small sample. This is a computational rendering of what social psychologists call 'thin-slicing,' a way people sample a short excerpt from social behavior to draw inferences about states, traits, and other personally relevant characteristics (Ambady and Rosenthal 1992). Prior research has demonstrated that initial moments of behavior have high correlations to the decisions made and actions taken in the future (Carrère and Gottman 1999; Gottman and Levenson 1992; Gottman and Notarius 2000). For instance, based on an observation of a three-minute video clip of a conflict between a married couple, Carrère and Gottman (1999) were able to predict the outcome of that marriage after six years. Using machine learning coupled with computer vision allows computers to mimic this human cognitive process; models are trained on a short sample of facial features and from those features automatically predict future behaviors.

Using this approach, Curhan and Pentland (2007) demonstrated that computers can use thin-slices of behavior in the first few minutes of a negotiation task to predict future outcomes. The methodology used in their study is comparable to ours in that they used data collected via computers and thin-slicing to predict behavior. In their study, computers were used in place of human coders to detect vocal behaviors (e.g., time spent speaking, influence over conversation partner, variation in pitch and volume, and behavior mirroring) during a negotiation task. Their results imply that the speech features extracted during the first five minutes of negotiation are highly predictive of future outcomes. The researchers also noted that using computers to code speech features offers advantages such as high test–retest reliability and real-time feedback.

As a cost-effective and relatively accurate method to detect, track, and create models for behavior classification and prediction, automatic facial expression analysis has the potential to be applied to multiple disciplines. Capturing behavioral data from participants such as facial expressions or head movements may be a more accurate representation of how and what they feel, and a better alternative to self-report questionnaires that interrupt participants' affective–cognitive processes and are subject to bias (Reynolds and Picard 2005; Picard and Daily 2005).

In the studies detailed in the following pages, we present some empirical applications of automatic facial detection coupled with learning algorithms. We begin with an introduction of the common methodology used in all of our experimental work and describe how we are able to derive behavior predictions from facial features.

Overview of proposed methodology

Facial expression recognition is a method of classifying facial motion and facial feature deformation into abstract classes based on visual information alone (Fasel and Luettin 2003). Our proposed method is a bottom-up approach in which the correlation between facial movement and a particular output (e.g., behavior, attitude, or emotion) becomes the formula for establishing not only a classification model but also a prediction model. As all of the raw facial feature movement data are processed to create these models, they reflect intricate combinations of feature movements and movement patterns which would be lost to the human coder using conventional measurement schemes.

Computer vision and machine learning form the cornerstones of our approach to modeling and predicting human behavior. Machine learning is a technique in which computers are programmed to modify and improve their performance based on novel data input (Bishop 2006). In this way the machines mimic the process of learning, autonomously adapting to external change in the environment (i.e., new input factors) and reducing the need for constant re-design.

Many of the problems approached with machine learning are *linearly separable*, which means that there is a clear distinction between one group and another in a given space. In this case, a model can be created to perform classifications in which parameters are established and modified until an optimal standard for categorizing different objects into a certain space is set. That is, through repeated trial and error, the model improves the method of calculating and estimating how to correctly classify individual instances of data until it reaches its optimum performance level. This is considered to be the *training* process through which these machine algorithms learn.

Although the computations behind this process are extremely complex, in greatly simplified terms, machine learning used in linearly separable classifications can be likened to a linear regression. There is an *input layer*, which is similar to independent variables, and an *output layer*, which is analogous to dependent variables. By detecting correlations between the input and output layers, the computer 'learns' how to derive the output with the given input. Once the computer goes through numerous sequences of input and output information, it is able to secure the parameters for a model. Then, as in regression analysis, this model can be used to predict a novel output with novel input data, thus successfully completing classification tasks.

However, not all problems in the real world are linearly distinguishable. The strength of certain machine learning methods, such as the Support Vector Machine (SVM), is that the machine classifier is effectively able to handle nonlinear problems (Elizondo 2006). This requires complex computations that map nonlinearly separable data into numerous independently linear dimensions until the data becomes linearly separable. This would be analogous to breaking down complex, multi-dimensional data into uni-dimensional data for ease of analysis. Using computers allows researchers to resolve this issue as the machine classifier is able to comb through large amounts of data and detect linear patterns from what seems to be a complex, nonlinear spread of data. This is a powerful alternative to conventional statistical methods in finding causal relationships in social science research.

To create our models we used facial feature data extracted from recorded video clips as the input layer and participant behavior gauged through behavioral measures or surveys as the output layer. Once the computer secured the parameters of its computational model, we were able to predict the behavior of a new person based solely on his or her new face feature data. Because we are looking at twenty-two or thirty-seven characteristic points on the face (depending on the specification of the tracking software) rather than at a single feature to derive correlations between facial features and behaviors, our input space is often complex and nonlinear. Using machine learning algorithms provides us with the analytic power to search for patterns on multiple dimensions.

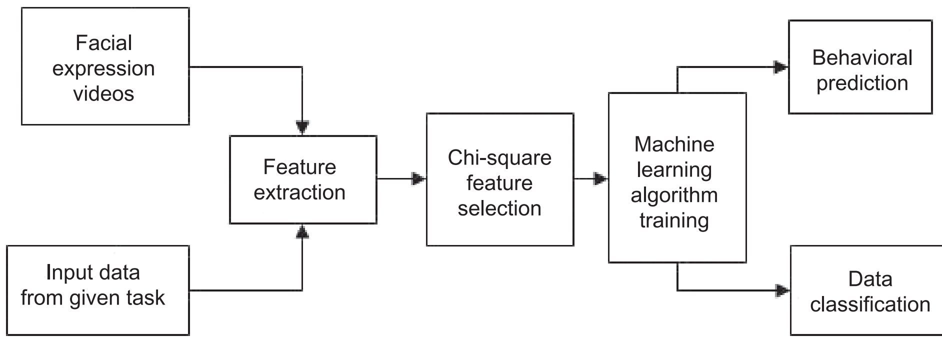

Figure 20.1 Data analysis procedure.

