

# Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications

Rafael A. Calvo, *Senior Member, IEEE*, and Sidney D'Mello, *Member, IEEE Computer Society*

**Abstract**—This survey describes recent progress in the field of Affective Computing (AC), with a focus on affect detection. Although many AC researchers have traditionally attempted to remain agnostic to the different emotion theories proposed by psychologists, the affective technologies being developed are rife with theoretical assumptions that impact their effectiveness. Hence, an informed and integrated examination of emotion theories from multiple areas will need to become part of computing practice if truly effective real-world systems are to be achieved. This survey discusses theoretical perspectives that view emotions as expressions, embodiments, outcomes of cognitive appraisal, social constructs, products of neural circuitry, and psychological interpretations of basic feelings. It provides meta-analyses on existing reviews of affect detection systems that focus on traditional affect detection modalities like physiology, face, and voice, and also reviews emerging research on more novel channels such as text, body language, and complex multimodal systems. This survey explicitly explores the multidisciplinary foundation that underlies all AC applications by describing how AC researchers have incorporated psychological theories of emotion and how these theories affect research questions, methods, results, and their interpretations. In this way, models and methods can be compared, and emerging insights from various disciplines can be more expertly integrated.

**Index Terms**—Affective computing, affect sensing and analysis, multimodal recognition, emotion detection, emotion theory.

## 1 INTRODUCTION

INSPIRED by the inextricable link between emotions and cognition, the field of Affective Computing (AC) aspires to narrow the communicative gap between the highly emotional human and the emotionally challenged computer by developing computational systems that recognize and respond to the affective states (e.g., moods and emotions) of the user. Affect-sensitive interfaces are being developed in a number of domains, including gaming, mental health, and learning technologies. The basic tenet behind most AC systems is that automatically recognizing and responding to a user's affective states during interactions with a computer can enhance the quality of the interaction, thereby making a computer interface more usable, enjoyable, and effective. For example, an affect-sensitive learning environment that detects and responds to students' frustration is expected to increase motivation and improve learning gains when compared to a system that ignores student affect [1].

Although scientific research in the area of emotion stretches back to the 19th century when Charles Darwin [2], [3] and William James [4] proposed theories of emotion that continue to influence thinking today, the injection of affect into computer technologies is much more recent. During most of the last century, research on emotions was conducted by philosophers and psychologists, whose work

was based on a small set of "emotion theories" that continue to underpin research in this area. However, in the 1980s, researchers such as Turkle [5] began to speculate about how computers might be used to study emotions. Systematic research programs along this front began to emerge in the early 1990s. For example, Scherer [6] implemented a computational model of emotion as an expert system. A few years later, Picard's landmark book *Affective Computing* [7] prompted a wave of interest among computer scientists and engineers looking for ways to improve human-computer interfaces by coordinating emotions and cognition with task constraints and demands.

Picard described three types of affective computing applications: 1) systems that detect the emotions of the user, 2) systems that express what a human would perceive as an emotion (e.g., an avatar, robot, and animated conversational agent), and 3) systems that actually "feel" an emotion. Although the last decade has witnessed astounding research along all three fronts, the focus of this survey is on the theories, methods, and data that support *affect detection*. Affect detection is critical because an affect-sensitive interface can never respond to users' affective states if it cannot sense their affective states. Affect detection need not be perfect but must be approximately on target. Affect detection is, however, a very challenging problem because emotions are constructs (i.e., conceptual quantities that cannot be directly measured) with fuzzy boundaries and with substantial individual difference variations in expression and experience.

Many have argued that relying solely on existing methods to develop computer systems with a new set of "affect-sensitive" functionalities would be insufficient [8] because user emotions are not currently on the radar of computing methods. This is where insights gleaned from a century and a half of scientific study on human emotions

• R.A. Calvo is with the School of Electrical and Information Engineering, The University of Sydney, Sydney NSW 2006, Australia.  
E-mail: Rafael.Calvo@sydney.edu.au.

• S. D'Mello is with the University of Memphis, 202 Psychology Building, Memphis, TN 38152. Email: sdmello@memphis.edu.

Manuscript received 11 Feb. 2010; revised 27 May 2010; accepted 6 July 2010; published online 19 July 2010.

Recommended for acceptance by B. Parkinson.

For information on obtaining reprints of this article, please send e-mail to: [taffc@computer.org](mailto:taffc@computer.org), and reference IEEECS Log Number TAFFC-2010-02-0008.

Digital Object Identifier no. 10.1109/T-AFFC.2010.1.

becomes especially useful for the successful development of affect-sensitive interfaces. Blending scientific theories of emotion with the practical engineering goal of developing affect-sensitive interfaces requires a review of the literature that contextualizes the engineering goals within the psychological principles. This is a major goal of this review.

### 1.1 Previous Surveys and Other Resources

The strong growth in AC is demonstrated by an increasing number of conferences and journals. IEEE's launch of the *Transactions on Affective Computing* is just one example. In 2009, two conferences were dedicated to AC: the International Conference on Affective Computing and Intelligent Interaction (bi-annual) and Artificial Intelligence in Education (affect was the theme of the 2009 conference). These journals and conferences describe new affect-aware applications in the areas of gaming, intelligent tutoring systems, support for autism spectrum disorders, cognitive load, and many others.

The research literature related to building affect-aware computer applications has been surveyed previously. These surveys have very different foci, as is expected in a highly interdisciplinary field that spans psychology, computer science, engineering, neuroscience, education, and many others. A review on affective computing was published in the First International Conference on Affective Computing and Intelligent Interaction [9]. Techniques specific to a single data type have been surveyed by experts in each specific modality. Pantic and Rothkrantz [10] and Sebe et al. [11] provide reviews of multimodal approaches (mainly face and speech) with brief discussions of data fusion strategies. Zeng et al. [12] reviewed the literature on audiovisual recognition approaches focusing on spontaneous expressions, while others [10], [13] focused on reviewing work on posed (e.g., by actors) emotions. *Sentiment analysis*, a set of text-based approaches to analyzing affect and opinions that we have included as part of the affective computing literature was reviewed in [14]. A number of new books on Affective Computing are being published, including [15], [16], [17], [18].