Although previous work has used machine learning to detect and classify the movement of facial features, our approach extends these efforts by focusing on the predictive function of these models. That is, although prior models stop at either detecting facial movements or classifying them into pre-designated categories, our model goes beyond to predict the future behavior within a given task (e.g., a virtual car accident or an error in performance). This opens up the possibility of such models becoming a common methodology in social scientific and behavioral research, as discussed in detail in the final pages of this chapter. Figure 20.1 summarizes the typical phases of data analysis used in our experiments.

Facial feature extraction

The first step in the construction of a typical dataset includes post-processing the videos of facial movements recorded during experimental tasks to extract key facial features and head movements. In our studies, a computer vision library, such as the Neven or the OKAO vision library, was used. The Neven vision library automatically detects, without any preset markers worn on the face, the x and y coordinates of twenty-two points along with eyes and mouth openness levels and head movements such as yaw, pitch, and roll (the measurement of object rotation for X, Y, and Z axes, respectively) for each frame of the video. The OKAO vision library, developed by OMRON Corporation, automatically detects and tracks thirty-seven points on the face, head movements such as pitch, yaw, and roll, as well as eye and mouth openness levels. In Figure 20.2, we present a screenshot of OKAO face tracking software.

Data synchronization and time series statistics calculation

In the next phase of analysis, video recordings are synchronized with data collected from experimental tasks such as surveys or simple motor tasks. This is done to map the extracted facial geometry information to behavioral output data. In the experiments described in this chapter, three to five second intervals of facial expressions were taken one to two seconds before each instance of the behavior to be predicted and used as the input data. After data synchronization we also computed a series of time-domain

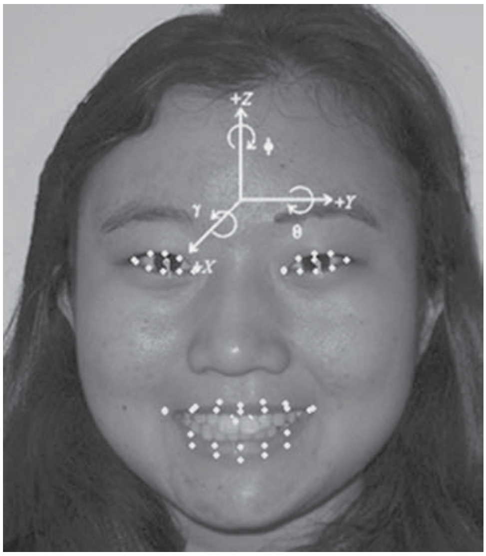

Figure 20.2 OKAO vision tracking points on participant's face.

statistics on the coordinates in each interval to use as additional inputs to our classifiers. Statistics calculated included averages, velocity, maximum value, minimum value, standard deviation, and range. All of these values were computed for each of the Neven or OKAO outputs. For some important facial characteristics, such as the eye and mouth openness levels, we created five-bin histograms from 0 percent to 100 percent to capture the distribution of eye and mouth state over the interval. We also calculated the percentage of the time the eyes were closed during each interval.

Statistical feature extraction

In our experiments, we discovered that classifiers using a reduced feature set tended to perform better because training the model with an excessive number of features leads to over-fitting. One reason for over-fitting is that there is often much overlap among our features. For instance, the left eye *x*-coordinate usually correlates well with the right eye *x*-coordinate. Including this redundant information results in reduced accuracy in our models. Thus, in order to reduce our datasets to only the most relevant features for each analysis, we performed a chi-square analysis for each of our datasets. A chi-square analysis evaluates the worth of an attribute by computing the value of the chi-squared statistic with respect to the class. By performing this analysis and ranking the obtained scores we were able to determine the most informative features for a given dataset and select only those above a certain threshold to use in our final classifications. This not only alleviated the problem of over-fitting, but also reduced computation costs and identified which features were most important indicators of the behavior we were aiming to predict, and sped up the process of training.

Dataset creation

To prepare our data for classification, we split each dataset into two subsets – an independent test and training set – by selecting test instances from a certain group of participants and training instances from a different, nonoverlapping set of participants. This was done so that we could secure the parameters of our model from one group of participants and be sure that none of those participants appeared in the dataset we tested our models on, which would result in falsely high reported accuracies.

Evaluation of classifier performance

Machine learning, as discussed above, was then used to analyze the dataset created. To select the type of classifier which would best perform each task, we tested some state of the art machine learning classifiers, including Support Vector Machines, Logitboost with a decision stump weak classifiers, Multilayer Perceptron Neural Networks, and Bayes Net Classifiers (Freund and Schapire 1996; Friedman et al. 2000) on each dataset. We built all classifiers using the freely distributed machine learning software package *Waikato Environment for Knowledge Analysis*.

To gauge the performance of our classifiers, we used two main measures: *precision* and *overall accuracy*. *Overall accuracy* is defined as the percentage of correctly classified instances. *Precision* is defined as the percentage of correctly classified instances in a given class. Note that *overall accuracy* is a gross measure of performance while *precision* is a measure of the accuracy within a given class. Thus in looking at both *overall accuracy* and *precision* we are able to obtain a comprehensive measure of the performance of our classifiers.

When viewing results, however, we also consider the *chance overall accuracy* and the *chance precision* for each class. We define a *chance classifier* as a classifier that would naively guess the class of an instance using the proportional split of the data; if the dataset consisted of twenty-five error instances and seventy-five nonerror instances, the *chance classifier* would classify an instance as error 25 percent of the time and as nonerror 75 percent of the time. The *chance overall accuracy* is defined as the percentage of instances such a naïve classifier would correctly classify and the *chance precision* for each class is the accuracy it would achieve in each class individually. By comparing our classification results to the results of this chance classifier, we are able to gauge the statistical significance of our models.

Empirical applications

In this section, we introduce some of our empirical studies that used computer vision coupled with machine learning algorithms to detect and classify facial features as well as predict future behavior based on these facial features. Compared to facial expression studies that use human coders, all of our studies were able to process more data faster with less labor, achieve high test–retest accuracy across all participants, and detect patterns in facial movements that predict future behaviors that may have been lost to the human eye. Along with research benefits for social scientists, implications for real-life applications will also be discussed for each experiment.