The psychological literature related to AC is probably the most extensive. Journals such as *Emotion*, *Emotion Review*, and *Cognition and Emotion* provide an outlet for reviews, and are good sources for researchers with an interest in emotion. Recent reviews of emotion theories include the integrated review of several existing theories (and the presentation of his core affect theory) by Russell [19], a review of facial and vocal communication of emotion also by Russell [20], a critical review on the theory of basic emotions and a review on the experience of emotion by Barrett [21], [22], and a review of affective neuroscience by Dalgleish et al. [23].

The *Handbook of Cognition and Emotion* [24], the *Handbook of Emotion* [25], and the *Handbook of Affective Sciences* [26] provide broad coverage on affective phenomena. Other resources include the International Society for Research on Emotions (ISRE) and the HUMAINE association (a Network of Excellence in the EU's Sixth Framework Programme).

### 1.2 Goals, Scope, and Overview of Present Survey

Despite the extensive literature in emotion research and affective science [26], the AC literature has been primarily

driven by computer scientists and AI researchers who have remained agnostic to the controversies inherent in the underlying psychological theory. Instead, they have focused their efforts on the technical challenges of developing affect-sensitive computer interfaces. However, ignoring the important debates has significant limitations because a functional AC application can never be completely divorced from underlying emotion theory.

It is not only informative, but arguably essential for computer scientists to more expertly integrate emotion theory into the design of AC systems. Effective design of these systems will rely on cross-disciplinary collaboration and an active sharing of knowledge. New perspectives on emotion research are surfacing in multiple disciplines from psychology and neuroscience to engineering and computing. Innovative research continues to tackle the perennial questions of understanding what emotions are as well as questions pertaining to accurate detection of emotions. Therefore, the aim of this review is to present to engineers and computer scientists a more complete picture of the affective research landscape. We also aim for this survey to be a useful resource to new researchers in affective computing by providing a taxonomy of resources for further exploration and discussing different theoretical viewpoints and applications.

The review needs to be sufficiently broad in order to encompass an updated taxonomy of emotion research from different fields as well as current work within the AC area. Although it is not possible to exhaustively cover a century of scientific research on emotion, we hope that this survey will illuminate some of the key literature and highlight where computer scientists can best benefit from the knowledge in this field. It is important to note that computer scientists and engineers can also influence the emotion research literature.

This review can be distinguished from previous reviews by its inclusion of a survey of multiple theoretical models, methods, and applications for AC rather than an exclusive focus on affect detection, and by the breadth of multimodal affect detection research reviewed (i.e., other channels in addition to the face and speech).

This survey is structured as a sequence of sections, with the scope being progressively narrowed from emotion theories to practical AC applications. Section 2 reviews emotion theories and how they have been used in AC. In order to provide a complete picture, this survey reviews models that emphasize emotions as expressions, embodiments, outcomes of cognitive appraisals, social constructs, and products of neural circuitry. The study of human emotions is the longest lived antecedent to AC. Our goal is to introduce the major theories that have emerged and make reference to key pieces of literature as a guide for further reading.

In Section 3, we focus on affect detection, particularly describing how different modalities (i.e., signals) and algorithms have been used. We review studies that use physiological signals, neuroimaging, voice, face, posture, and text processing techniques. We also discuss some of the precious few multimodal approaches to affect detection. Computational models are very much dependent on the

types of data used and their statistical properties and features, such as time resolution impact on the user and usability, etc. These different types of data and the techniques used to process them are reviewed in Section 3. Each technique presents its own challenges and comes with its own constraints (e.g., time resolution) that may limit its utility for affect detection. Each of these techniques has their own distinct research communities (i.e., signal processing versus natural language processing), so this review aims to provide researchers that seek multimodal approaches with landmark techniques that have been applied to AC.

We conclude by discussing areas where emotion theory and affective computing diverge, areas where these fields should converge, and proposing areas that are ripe for future research.

## 2 MODELING AFFECT

Affective phenomena that are relevant to AC research include emotions, feelings, moods, attitudes, affective styles, and temperament. We focus on “emotion,” and consider six perspectives that have been used to describe this complex, fuzzy, indeterminate, and elusive, yet universal, essential, and exciting scientific construct. The first four perspectives (expressions, embodiments, cognitive appraisal, and social constructs) are derived from traditional emotion theories that have been dominant for several decades. These should now be extended with theoretical contributions from the affective neurosciences. We also include a sixth theoretical front pioneered by Russell because it integrates the different perspectives and has been influential to many AC researchers. For each perspective, we provide an introduction and cite key (but nonexhaustive) literature along with examples on how the theory has been used or has influenced AC research.

### 2.1 Emotions as Expressions

Darwin was the first to scientifically explore emotions [2]. He noticed that some facial and body expressions of humans were similar to those of other animals, and concluded that behavioral correlates of emotional experience were the result of evolutionary processes. In this respect, evolution was probably the first scientific framework to analyze emotions. An important aspect of Darwin’s theory was that emotion expressions (e.g., a disgusted face) that he called “serviceable associated habits” did not evolve for the sake of expressing an emotion, but were initially associated with other more essential actions (e.g., a disgusted face initially associated with rejecting an offensive object from consumption eventually accompanies disgust even in the absence of such an object).

Although the Darwinian theory of emotion fails to explain a number of emotional behaviors and expressions, there is some evidence that six or seven facial expressions of emotions are universally recognized; however, some researchers have challenged this view [20], [27], [28]. The existence of universal expressions for some emotions has been interpreted as an indication that these six emotions are “basic” in the sense that they are innate and cross-cultural boundaries [29], [30], [31], [32], [33].

Other researchers have expanded Darwin’s evolutionary framework toward other forms of emotional expression.

Frijda [34] studied “action tendencies” or states of readiness to act in a particular way when confronted with an emotional stimulus. These action tendencies were linked to our need to solve the problems that we find in our environment. For example, the action tendency “approach” permits the consumption of something “wanted” and it would be associated with the emotion normally called “desire.” On the other hand, the purpose of “avoidance” is to protect and would often be linked to what is called “fear.”