Real-time classification of evoked emotions

Our initial study (Bailenson et al. 2008) demonstrated the application of automatic facial recognition and machine learning to detect facial patterns and movements that correlate to discrete emotions. The processes and results of this study will further help readers understand how automated facial tracking and prediction based on computer vision and machine learning algorithm can be used to predict emotion.

The input data for this study consisted of videotapes of forty-one participants watching films that elicited the emotions of either amusement or sadness, along with measures of their cardiovascular activity, somatic activity, and electrodermal responding. It should be noted that the recorded expressions were spontaneously elicited expressions, unlike the photographs of deliberately posed faces often used in prior facial expression research (e.g., Ekman and Friesen 1971). Therefore, the probability that we accessed veridical emotions is higher than in studies that used deliberately posed faces (see Nass and Brave 2005 for further discussion of this distinction).

The output criteria was the intensity of the two emotions rated by trained coders looking at a second-by-second assessment of sadness or amusement of the recorded facial expressions using coding software informed by Ekman and Friesen's (1978) FACS. We correlated these input layers to the output layer (i.e., emotion) by building algorithms that used extracted points from the participants' faces as well as their physiological responses to predict the level of amusement or sadness for every person at every second.

We first created classifiers to detect discrete emotions (amusement, sadness, and neutral). These models were highly successful, demonstrating 90 percent overall accuracy in classifying amusement and 72 percent overall accuracy in classifying sadness, using only facial feature data. Furthermore, the overall accuracy rates increased to 91 percent and 75 percent respectively when physiological data was included with facial expressions as input layers.

Given the success of our initial binary classifiers, we also attempted the more difficult task of predicting the intensity of emotion. In predicting the emotional intensity of amusement, in terms of variance covered the model performed with a correlation coefficient of 0.53 ($MSE = 1.44$) based on just facial features. For predicting the intensity of sadness, the linear regression function performed with a correlation coefficient of 0.23 ($MSE = .79$). This suggests that machine learning algorithms may be successfully used to predict emotional intensity of amusement and sadness based solely on facial expression.

The results imply that computer vision coupled with machine learning algorithms may be used to detect, classify, and predict emotion and emotion intensity with relatively high accuracy. Along with important academic implications, the possibilities of industry applications are infinite. Examples of interested parties may include advertisers wanting to know the emotions of consumers, participants on dating sites interested in the emotion felt by their partners, and television producers interested in the emotions of their viewers.

Predicting unsafe driving behavior

Driving is an integral activity in the lives of many and emotions may well play a large role in driver safety (Lerner et al. 2008; Ranney 2008). There have been many recent research efforts to develop vehicle collision avoidance systems (Bertozzi et al. 2005; Kurihata et al. 2005; Suzuki et al. 2005; Williamson and Chamberlain 2005). These studies demonstrated that systems inside vehicles can use information collected about the driver, such as emotional states or distraction levels, to predict unsafe driving behavior. We extend the previous work by attempting to use facial expressions in not only detecting mental and emotional states but also in predicting driving accidents two full seconds before they occur.

For our study, participants were asked to drive through a 40-minute course in a driving simulator. The driving simulator was set up to run on a personal computer and the simulated image of the roadway was projected onto a white wall in the lab. The course simulated driving in a suburban environment with conditions varying from light to intense traffic (see Figure 20.3).

Based just on facial features, our system was able to predict minor and major automobile accidents with an average overall accuracy of 78.8 percent, which is over 25 percent better than the performance of a chance classifier. As explained earlier, chance classifiers are naïve methods of classification based purely on the proportions of instances. We also found that including supplementary information from the car, such as brake and accelerator maneuvers, improved the accuracy by 10 percentage points or more.

The results indicate that it would be relatively easy to create and apply a driver safety system using a small camera and a small computer to detect and analyze driver emotions and behaviors through facial expressions in real-time. Prior studies have shown that drivers who are more likely to become angry (e.g., those with high trait anger rates) tend to engage in more aggressive behavior on the road, which can result in negative outcomes such as crashes (Deffenbacher et al. 2003). Our model of accident prediction in this study and the model to predict emotional

Figure 20.3 Participant being monitored in STISIM driving simulator.

state and intensity discussed earlier could be coupled and applied to predict such emotional states *and* accidents before they occur.

Monitoring operator fatigue

Another area that we focused on is performance prediction, in particular, predicting the learning process of people tackling new and unfamiliar motor tasks. Many activities in everyday life require learning and performing some sort of motor skill, as do many jobs at workplaces. The ability to gauge the process of learning and predict future performance would lead to solutions that could optimize the productivity of both the individual and the organization.

This study (Ahn et al. 2009) aimed to use computer vision coupled with machine learning algorithms to predict errors before they occurred, to correctly identify whether a participant is learning or has already mastered a given task based on facial expressions, and also to predict overall participant performance quality based only on facial features. Moreover, we used thin-slicing to base our prediction of overall performance quality on facial data from the first seven minutes of participation.

The experimental task in this study was a simulation of a factory assembly line in which participants fit screws into designated holes using a *haptic device*, an input-equipment which allows users to feel, touch, and manipulate objects in virtual space. On the left-hand side of the monitor, participants saw three boxes, each containing a screw with a different part number. In the center of the screen, there was a wooden box with seven holes labeled with part numbers. The wooden boards were programmed to refresh to a new board after a pre-programmed amount of time. Every time the participant successfully filled out two consecutive boards without any errors (indicating that the level of difficulty was too low), the given time was curtailed by three seconds. This ensured that the participants were working at a rate of difficulty that corresponded to their level of learning. As can be seen in Figure 20.4, we recorded videos of the participants' faces and the movements of the haptic pen.

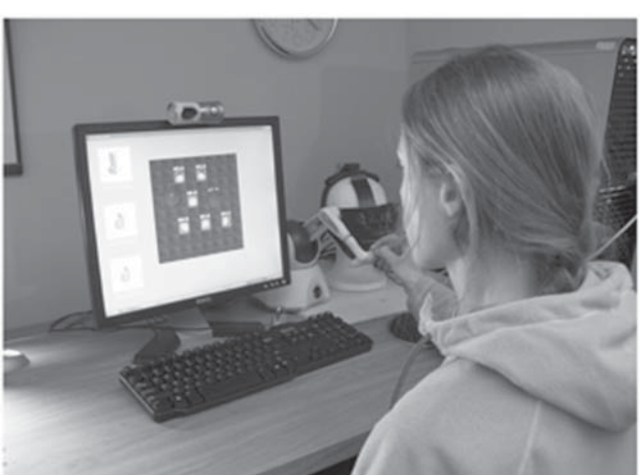

Figure 20.4 Experimental setup.