Inspired by the *emotion as expressions* view and informed by considerable research that identifies the facial correlates of emotion, AC systems that use facial expressions for affect detection are increasingly common (see Section 3.1). Camera-based facial expression detection also enjoys some advantages because it is nonintrusive (in the sense that it does not involve attaching sensors to a user), has reasonable accuracy, and current systems do not require expensive hardware (e.g., most laptops have Webcams that can be used for monitoring facial movements). Other AC systems use alternate bodily channels such as voice, body language and posture, and gestures; these are described in detail in Section 3.

### 2.2 Emotions as Embodiments

James proposed a model that combined expression (as Darwin) and physiology, but interpreted the perception of physiological changes as the emotion itself rather than its expression. James’ theory is summarized by a quote from James:

If we fancy some strong emotion, and then try to abstract from our consciousness of it all the feelings of its characteristic bodily symptoms, we find that we have nothing left behind [4].

Lange, a contemporary of James and a pioneer of psychophysiology, studied the bodily manifestations that accompany some of the common emotions (e.g., sorrow, joy, anger, and fright). The “James-Lange theory” focused on emotion being “changes” in the Sympathetic Nervous System (SNS) a part of the Autonomic Nervous System (ANS).

The James and Lange theories emphasize that emotional experience is embodied in peripheral physiology. Hence, AC systems can detect emotions by analyzing the pattern of physiological changes associated with each emotion (assuming a prototypical physiological response for each emotion exists).

The amount of information that the physiological signals can provide is increasing, mainly due to major improvements in the accuracy of psychophysiology equipment and associated data analysis techniques. Still, physiological signals are currently recorded using equipment and techniques that are more intrusive than those recording facial and vocal expression. Fortunately, some of the challenges associated with deploying intrusive physiological sensing devices in real-world contexts are being mitigated by recent advances in the design of wearable sensors (e.g., [35]). Even more promising is the integration of physiological sensors with equipment that records the widely varying pattern of actions associated with emotional experience. When statistical learning methods are applied to these vast repositories of affective data, they can be used to find patterns associated with different action tendencies in different contexts or individuals, linking them to

corresponding differences in autonomic changes that are associated with these actions.

### 2.3 Cognitive Approaches to Emotions

Arnold [36] is considered to be the pioneer of the cognitive approach to emotions, which is probably the leading view of emotion in cognitive psychology. Cognitivists believe that in order for a person to experience an emotion, an object or event must be appraised as directly affecting the person in some way, based on a person's experience, goals, and opportunity for action [37], [38], [39]; see the classic Schachter-Singer experiment [40]. Appraisal is a presumably unconscious process that produces emotions by evaluating an event along a number of dimensions such as novelty, urgency, ability to cope, consistency with goals, etc.

The cognitivist theories of emotion have been extensively described in the *Handbook of Cognition and Emotion* [24], so only some of the work used by AC researchers is discussed here. The work of psychologists such as Bower, Mandler, Lazarus, Roseman, Ortony, Scherer, and Frijda has been most influential in the AC community.

The cognitive-motivational-relational theory [41] states that in order to predict how a person will react to a situation, the person's expectations and goals in relation to the situation must be known. The theory describes how specific emotions arise out of personal conceptions of a situation. Roseman et al. [39], [42] have developed structural theories where a set of discrete emotions is modeled as direct outcomes of a multidimensional appraisal process. Roseman's 14 emotions are associated with a cognitive appraisal process that can be modeled with five dimensions [39], [42]:

1. *Consistency motives*. Also referred to as the "beneficial/harmful" variable by Arnold (1960), this pertains to an appraisal of how well an affect-inducing event helps or hinders one's intentions.
2. *Probability*. This dimension refers to the certainty that an event will actually occur. For example, a probable and negative event will be appraised differently than an improbable one.
3. *Agency*. This refers to the entity (i.e., self or other) that produces or is responsible for the event. For example, if a negative event is caused by oneself, it would be appraised differently than one caused by someone else.
4. *Motivational state*. An event can be "appetitive" (an event leading to a reward) or "aversive" (one leading to punishment).
5. *Power*. Refers to whether the subject is (or feels) in control of the situation or not.

The cognitive theory by Ortony, Clore, and Collins (OCC) views emotions as reactions to situational appraisals of events, actors, and objects [43]. The emotions can be positive or negative depending on the desirability of the situation. They identified four sources of evidence that can be used to test emotion theories: language, self-report, behavior, and physiology. The goal of much of their work was to create a computationally tractable model of emotion.

It is important to note that Roseman's and Ortony's models are similar in many ways. They both converge upon

the "universality" of the appraisal process. These models can be used to automatically predict a user's emotional state by taking a point in the multidimensional appraisal space (think of each contextual feature or appraisal variable as a dimension) and returning the most probable emotion.

Despite its success, the appraisal theories are not able to explain a number of questions that are discussed in detail elsewhere [44]. Briefly, these include issues such as

1. The minimal number of appraisal criteria needed to explain emotion differentiation.
2. Should appraisal be considered a process, an ongoing series of evaluations, or separate emotion episodes.
3. How can the social functions of emotion be considered within the cognitive approach?
4. How do appraisal theories explain evidence that preferences (simple emotional reactions) do not need any conscious registration, or that emotions can exist without cognition [45], [46]?

Despite these questions and limitations, the computational models derived from appraisal theories have proven to be useful to Artificial intelligence (AI) researchers and psychologists by providing both groups with a way to evaluate, modify, and improve their theories. Computational models of emotion had early success exemplified by Scherer's GENESE expert system. GENESE was built on a knowledge base that mapped appraisals to different emotions as predicted by Scherer's heuristics and theory [6]. The system was successful in achieving a high accuracy (77.9 percent) at predicting the target emotions for 286 emotional episodes described by subjects as a series of responses to situational questions. Scherer's model made assumptions about what emotions are (i.e., one of 14 labels), what an emotional event is (i.e., the sequence of events that lead to the appraisal), and the role of the computer system (i.e., to provide a classification label and assess its accuracy). Despite its value as a tool to test emotion theories, the system was too limited for most real-world applications. It is too difficult, if not impossible, to construct the significant knowledge base needed for complex realistic situations.

More recently, the OCC model has been used by Conati [47] in his work on emotions during learning. The model is used to combining evidence from situational appraisals that produce emotions (top-down) with bodily expressions associated with emotion expression (bottom-up). In this fashion, both a predictive (from the OCC model) and a confirmatory (from bodily measures) analysis of emotional experience is feasible (this work is discussed in detail in Section 4.1).

### 2.4 Emotions as Social Constructs

Averill put forward the idea that emotions cannot be explained strictly on the basis of physiological or cognitive terms. Instead, he claimed that emotions are primarily social constructs; hence, a social level of analysis is necessary to truly understand the nature of emotion [48]. The relationship between emotion and language [49] and the fact that the language of emotion is considered a vital part of the experience of emotion has been used by social constructivists and anthropologists to question the "universality" of Ekman's studies, arguably because the

language labels he used to code emotions are somewhat US-centric. In addition, other cultures might have labels that cannot be literally translated to English (e.g., some languages do not have a word for fear [19]).

Researchers like Darwin only occasionally mentioned the social function of emotional expressions, but contemporary researchers from social psychology have highlighted the importance of social processes in explaining emotional phenomena [50]. For example, Salovey [51] describes how social processes influence emotional experiences due to adaptation (adjustments in response to the environment), social coordination (reactions in response to expressions by others), and self-regulation (reactions based on our understanding of our own emotional state and relationship with the environment).

Stets and Turner [50] have recently reviewed five research traditions pertaining to the sociology of emotions that have emerged in the literature; two of these (dramaturgical and structural approaches) are discussed here (see [50] for more details). *Dramaturgical approaches* emphasize the importance of cultural guidelines in the experience and expression of emotions [52], [53]. For example, cultural norms play a significant role in specifying when we *feel* and how we *express* emotions such as grief and sympathy (i.e., display rules [54], [55]).

Alternatively, *structural approaches* focus on social structure when analyzing emotions. Along these lines, Kemper's power-status theory considers *power* and *status* in society as the two particularly relevant dimensions [56]. Gain in power is diagnostic of confidence and satisfaction, while a loss in power or an unexpected failure to gain power is predictive of increased anxiety and fear and loss of confidence. Similarly, an expected gain in status is associated with well-being and positive emotions. A loss in status will trigger negative emotions, the nature and intensity of which are based on how the loss of status is appraised by the individual.

With a few exceptions, such as Boehner et al.'s [8] Affecter system, AC researchers have preferred the first three perspectives of emotion (expressions, embodiments, and cognitive appraisals), over the social constructivist or "interactionist" perspectives; this is an important limitation that will be addressed in Section 4.

## 2.5 Neuroscience

Until quite recently, the study of emotional phenomena was divorced from the brain. But, in the last two decades, neuroscience has contributed to the study of emotion by proposing new techniques and methods to understand emotional processes and their neural correlates. In particular, the field of affective neuroscience [23], [26], [57], [58] is helping us understand the neural circuitry that underlies emotional experience, the etiology of certain mental health pathologies, and is offering new perspectives pertaining to the manner in which emotional states (and their development) influence our health and life outcomes.

The methods used in affective neuroscience include imaging (e.g., fMRI), lesion studies, genetics, and electrophysiology. A key contribution in the last two decades has been to provide evidence against the notion that emotions are subcortical and limbic, whereas cognition is cortical. This notion was reinforcing the flawed Cartesian dichotomy

between thoughts and feelings [58]. There is now ample evidence that the neural substrates of cognition and emotion overlap substantially [23]. Cognitive processes such as memory encoding and retrieval, causal reasoning, deliberation, goal appraisal, and planning operate continually throughout the experience of emotion. This evidence points to the importance of considering the affective components of any human-computer interaction.

Affective neuroscience has also provided evidence that elements of emotional learning can occur without awareness [59] and elements of emotional behavior do not require explicit processing [60]. Despite some controversy about these two findings (cf. [61]), specifically that they might only apply to specific types of emotional stimuli, this work is of great importance to AC research. It suggests that some emotional phenomena might not be consciously experienced (yet influence memory and behavior), and provides further evidence that studies based solely on self-reports of emotion might not reflect more subtle phenomena that do not make it to consciousness.

Neuroscientific methods have also provided an alternative to self-reports or facial expressions as a data source for parsing emotional processes. Although not often used in the AC literature, EEG-based techniques are being increasingly used [62], [63]. Other studies in affect detection using EEG are discussed in Section 3.2.

Although neuroscience contributions have run in parallel with cognitive theories of emotion [64], increasing collaboration has been fruitful. For example, Lewis recently proposed a framework grounded in dynamical systems theory to integrate low-level affective neuroscience with high-level appraisal theories of emotion.

The numerous perspectives on conceptualizing emotions are being further challenged by emerging neuroscience evidence. Some of this evidence [65] challenges the common view that an organizing neural circuit in the brain is the reason that indicators of emotion covary. The new evidence, together with progress in complex systems theory, has increased the interest in emergent variable models where emotions do not cause, but rather are caused by, the measured indicators of emotion [65], [66].

## 2.6 Core Affect and Psychological Construction of Emotion

Russell and others have argued that the different emotion theories are different because they actually talk about different things, not necessarily because they are presenting different positions of the same concept. He posited the idea that if "emotion" is actually a "*heterogeneous cluster of loosely related events, patterns and dispositions, then these diverse theories might each concern a somewhat different subset of events or different aspects of those events*" [19].

Russell proposed a framework that could help bridge the gaps between the different emotion theories. Although Russell's theory introduces a number of important terms and concepts that cannot be elaborated here, his major contributions include

1. an emotion theory centered around "core affect" (i.e., a consciously accessible neurophysiological state described as a point in a valence (pleasure-displeasure) and arousal (sleepy-activated) space, somewhat similar to what is often called feeling,

2. the importance of context and separating emotional episodes (i.e., a particular episode of anger) from emotion categories (i.e., the anger class),
3. a realization that emotions are not produced by "affect programs" but instead consist of loosely coupled components, such as physiological responses, bodily expressions, appraisals, etc., and
4. a way to unite "categorical" models that use "folk" terminology to define "emotions" with dimensional models.