Using the collected facial expression data from the web-camera recordings and the performance data from the haptic pen, datasets were then constructed to build our computational models according to the processes discussed above.

We then built five separate prediction models. The first used facial data collected two seconds before an error occurred to predict the impending error. The second model classified whether or not a participant was in a 'learning' phase or 'mastery' phase of the experiment. *Learning* was defined as sessions in which the given time continuously decreased. Decreasing completion time indicated that participants were adjusting to the task and that they were gaining speed and accuracy. *Mastery* was defined as sessions in which the given time had stopped decreasing and had leveled off. This indicated that a participant reached his or her maximum performance capacity and was unable to complete the task any faster. The final three models predicted the overall quality of performance, learning capacity, and the rate of fatigue for each participant. In making the last three models we used only data from the first seven minutes of facial features, mimicking cognitive thin-slicing.

Results indicated that the face does indeed reveal information about impending errors and participant performance. The three models built to predict overall performance, learning capacity, and the rate of fatigue for each participant based on only a thin slice of data performed with overall accuracies of 91 percent, 78 percent, and 99 percent. The overall accuracy rate for error prediction two seconds before the event was 88.7 percent, and the overall accuracy of classifying learning and mastery phases were 66 percent.

Thus, we have evidence that computer vision coupled with machine learning algorithms is not only able to predict complex behaviors such as motor tasks and learning, but is also able to mimic cognitive thin-slicing. That is, the patterns of facial expressions detected within the first few minutes of the task could be used to predict the overall performance outcomes of participants at the end of the task. This further demonstrates the predictive power of patterns within fleeting facial movements. An example of a real-life implementation of this finding may be a system with which the facial expressions of workers are monitored to assess fatigue levels and establish the most effective work–rest schedules.

Online shopping

This study (Ahn et al. 2008) explored the possibility of predicting whether or not a consumer on an online shopping site will purchase a certain product based on only his or her facial expressions moments prior to the purchase. Online behavior involves countless dispositional and environmental variables. By predicting consumer behavior based on facial expression alone, our system would not only shed light onto online consumer decision-making processes but also yield important implications for developing advertisements or sales aids that react to the facial expressions of consumers. This could also lead to the establishment of an interactive shopping experience.

Until recently, clickstream has been the most widely used measure of online consumer behavior. Clickstream is a tracking technique that denotes the path a visitor takes through websites (Bucklin et al. 2002). Two aspects of browsing behavior are tracked: the visitor's decision to continue browsing or exit the site, and the length of time spent in each page. However, Fox and Spencer (2006) point out that although clickstream data can point out the ‘when’ and ‘what’ of web visits, it fails to answer questions about the ‘how’ and ‘why’ of consumer site use. In this light, supplementing facial data with click-through data will provide greater accuracy in predicting online consumer behavior.

The first part of the experimental task was to answer two rating questions regarding the twenty-four products selected from the pretest. The first question measured how much a participant liked the product and the second measured how likely the participant was to purchase it. The second part of the task showed a complete list of the twenty-four products rated by the participant in thumbnail pictures and required the participant to choose the final set of five products that he or she would like to buy. This was to relate the purchase intention measured in the first task to actual purchase behavior measured by this second task.

The dataset construction for this study first required extracting the facial features of the participants from the web camera videos taken during the study and then synchronizing this data with the user responses to the survey. Statistics were then computed on these facial points.

Overall, classifiers we subsequently created demonstrated better accuracies when predicting purchasing behavior than when predicting liking. For all categories of products, the classifiers performed from 8.35 percent to 44.4 percent above the performance of a chance classifier. This implies that consumers incorporate different information processing tactics when they evaluate products versus when they make purchasing decisions. In relation, previous work by Sood and Forehand (2005) demonstrated that memory tracing for choice tasks is higher than for judgment tasks. Choice tasks can be paralleled to the buying ratings in our study and judgment tasks to our liking ratings. The authors note that choice tasks encourage consumers to rely on heuristics which implies that the buying rating may be based more on heuristics than the liking ratings. Because heuristic decision-making takes into account peripheral cues such as emotions, and an increase in emotional stimuli leads to an increase in facial expressions, participants are likely to show richer facial expressions for buying ratings than for liking ratings, thus increasing the probability of better predictions.

Among products with a high likelihood of purchase, products that invoked amusement had a purchase intention prediction rate of 10 percent above chance. This could be accounted for by the fact that humor triggers an orienting response where individuals display a certain level of arousal and increased attention to process the novel stimulus. This could lead to leakage in facial expressions as the individual loses cognitive control of the face and the variance of facial feature movement may be intensified. In other words, the orienting response toward humorous products may have enriched the expressions displayed on the face, yielding a greater variance within the data for analysis.

Another category that yielded significant results among products that the participants wanted to buy was the high involvement products, with a purchase intention prediction rate of 44.4 percent above chance. Intuitively, high involvement products encourage consumers to become personally engaged in the decision-making process (Greenwald and Leavitt 1984; Zaichkowsky 1985) and such personal connections are bound to bring in heuristic cues heavily loaded with emotions and personal memories. Similar to orienting responses, high involvement in a product may then yield richer facial expressions, which would explain the higher predictive power.

Although this study did not focus on detecting the emotions of consumers as they shopped online, it is highly likely that some level of affect was involved in producing facial expressions, especially for high involvement and humorous products. One of the strengths of our approach is the system's ability to predict behavior without having to define emotions or to decide their relationships with facial expressions. The model is able to objectify facial movement regardless of the specificities of the felt emotion and establish a real-time system that can predict the purchasing behavior of consumers.