Yet another significant idea in the theory is that the various components that underlie an emotional episode (i.e., appraisals, attributions, etc.) need not be coherent. In fact, most often they are not, yet the person is generally able to make sense of these different sources of information. When the components are coherent, the person finds what Russell called the "prototypical case," those that a layman would call "emotion."

### 3 AFFECT DETECTION: SIGNALS, DATA, AND METHODS

The foci of emotion research (or affective science) and affective computing are often different, arguably the former being the human and the latter being the computer (although successful AC research should view the human and computer as one interacting system). In addition, in the areas of common interest (e.g., affect expression and detection by humans and computers), disciplinary differences are still to be surpassed before a more coherent body of literature can be developed. In this section, we discuss the AC approaches to affect detection. In the spirit of fostering interdisciplinary discussions between the emotion theorists and the AC practitioners, we link each approach to the emotion theory that most closely supports it, knowing full well that, in many cases, there is no clear one-to-one mapping between AC approach and emotion theory.

The review of current affect detection systems is organized with respect to individual modalities or channels (e.g., face, voice, and text). Each modality has advantages and disadvantages toward its use as a viable affect detection channel. Some of the factors affecting the value of each modality include:

1. the validity of the signal as a natural way to identify an affective state,
2. the reliability of the signals in real-world environments,
3. the time resolution of the signal as it relates to the specific needs of the application, and
4. the cost and intrusiveness for the user.

Each modality has its own body of literature, and extensively reviewing this literature would be beyond the scope of this survey. Hence, we provide a broad overview of the research in each modality and focus on the major accomplishments and open issues. We also provide a synthesis of existing reviews if they are available.

#### 3.1 Facial Expressions

Inspired by the "emotions as expressions" view described in Section 2, perhaps the majority of affect detection

research has focused on detecting the basic emotions from the face. Recall that this view posits that there is a distinctive facial expression associated with each basic emotion [30], [67], [68]. The expression is triggered for a short duration when the emotion is experienced, so detecting an emotion is simply a matter of detecting its prototypical facial expression.

Ekman and Friesen [69] developed the Facial Action Coding System (FACS) to measure facial activity using objective component facial motions or "facial actions" as a method to identify emotion. They identified a set of action units (AUs) that identify independent motions of the face. Trained human coders decompose an expression into a set of AUs, and this coding technique has become the leading method for objectively classifying facial expressions in the behavioral sciences. The classified expressions are then linked to the six basic emotions: anger, disgust, fear, joy, sadness, and surprise [67].

Manually coding a segment of video is an expensive task. It requires specially trained coders that spend approximately 1 hour for each minute of video [70]. A number of researchers have tried to automate this task and some of their work is listed in Table 1. It should be noted, however, that the development of a system that automatically detects the action units is quite a challenging task because the coding system was originally created for static pictures rather than for changing expressions over time. Although there has been remarkable progress in this area, the reliability of current automatic AU detection systems does not match humans [71], [72], [73]. The evaluation of these systems is normally done through "1st" person reports (the subjects themselves) or "3rd" person reports (e.g., experts or peers). These are indicated as "1st" and "3rd" in the tables below.

Despite the challenges, some important progress is being made toward the development of fully automated face-based affect detection. Peter Robinson and his team have worked on the use of facial expressions for affect detection [74], [75]. Other projects in the area of ITS aim to use facial expression recognition techniques to improve the interaction between students and learning systems [35], [76]. For example, Arroyo et al. [35] showed how features from a student's facial expressions (recorded with a Webcam and processed by third party software) together with physiological activity can predict students' affective states.

Zeng et al. [12] recently reviewed 29 state-of-the-art vision-based affect detection methods; hence, an extensive review will not be presented here. However, a meta-analysis of the 29 systems reviewed provided some important insights pertaining to current vision-based affect detection systems. First, almost all systems were concerned with detecting the six basic emotions, irrespective of whether these emotions are relevant to AC applications. Second, approximately half of the systems relied on data sets with posed facial expressions. This has important implications for applicability in real-world scenarios. Third, and of more concern, is the fact that only 6 out of the 29 systems could operate in real time, which is an obvious requirement for practical AC applications. Fourth, the vast majority of the systems required presegmented emotion expressions rather than naturalistic video sequences.

TABLE 1  
Facial Expressions Analysis for Affect Recognition

Study	Signal	Classifier	Model	Evaluation	Stimulus
[75]	Face expressions and head movement	DBN (real time)	6 categories	3rd: 10 annotators	Acting
[78]	Face	Gabor wavelets/ SVM, AdaBoost, LDA	20 AU	3rd (2 FACS coders)	Self (false-opinion + other)
[79]	Face	temporal rules	27 AU	3rd: 2 FACS coders	Self
[80]	Face, body	C4.5, BN (fusion)	6 categories	1	Self
[76]	Face	33 AU / DA	6 categories	1st + 3rd (2 FACS coders)	ITS interactions

*Dynamic Bayesian Networks (DBNs), Bayesian Networks (BNs), Support Vector Machines (SVMs), Linear Regression (LR), Discriminant Analysis (DA).*

Finally, although affective expressions can never be divorced from context [22], [77], almost none of the systems integrated contextual cues with facial feature tracking.

As evident from our synthesis of the recent Zeng et al. [12] survey, additional technological development is necessary before vision-based affect detection can be functional for real-world applications. Nevertheless, when compared to an earlier review by Pantic and Rothkrantz [10], remarkable progress has been made along a number of dimensions. The most important progress has been made toward larger and more naturalistic emotion databases. In particular, Zeng et al. [12] reviewed six video, seven audio, and five audiovisual databases of affective behavior. The data sets described in the 2009 survey represent a marked improvement compared to the 2003 survey where the training sets of 61 percent of the studies on affect detection from static facial images included 10 or fewer participants. Similarly, 78 percent of studies on affect detection from sequences of facial images had 10 or fewer participants.

### 3.2 Voice (Paralinguistic Features of Speech)

Speech transmits affective information through the explicit linguistic message (*what* is said) and the implicit paralinguistic features of the expression (*how* it is said). The explicit part will be discussed in Section 4.6. The decoding of paralinguistic messages has not yet been fully understood, but listeners seem to be able to decode the basic emotions (particularly from prototypical and acted productions) using prosody [81] and nonlinguistic vocalizations (e.g., laughs and cries); these can also be used to decode other affective signals such as stress, depression, boredom, and excitement.

An extensive analysis of the literature on vocal communication of emotion has yielded some conclusions with important implications for AC applications [20], [82]. First, affective information can be encoded and decoded through speech. The most reliable finding is that pitch appears to be an index into arousal. Second, affect detection accuracy rates from speech are somewhat lower than facial expressions for the basic emotions. Sadness, anger, and fear are the emotions that are best recognized through voice; disgust is the worst. Finally, there is some ambiguity with respect to how different acoustic features communicate the different emotions.

The most recent survey of audio-based affect recognition [12] provides a comprehensive discussion of the literature, so it will not be repeated here. However, a meta-analysis of the 19 systems reviewed by Zeng et al. gives an indication of the current state of the field. When compared to vision-based detection, speech-based detection systems are more apt to meet the needs of real-world applications since 16 out of 19 systems were trained on spontaneous speech. Second, although many systems still focus on detecting the basic emotions, there are some marked efforts aimed at detecting other states, such as frustration. Third, similarly to facial feature tracking, context is often overlooked in acoustic-prosodic-based detection. Nevertheless, the focus on realistic scenarios such as call center logs, tutoring sessions, and some Wizard-of-Oz studies has yielded rich sources of data that will undoubtedly improve next-generation acoustic-prosodic-based affect detection systems.

Table 2 compares sample projects in which voice was used to recognize affect in a number of applications. Far from being a comprehensive list, the studies were selected to show some of the current trends and applications.

Just as with facial affect detection, voice is a promising signal in AC applications: It is low-cost, nonintrusive, and has fast time resolution.

### 3.3 Body Language and Posture

Although Darwin's research on emotion expression heavily focused on body language and posture, state-of-the-art affect detection systems have overlooked posture as a serious contender when compared to facial expressions and acoustic-prosodic features. This is somewhat surprising because there are some benefits to using posture as a means to diagnose the affective states of a user [91], [92], [93]. Human bodies are relatively large and have multiple degrees of freedom, thereby providing them with the capability of assuming a myriad of unique configurations [94]. These static positions can be concurrently combined and temporally aligned with a multitude of movements, all of which makes posture a potentially ideal affective communicative channel [95], [96]. Posture can offer information that is sometimes unavailable from the conventional nonverbal measures such as the face and

TABLE 2  
Voice Analysis for Affect Recognition

Study	Description	Evaluation and Application
[83]	14 categories. Confusion matrix of DA and Jackknifing are similar to those of expert annotators	Self elicitations (12 actors)
[84]	2 categories. Used prosody, lexical and discourse with LDC, KNN	3rd - Call-Center
[85]	2 categories. ADTree and SVM using lexical, acoustic and combination	Medical-help call center
[86]	4 categories. Used prosody, Lexical, PSD, etc., with SVM Angry and neutral sentences from WSJ corpus. SVM, NB, KNN, NN	3rd: 2 annotators - Emergency Call Center Shows that Automatic Speech Recognition that adapts to affect is more accurate.
[87]		
[88]	3 categories: negative, positive neutral. Used prosody, lexical, syntactic, etc with Boost decision tree	3rd: 2 annotators - ITS spoken dialogue
[89]	7 categories. Used prosody, linguistic with MLR, NB, ND, SVM, C4.5	3rd (2 annotators) - Car-user dialogues
[90]	4 Stress categories. 61% on speaker-dependent and 51% on speaker-independent test sets.	Drivers under stress

Evaluations by "1st" and/or "3rd" person reports.

paralinguistic features of speech. For example, the affective state of a person can be decoded over long distances with posture, whereas recognition at the same distance from facial features is difficult or unreliable [97].

Perhaps the greatest advantage to posture-based affect detection is that gross body motions are ordinarily unconscious, unintentional, and thereby not susceptible to social editing, at least compared with facial expressions, speech intonation, and some gestures. Ekman and Friesen [98], in their studies of deception, have coined the term *nonverbal leakage* to refer to the increased difficulty faced by liars, who attempt to disguise deceit, through less controlled channels such as the body when compared to facial expressions.

Mota and Picard [99] reported the first substantial body of work that used the automated posture analysis via Tekscan's Body Pressure Measurement System (BPMS) to infer the affective states of a user in a learning environment. The BPMS consists of a thin-film pressure pad with a rectangular grid of sensing elements that can be mounted on a variety of surfaces, such as the seat and back of a chair. They analyzed temporal transitions of posture patterns to classify the interest level of children, while they performed a learning task on a computer. A neural network provided real-time classification of nine static postures (leaning back, sitting upright, etc.) with an overall accuracy of 87.6 percent. Their system then recognized interest (high interest, low interest, and taking a break) by analyzing posture sequences over a 3 second interval, yielding an overall accuracy of 82.3 percent.

D'Mello and Graesser [100] recently extended Mota and Picard's work by developing systems to detect boredom, engagement/flow, confusion, frustration, and delight from students' gross body movements during a learning task. They extracted two sets of features from the pressure maps that were automatically computed with the BPMS. The first set focused on the average pressure exerted, along with the magnitude and direction of changes in the pressure during emotional experiences. The second set of features monitored the spatial and temporal properties of naturally occurring pockets of pressure. Machine learning experiments yielded affect detection accuracies of 73, 72, 70, 83, and 74 percent (chance = 50%) in detecting boredom, confusion, delight, flow, and frustration from neutral, respectively.

Although there are not yet enough studies to provide a meta-analysis on posture-based affect detection, posture-based affect detection is growing [101], [102] and is indeed a field that is ripe for additional research. Posture tracking is not intrusive to the user's experience, but the equipment required is more expensive and not practical unless the user is sitting. Recent advances in gestural interfaces might open further opportunities.