Discussion and implications

The face is an important nonverbal channel to signal emotions and its expressions hold important implications in communication and social interactions. Although more than 10,000 expressions can be made using just the facial muscles (Ekman 2003), scholars are still actively debating about a handful of basic emotions expressed through spontaneous facial movements (e.g., Fridlund 1994; Izard 1990; Matsumoto 1987; Russell 1997).

The difficulty of the topic of facial expression and emotions lies in the fact that, as noted by scholars such as Ruch (1997) and Zajonc (1998), emotions tend to be a slippery, hypothetical concept that cannot be defined and observed directly, but can only be inferred from several indicators such as behavior, physiological changes, or self-reports. Using a systematical measurement tool such as FACS helps researchers by breaking down facial expressions into distinctively measurable units, providing a tangible approach to investigating the causes and effects of facial expressions. However, FACS is labor-intensive, susceptible to problems of inter-coder reliability, and limits the usage of data.

Computer vision coupled with machine learning offers a cost effective and reliable alternative to human coding of emotion. Computers are able to offer measurement of facial feature movements with precision that would be difficult to match by human coders and provide accurate data on the duration and intensity of the movement. With the incredible rate of technical development, the machine's ability to detect and analyze facial expressions is likely to increase in the future. Just as computer programs now execute most statistical calculations, it may not be long before we rely heavily on computer power to plough through the massive amounts of data to figure out the relationship between facial expression, emotion, and behavior.

Our empirical studies demonstrate that coupling computer vision and machine learning algorithms can take facial expression research beyond the limited capacities of prior methodologies and help researchers obtain a firmer grip on this slippery topic. Based only on the raw data of facial feature movements, our models were able to: classify different emotions; predict an automobile accident before it happens; classify whether a person is learning or has mastered a task; predict an error for a given task before it occurs; predict whether a person will excel at a task, learn quickly, or fatigue quickly; and predict whether a consumer will purchase a product or not. All of these predictions were made via unobtrusive real-time collection of facial features. The models were even able to mimic thin-slicing and make predictions based on only a short sample of data.

In addition to some of the implications discussed with the empirical studies, our models have application potentials for other studies of mass media that involve emotion and media effects. One area that would benefit greatly from adopting automated facial detection technology would be research on children and media. Most studies use self-reports to measure responses from children after they view television programs (Cantor et al. 1993; Cantor and Nathanson 1996), but children have immature cognitive abilities compared to adults (Flavell et al. 2002). It may be difficult for them to fully comprehend some of the questions used in scales developed for adults. Measurement of children's emotions and responses via spontaneous facial expressions would be much more accurate than these self-reports.

This technology would also facilitate research of other media effects, such as those investigating the motivation and selection of media content. For instance, there have been numerous studies that incorporate Zillmann's concept of mood management (Zillmann 1988a; Zillmann, this volume), which explains media selection and consumption as a means for users to sustain positive moods or to terminate negative moods (Zillmann 1988b; Knobloch and Zillmann 2002). Using our system, researchers would be able to monitor emotional fluctuations in real time as participants make their selection of media content and verify whether or not the choices are indeed contributing toward managing an optimal state of happiness based on their facial expressions. Moreover, predictions of future media selection could be made based on the theory of mood management supplemented by analyses of facial expressions, further progressing current findings.

Another possible application of our proposed method in media research would be as a tool that supplements another methodology. One such example would be physiological measures that are used to gauge real-time responses to media exposure to assess such variables as attention, arousal, and orienting responses. In addition to measures of facial electromyography (EMG), skin conductance, and heart rate, which are often the standards in physiological measurements (Bradley et al. 1996; Bradley et al. 2001), automatic facial expression tracking could either successfully replace any one of these measures or complement them. As facial EMG involves obtrusive electrodes being placed directly on the face, and skin conductance and heart rate can reliably measure arousal but not valence (Winton et al. 1984), the use of our proposed system could offer simple solutions to overcome these limitations.

Beyond gauging intrapersonal responses, this technology could also serve to assist research endeavors in interpersonal communication. Interpersonal communication is vital to relationship management as it is a skill essential to the initiation, negotiation, maintenance, and termination of relationships between individuals (Burleson et al. 2000). For instance, prior research has demonstrated that communication success is one of the most important indicators of marital satisfaction (Steggell and Harper 1991). Since interpersonal competence, or the quality and skills of interaction, differ from individual to individual (Spitzberg and Cupach 1984), our system could improve the veracity and fidelity of interpersonal interactions by providing information from facial expressions that supplements the verbal message. More importantly, the system could aid in the resolution of interpersonal conflict, particularly when the complaint is withheld rather than confronted. Such unexpressed conflicts are harmful as they can lead to delayed and excessively emotional interactions that the confronted person may think are unjustified (Baumeister et al. 1990). Our proposed methodology could help detect dissatisfactions even when they are not expressed verbally and prevent relationship failures.

Finally, we discussed earlier how facial expressions could yield meaningful cues for deception detection and consumer behavior, which are highly relevant to influence and persuasion. Studying interpersonal influences is important since data suggest that influence attempts are more often successful than not (Cody et al. 1994). Our method of facial expression analysis could facilitate the efforts in compliance-gaining research. Many scholars have lamented about the shortcomings of relying on paper-and-pencil measures in lieu of behavioral measures, and basing their observations only on a single episode of message delivery rather than extended interactions (Berger 1985; Miller 1987). The nonverbal response data from tracking facial expressions will provide more concrete cues to behavioral changes induced by interpersonal influences compared to the use of self-reports alone. Furthermore, our system will endow researchers with the power to closely observe and analyze participant behavior for an extended period of time.

The face, as rich as its expressions are, is in no way a singular source of emotional expression that gives us insight into the human mind. Human emotion and its expressions are intricate combinations of mental, physical, internal, and external processes. Nevertheless, we focus on the face because it is the most evident signal of emotion and there is much to benefit from studying the face and its gestures. In the same way, the computer, lacking instinct or 'gut feelings,' may never perform entirely like the human mind. In fact, using computer vision and machine learning to predict human behavior is far from flawless. But we focus on the system because such shortcomings pale in comparison to the obvious advantages offered. If the face has a story to tell, systems of real-time behavior prediction based on facial expressions will be an effective methodological tool to help detect and interpret its message.