### 3.4 Physiology

Inspired by the theories that highlight the embodiment of emotion, several AC applications focus on detecting affect by using machine learning techniques to identify patterns in physiological activity that correspond to the expression of different emotions. Much of this AC research is guided by the strong research traditions of physiological psychology and psychophysiology [103]. Physiological psychology studies how physiological variables such as brain stimulation or removal of brain tissue (independent variables) affect other (dependent) variables such as learning or perceptual accuracy. When the independent variable is a stimulus (e.g., an image of a spider) and the dependent a physiological measure (e.g., heart rate), the research is commonly known as psychophysiology.

Although physiological psychology and psychophysiology share the common goal of understanding the physiology of behavior, the latter has stronger implications for affective computing research. Behavior in this body of literature extends beyond emotional processes; it also includes cognitive processes such as perception, attention, deliberation, memory, and problem solving.

Most of the measures used to monitor physiological states are "noninvasive," based on recordings of electrical signals produced by brain, heart, muscles, and skin. The measures include the Electromyogram (EMG) that measures muscle activity, Electrodermal Activity (EDA) that measures electrical conductivity as a function of the activity of sweat glands at the skin surface, Electrocardiogram (EKG or ECG) that measures heart activity, and Electrooculogram (EOG) measuring eye movement. Electroencephalography (EEG, measuring brain activity) and other more recent techniques in neuroimaging are discussed in Section 4.2.

Some important results from psychophysiology reviewed by Andreassi [104] that are of relevance to physiological-based affect detection include:



**Law of initial values** (LIVs) states that the physiological response to a stimulus depends on prestimulus physiological level. For example, for a higher initial level, a smaller response is expected if the stimulus produces an increase and a larger response if it produces a decrease. LIV is questioned by some psychophysiologicalists, pointing out that it does not generalize to all measures (e.g., skin conductance) and can be influenced by other variables.

**Arousal measures** (as discussed elsewhere in this review) refer to quantities that indicate if the subject is calm/sleepy in one extreme or excited in the other. Stemming from [105] classical research, Malmö [106] and others have found performance to be a function of arousal with an inverted-U shape (i.e., poor performance when arousal is too low or high). Performance is optimal at a critical level of arousal and the arousal-performance relationship is influenced by task constraints [107].

**Stimulus-response (SR) specificity** is a theory that states that for particular stimulus situations, subjects will go through specific physiological response patterns. Ekman et al. [108], as well as a number of other researchers, found that autonomic nervous system specificity allowed certain emotions to be discriminated.

**Individual-response (IR) specificity** complements SR specificity but with one important difference. While SR specificity claims that the pattern of responses is similar for most people, IR specificity pertains to how consistent an individual's responses are to different stimulations.

**Cardiac-somatic features** refer to changes in heart responses caused by the body getting ready for a behavioral response (e.g., a fight). The effect might be the cause of physiological changes in situations such as the dissonance effect described by [109]. Here, subjects writing an essay that puts forward an argument that is in agreement with their preexisting attitudes showed lower arousal in the form of electrodermal activity than those writing an essay in disagreement with their preexisting attitudes.

**Habituation and rebound** are two effects appearing in sequences of stimuli. When the stimulus is presented repeatedly, the physiological responsiveness decreases (habituation). When the stimulus is presented, the physiological signals change and, after a while, return to prestimulus levels (rebound) [104].

Despite evidence for SR specificity, accurate recognition requires models that are adjusted to each individual subject, and researchers are looking into efficient ways of accomplishing this goal. For example, Nasoz et al. compared different algorithms for mapping physiological signals to emotions in their MAUI system [110]. More recently, they proposed a Psychophysiological Emotional Map (PPEM) [111] that creates user-dependent mappings between physiological signals (heart rate and skin conductance) and (*valence, arousal*) dimensions.

Table 3 summarizes some of the studies where physiological signals, including EEG, were used to recognize emotions. Some of these studies combined multiple signals, each with advantages and disadvantages for laboratory and real-world applications. Feature selection techniques and the model used (categories or dimensional) for each study are also included in the table, and it appears that most

studies use categorical models. The type of evaluation refers to methods(s) used by the researchers to annotate the data with labels or dimensional data; it can either be a 1st person (e.g., self-report) or 3rd person (an external observer).

The stimulus or elicitation technique used in most of these studies was "self-enactment," where subjects use personal mental images to trigger (or act out) their emotions. Psychophysiologicalists have also performed many studies using databases of photographs, videos, and text. As more applications are built, studies that adopt more realistic scenarios will become more common. A challenge to using these techniques in real applications includes the intrusiveness and the noise in the signals produced by movement. Although standard equipment can record signals at 1,000 Hz, the actual time resolution of each signal depends on the specific features used and of the stimulus-response phenomena described earlier, see Andreassi [104] for details.

### 3.5 Brain Imaging and EEG

Affective neuroscience has utilized recent advances in brain imaging in an attempt to map the neural circuitry that underlies emotional experience [23], [58], [64], [122], [123]. In addition to providing confirmatory evidence for some of the existing emotion theories, recent advances in affect neuroscience have highlighted some new perspectives into the nature of emotional experience and expression. For example, affective neuroscience is also contributing evidence to the discussions on dimensional models where valence and arousal might be supported by distinct neural pathways.

The techniques used by neuroscientists, particularly Functional Magnetic Resonance Imaging (fMRI) [44], are providing new evidence into emotional phenomena. Immordino-Yang and Damasio [124] used evidence from brain-damaged subjects to show that emotions are essential in decision making. Patients with a lesion in a particular section of the frontal lobe showed normal logical reasoning yet they were blind to the consequences of their actions and unable to learn from mistakes. Thus, the authors showed how emotion-related processes were required for learning, even in areas that had previously been attributed to cognition. To AC researchers interested in learning technologies, this work provides some evidence of the effect of emotions on learning.

Despite its limitations, EEG studies, often supported by fMRI approaches, have been successful in many other ways. Due to the lack of a neural model of emotion, few studies in AC have used EEG or other neuroimaging techniques. Although most of the focus has been on Event-Related Potentials (ERPs), which have recently been reviewed by Olofsson et al. [125], machine learning techniques have shown promise for building automatic recognition systems from EEG signals [120].