References

Ahn, S.J., Jabon, M.E. and Bailenson, J.N. (2008) 'Facial expressions as predictors of online buying intention', *Proceedings of the 58th Annual International Communication Association Conference*, May 22–26, Montreal, Canada.

Ahn, S.J., Jabon, M.E. and Bailenson, J.N (2009) 'Judging a book by the cover: Using facial expressions to predict performance', *Proceedings of the 59th Annual International Communication Association Conference*, May 21–25, Chicago, IL.

Ambady, N. and Rosenthal, R. (1992) 'Thin slices of behavior as predictors of interpersonal consequences: A meta-analysis', *Psychological Bulletin*, 2: 256–74.

Bailenson, J.N., Pontikakis, E.D., Mauss, I.B., Gross, J.J., Jabon, M.E., Hutcherson, C.A., Nass, C. and John, O. (2008) 'Real-time classification of evoked emotions using facial feature tracking and physiological responses', *International journal of Human Machine Studies*, 66: 303–17.

Barrett, L.F., Ochsner, K.N. and Gross, J.J. (2007) 'Automaticity and emotion', in J. Bargh (ed.) *Social Psychology and the Unconscious* (pp. 173–218), New York: Psychology Press.

Baumeister, R.F., Stillwell, A.M. and Wotman, S.R. (1990) 'Victim and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger', *Journal of Personality and Social Psychology*, 59: 991–1005.

Bente, G., Krämer, N.C. and Eschenburg, F. (2008) 'Is there anybody out there? Analyzing the effects of embodiment and nonverbal behavior in avatar-mediated communication', in E.A. Konijn, S. Utz, M. Tanis and S. Barnes (eds) *Mediated Interpersonal Communication* (pp. 131–57), New York: Routledge.

Berger, C.R. (1985) 'Social power and communication', in M.L. Knapp and G.R. Miller (eds) *Handbook of Interpersonal Communication* (pp. 439–99), Beverly Hills, CA: Sage.

Bertozzi, M., Broggi, A. and Lasagni, A. (2005) 'Infrared stereo vision-based pedestrian detection', *Proceedings of the 2005 IEEE Intelligent Vehicles Symposium* (pp. 24–29), Las Vegas, NV: IEEE Press.

Bishop, C. (2006) *Pattern Recognition and Machine Learning*, New York: Springer.

Bradley, M.M., Cuthbert, B.N. and Lang, P.J. (1996) 'Picture media and emotion: Effects of a sustained affective context', *Psychophysiology*, 33: 662–70.

Bradley, M.M., Codispoti, M., Cuthbert, B.N. and Lang, P.J. (2001) 'Emotion and motivation I: Defensive and appetitive reactions in picture processing', *Emotion*, 1: 276–98.

Bucklin, R.E., Lattin, J.M., Ansari, A., Gupta, S., Bell, D. and Coupey, E. (2002) 'Choice and the internet: From clickstream to research stream', *Marketing Letters*, 13: 243–56.

Burgoon, J.K. and Hoobler, G. (2002) 'Nonverbal signals', in M.L. Knapp and J. Daly (eds) *Handbook of Interpersonal Communication* (pp. 240–99), Thousand Oaks, CA: Sage.

Burleson, B.R., Metts, M.W. and Kirch, M.W. (2000) 'Communication in close relationships', in C. Hendrick and S.S. Hendrick (eds) *Close Relationships: A Sourcebook* (pp. 244–58), Thousand Oaks, CA: Sage.

Cacioppo, J.T., Tassinary, L.G. and Fridlund, A.J. (1990) 'The skeletomotor system', in J.T. Cacioppo and L.G. Tassinary (eds) *Principles of Psychophysiology: Physical, Social, and Inferential Elements* (pp. 325–84), Cambridge: Cambridge University Press.

Cantor, J. and Nathanson, A.I. (1996) 'Children's fright reactions to television news', *Journal of Communication*, 46: 139–52.

Cantor, J., Mares, M.L. and Oliver, M.B. (1993) 'Parent's and children's emotional reactions to TV coverage of the Gulf War', in B.S. Greenberg and W. Gantz (eds) *Desert Storm and the Mass Media* (pp. 325–40), Cresskill, NJ: Hampton Press.

Carrère, S. and Gottman, J. (1999) 'Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion', *Family Processes*, 38: 293–301.

Cody, M.J., Canary, D.J. and Smith, S.W. (1994) 'Compliance-gaining goals: An inductive analysis of actors' and goal types, strategies, and successes', in J.A. Daly and J.M. Wiemann (eds) *Strategic Interpersonal Communication* (pp. 33–90), Hillsdale, NJ: Erlbaum.

Collier, G. (1985) *Emotional Expression*, Hillsdale, NJ: Erlbaum.

Colmenarez, A.J., Xiong, Z. and Huang, T.S. (2004) *Facial Analysis from Continuous Video with Applications to Human-Computer Interface*, Boston, MA: Kluwer.

Curhan, J.R. and Pentland, A. (2007) 'Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes', *Journal of Applied Psychology*, 92: 802–11.

Deffenbacher, J.L., Lynch, R.S., Filetti, L.B., Dahlen, E.R. and Oetting, E.R. (2003) 'Anger, aggression, risky behavior, and crash-related outcomes in three groups of drivers', *Behaviour Research and Therapy*, 41: 333–49.

DePaulo, B.M. and Friedman, H.S. (1998) 'Nonverbal communication', in D. Gilbert, S. Fiske, and G. Lindzey (eds) *Handbook of Social Psychology*, vol. 2 (pp. 3–40), 4th edn, Boston, MA: McGraw-Hill.

D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., el Kaliouby, R., Picard, R., and Graesser, A. (2008) 'AutoTutor detects and responds to learners affective and cognitive states', *Proceedings of the Workshop on 'Emotional and Cognitive Issues', International Conference of Intelligent Tutoring Systems*, June 23–27, Montreal, Canada.

Ekman, P. (1972) 'Universals and cultural differences in facial expressions of emotion', in J. Cole (ed.) *Nebraska Symposium on Motivation* (pp. 207–83), Lincoln, NE: University of Nebraska Press.

Ekman, P. (1984) 'Expression and the nature of emotion', in K. Scherer and P. Ekman (eds) *Approaches to Emotion* (pp. 319–43), Hillsdale, NJ: Erlbaum.

Ekman, P. (1997) 'Should we call it expression or communication?', *Innovations in Social Science Research*, 10: 333–44.

Ekman, P. (2003) *Emotions Revealed*, New York: Times Books.

Ekman, P. and Friesen, W.V. (1969a) 'Nonverbal leakage and cues to deception', *Psychiatry*, 32: 88–105.

Ekman, P. and Friesen, W.V. (1969b) 'The repertoire of nonverbal behavior: Categories, origins, and coding', *Semiotica*, 1: 49–97.

Ekman, P. and Friesen, W.V. (1971) 'Constants across cultures in the face and emotions', *Journal of Personality and Social Psychology*, 17: 124–9.

Ekman, P. and Friesen, W.V. (1978) *The Facial Action Coding System*, Palo Alto, CA: Consulting Psychologists Press.

Ekman, P. and Friesen, W.V. (1982) 'Rationale and Reliability for EMFACS Coders', unpublished manuscript.

Ekman, P. and Rosenberg, E.L. (1997) *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS)*, New York: Oxford University Press.

Elizondo, D. (2006) 'The linear separability problem: Some testing methods', *IEEE Transactions on Neural Networks*, 17: 330–44.

Fasel, B. and Luettin, J. (2003) 'Automatic facial expression analysis: A survey', *Pattern Recognition*, 36: 259–75.

Flavell, J.H., Miller, P.H. and Miller, S.A. (2002) *Cognitive Development*, 4th edn, Englewood Cliffs, NJ: Prentice Hall.

Fletcher, L., Apostoloff, N., Chen, J. and Zelinsky, A. (2001) 'Computer vision for vehicle monitoring and control', *Proceedings of the 2001 Australian Conference on Robotics and Automation* (pp. 67–72), November 14–16, Sydney, Australia.

Fox, M. and Spencer, B. (eds) (2006) *Proceedings of the 8th International Conference on Electronic Commerce: The New E-commerce – Innovations for Conquering Current Barriers, Obstacles and Limitations to Conducting Successful Business on the Internet* (pp. 463–70), August 13–16, Fredericton, New Brunswick, Canada, ACM.

Frank, M.G. (1997) 'Some thoughts on FACS, dynamic markers of emotion, and baseball', in P. Ekman and E.L. Rosenberg (eds) *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS)* (pp. 239–42), New York: Oxford University Press.

Freund, Y. and Schapire, R. (1996) 'Experiments with a new boosting algorithm', *Proceedings of the Thirteenth International Conference on Machine Learning*, (pp. 148–56), July 3–6, Bari, Italy.

Fridlund, A.J. (1994) *Human Facial Expression: An Evolutionary View*, New York: Academic Press.

Friedman, J., Hastie, T. and Tibshirani, R. (2000) 'Additive logistic regression: a statistical view of boosting (with discussion)', *Annals of Statistics*, 28: 337–407.

Gilbert, D.T. (1991) 'How mental systems believe', *American Psychologist*, 46: 107–19.

Gottman, J.M. and Levenson, R.W. (1992) 'Marital processes predictive of later dissolution: Behavior, physiology, and health', *Journal of Personality and Social Psychology*, 63: 221–33.

Gottman, J.M. and Notarius, C.I. (2000) 'Decade review: Observing marital interaction', *Journal of Marriage and the Family*, 62: 927–47.

Greenwald, A.G. and Leavitt, C. (1984) 'Audience involvement in advertising: Four levels', *Journal of Consumer Research*, 11: 581–92.

Izard, C.E. (1979) *The Maximally Discriminative Facial Movement Coding System (MAX)*, Newark, DE: University of Delaware Office of Instructional Technology.

Izard, C.E. (1990) *The Psychology of Emotions*, New York: Plenum.

Izard, C.E. (1997) 'Emotions and facial expressions: A perspective from differential emotions theory', in J.A. Russell and J.M. Fernandez-Dols (eds) *The Psychology of Facial Expression* (pp. 57–77), New York: Cambridge University Press.

Izard, C.E. (2009) 'Emotion theory and research: Highlights, unanswered questions, and emerging issues', *Annual Review of Psychology*, 60: 1–25.

Jakobs, E., Manstead, A.S.R. and Fischer, A.H. (1996) 'Social context and the experience of emotion', *Journal of Nonverbal Behavior*, 20: 123–42.

Kapoor, A., Burleson, W. and Picard, R. (2007) 'Automatic prediction of frustration', *International Journal of Human-Computer Studies*, 65: 724–36.

Knobloch, S. and Zillmann, D. (2002) 'Mood management via the digital jukebox', *Journal of Communication*, 52: 351–66.

Konijn, E.A. and Van Vugt, H.C. (2008) 'Emotions in mediated interpersonal communication: Toward modeling emotion in virtual humans', in E.A. Konijn, S. Utz, M. Tanis, and S. Barnes (eds) *Mediated Interpersonal Communication* (pp. 100–130), New York: Routledge.

Kurihata, H., Takahashi, T., Ide, I., Mekada, Y., Murase H., Tamatsu, Y. and Miyahara, T. (2005) 'Rainy weather recognition from in-vehicle camera images for driver assistance', *Proceedings of the IEEE Intelligent Vehicles Symposium* (pp. 205–10), June, Las Vegas, IEEE Press.

LeDoux, J.E. (1996) *The Emotional Brain*, New York: Simon & Schuster.

Lerner, H., Singer, J. and Huey, R. (2008) *Driver Strategies for Engaging in Distracting Tasks Using In-Vehicle Technologies. (National Highway Traffic Safety Administration Report DOT HS 810 919)*, Washington, DC: U.S. Department of Transportation.

Madsen, M., el Kaliouby, R., Goodwin, M., and Picard, R.W. (2008) '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, Halifax, Canada.

Manstead, A.S.R. and Fischer, A.H. (2002) 'Beyond the universality-specificity dichotomy', *Cognition and Emotion*, 16: 1–9.

Matsumoto, D. (1987) 'The role of facial response in the experience of emotion: More methodological problems and a meta-analysis', *Journal of Personality and Social Psychology*, 52: 769–74.

Matsumoto, D. and Willingham, B. (2009) 'Spontaneous facial expressions of emotion in congenitally and non-congenitally blind individuals', *Journal of Personality and Social Psychology*, 96: 1–10.

Meservy, T.O., Jensen, M.L., Kruse, J., Burgoon, J.K. and Nunamaker Jr., J.F. (2005) 'Deception detection through automatic, unobtrusive analysis of nonverbal behavior', *IEEE Intelligent Systems*, 20: 36–43.

Meservy, T.O., Jensen, M.L., Kruse, W.J., Burgoon, J.K. and Nunamaker Jr., J.F. (2008) 'Automatic extraction of deceptive behavioral cues from video', in H. Chen, E. Reid, J. Sinai, A. Silke and B. Ganor (eds) *Terrorism Informatics* (pp. 495–516), New York: Springer.

Miller, G.R. (1987) 'Persuasion', in C.R. Berger and S.H. Chaffee (eds) *Handbook of Communication Science* (pp. 446–83), Newbury Park, CA: Sage.

Nass, C. and Brave, S. (2005) *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*, Cambridge, MA: MIT Press.

Osgood, C.E. (1953) *Method and Theory in Experimental Psychology*, New York: Oxford University Press.

Panksepp, J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions*, New York: Oxford University Press.

Panksepp, J., LeDoux, J.E., Phelps, E.A., Cacioppo, J.T., Berntson, G.G., Larsen, J.T., et al. (2000) 'Part II: Biological and neurophysiological approaches to emotion', in M. Lewisand and J.M. Haviland-Jones (eds), *Handbook of Emotions*, 2nd edn (pp. 137–249), New York: Guilford.

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

Picard, R.W. and Daily, S.B. (2005) 'Evaluating affective interactions: Alternatives to asking what users feel', paper presented at *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*, April 2–7, Portland, OR.

Ranney, T. (2008) *Driver Distraction: A Review of the Current State-of-Knowledge. (National Highway Traffic Safety Administration Report DOT HS 810 787)*, Washington, DC: U.S. Department of Transportation.

Reynolds, C. and Picard, R.W. (2005) 'Evaluation of affective computing systems from a dimensional metaethical position', *1st Augmented Cognition Conference in conjunction with the 11th International Conference on Human-Computer Interaction*, July 22–27, Las Vegas, NV.

Ruch, W. (1997) 'Will the real relationship between facial expression and affective experience please stand up: The case of exhilaration', in P. Ekman and E.L. Rosenberg (eds) *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS)* (pp. 89–111), New York: Oxford University Press.

Russell, J.A. (1997) 'Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective', in J.A. Russell and J.M. Fernandez-Dols (eds) *The Psychology of Facial Expression* (pp. 295–320), New York: Cambridge University Press.

Russell, J.A. and Fernandez-Dols, J.M. (eds) (1997) *The Psychology of Facial Expression*, New York: Cambridge University Press.

Sayette, M.A., Cohn, J.F., Wertz, J.M., Perrott, M.A. and Parrott, D.J. (2001) 'A psychometric evaluation of facial action coding system for assessing spontaneous expression', *Journal of Nonverbal Behavior*, 25: 167–85.

Sood, S. and Forehand, M. (2005) 'On self-referencing differences in judgment and choice', *Organizational Behavior and Human Decision Processes*, 98: 144–54.

Spitzberg, B.H. and Cupach, W.R. (1984) *Interpersonal Communication Competence Research*, New York: Springer.

Springer, K. (2001) 'Perceptual boundedness and perceptual support in conceptual development', *Psychological Review*, 108: 691–708.

Steggell, G.L. and Harper, J.M. (1991) 'Family interaction patterns and communication processes', in S.J. Bahr (ed.) *Family Research: A Sixty-year Review, 1930–1990*, vol. 1 (pp. 97–170), New York: Lexington.

Suzuki, Y., Fujii, T. and Tanimoto, M. (2005) 'Parking assistance using multi-camera infrastructure', *Proceedings IEEE Intelligent Vehicles Symposium, IEEE Press* (pp. 106–10), June 6–8.

Trope, Y. (1986) 'Identification and inferential processes in dispositional attribution', *Psychological Review*, 93: 239–57.

Wagner, H.L. (1997) 'Methods for the study of facial behavior', in J.A. Russell and J.M. Fernandez-Dols (eds) *The Psychology of Facial Expression* (pp. 31–54), New York: Cambridge University Press.

Webb, E.J., Campbell, D.T., Schwartz, R.D. and Sechrest, L. (2000) *Unobtrusive Measures*, Thousand Oaks, CA: Sage.

Williamson, A. and Chamberlain, T. (2005) Review of on-road driver fatigue monitoring devices, NSW Injury Risk Management Research Centre, April, University of New South Wales.

Winton, W.M., Putnam, L.E. and Krauss, R.M. (1984) 'Facial and autonomic manifestations of the dimensional structure of emotion', *Journal of Experimental Social Psychology*, 20: 195–216.

Zaalberg, R., Manstead, A.S.R. and Fischer, A.H. (2004) 'Relations between emotions, display rules, social motives, and facial behavior', *Cognition and Emotion*, 18: 182–207.

Zaichkowsky, J.L. (1985) 'Measuring the involvement construct', *Journal of Consumer Research*, 12: 341–52.

Zajonc, R.B. (1998) 'Emotion', in D. Gilbert, S. Fiske, and G. Lindzey (eds) *Handbook of Social Psychology*, vol. 2 (pp. 591–632), 4th edn, Boston: McGraw-Hill.

Zillmann, D. (1988a) 'Mood management through communication choices', *American Behavioral Scientist*, 31: 327–40.

Zillmann, D. (1988b) 'Mood management: Using entertainment to full advantage', in L. Donohew, H.E. Sypher and E.T. Higgins (eds) *Communication, Social Cognition, and Affect* (pp. 147–71), Hillsdale, NJ: Erlbaum.