Regrettably the cost, time resolution, and complexity of setting up experimental protocols that resemble real-world activities are still problematic issues that hinder the development of practical applications that utilize these techniques. Signal processing [126] and classification algorithms [127] for EEG have been developed in the context of building Brain Computer Interfaces (BCIs), and researchers are constantly seeking ways for developing new approaches to recognizing affective states from EEG and other physiological signals.

TABLE 3  
Physiological Signals for Affect Recognition, Including EEG

Study	Signal	Description	Stimulus and evaluation
[112]	EKG, SC, EMG	8 categories (baseline 12.5%). Best acc: 81.25% using 40 features on the data set II. Fisher Proj with SFFS / KNN.	Self elicitation
[113]	EKG, SC, EMG	4 categories. Acc: 80-90% using 120 features, then PCA.	Self-selected songs
[114]	EKG, ST, SC	3&4 categories with acc. of 78% for 3 and 62% for 4 using SVM. Features included RR, HRV.	Audio-visual, self-evaluation
[110]	SC, HR, ST	6 categories + intensity of each. / KNN (71%), DFA (74%), MBP (83%) for individual emotions. No results reported for overall.	Movie clips selected by panel
[111]	HR+SC	Dimensional model, several signal processing techniques discussed, no classification results.	No subjects evaluated
[115]	EKG, SC, EMG	8 categories, 68-80% for individual subjects, 42% for all-subject SVM classifier. 120 features / NB, FT, BN, MLP, LLR, SVM	Self elicited
[116]	EKG, SC, + Face	2 categories (amusement versus sadness) + intensity. Chi-square/SVM and Logistic Reg.	3rd & Films
[117]	EKG, SC, EMG, ST	3 categories. 83% using SVM. Aimed at children with Autism	1st and 3rd person (parent and therapist) evaluations. & Computer based tasks, Pong game and anagram
[118]	EMG, BVP, GSR, RESP	8 categories, 40-46% using SFFS/KNN, FP/MAP, Hybrid SFFS/ FP	Self elicited
[119]	EKG, EMG, SC, ST, BVP, RESP	Dimensional model. Acc for 10-20% bands: Valence (90-97%), Arousal (63-90%). Used MLP	IAPS
[120]	EEG	10 categories. 33-55% for individual subjects using SVM. Evaluated static and adaptive NB, KNN, SVN	Self elicited
[121]	EEG	3 dimensions: valence, arousal, dominance. 17 subjects. Evaluated Nearest Neighbours, DT, Bagging	IAPS

Signals include EKG, Skin Conductivity, Electrodermal activity (SC), Skin Temperature (ST). Feature selection methods include Statistical (PCA, Chi-square, PSD) and Physiological (interbeat interval or IBI, Heart Rate Variability (HRV), and others). Classifiers include Naive Bayes (NB), Function Trees (FT), BNs, Multilayer Perceptron (MLP), Linear Logistic Regression (LLR), SVMs, Discriminant Function Analysis (DFA), and Marquardt Backpropagation (MBP).

### 3.6 Text

The research on detecting emotional content in text refers to written language and transcriptions of oral communication. Both cases represent humans' need to communicate with others, so it is no surprise that the first researchers to try linking text to emotions were social psychologists and anthropologists trying to find similarities on how people from different cultures communicate [128], [129]. This research was also triggered by a dissatisfaction with the dominant cognitive view centered around humans as "information processors" [129].

Early work on understanding how people express emotions through text, or how text triggers different emotions, was conducted by Osgood et al. [128], [130]. Osgood used multidimensional scaling (MDS) to create visualizations of affective words based on similarity ratings of the words provided to subjects from different cultures. The words can be thought of as points in a multidimensional space, and the similarity ratings represent the distances between these words. MDS projects these distances to points in a smaller dimensional space (usually two or three dimensions).

The emergent dimensions found by Osgood were "evaluation," "potency," and "activity." *Evaluation* quantifies how a word refers to an event that is pleasant or unpleasant, similar to hedonic valence. *Potency* quantifies how a word is associated to an intensity level, particularly

strong versus weak. *Activity* refers to whether a word is active or passive. These dimensions are qualitatively similar to valence and arousal, which are considered to be the fundamental dimensions of affective experience [19], [21].

Lutz and others have found similar dimensions but differences in the similarity matrices produced by people of different cultures. This anthropological work has been controversial [129] and progress may be limited by the cost of such field studies. However, more recently, Samsonovich and Ascoli [131] used English and French dictionaries to generate what they called "conceptual value maps," a type of "cognitive map" similar to Osgood's, and found the same set of underlying dimensions.

Another strand of research involves a lexical analysis of the text in order to identify words that are predictive of the affective states of writers or speakers [132], [133], [134], [135], [136], [137]. Several of these approaches rely on the Linguistic Inquiry and Word Count (LIWC) [138], a validated computer tool that analyzes bodies of text using dictionary-based categorization. LIWC-based affect detection methods attempt to identify particular words that are expected to reveal the affective content in the text [132], [134], [137]. For example, first person singular pronouns in essays (e.g., "I" and "me") have been linked to negative emotions [139], [140].

Corpora-based approaches shown in Table 4 have been used by researchers who assume that people using the same

TABLE 4  
Text Analysis for Affect Recognition

Study	Feat. Sel. / Classifier	Model	Evaluation	Corpora	Other
[155]	Winnow Linear	7 categories	1st	Children stories	
[156]	/ VSM, SVM, NB	5 categories	1st	ISEAR questionnaire	ontology: ConceptNet
[157]	WN presence, LSA	6 categories	3rd (6 annotators)	News stories	Ontology: Wordnet Affect
[158]	LSA / MLP, NB, KNN, C4.5, ALG	3 categories	1st	Dialogue turns	ITS interaction
[159]	Cohesion Indices	4 categories	1st and 3rd	dialogue logs	ITS interaction
[160]	WN presence, sentence	6 categories	N/A	chat logs	Ontology: Wordnet Affect
[161]	WN presence, LSA	WN a-labels	N/A	N/A	Ontology: Wordnet Affect
[162]	/ NB, SVM	2 categories	3rd	Political articles	

*International Survey on Emotion Antecedents and Reactions (ISEAR) and Additive Logistic Regression (ALR).*

language would have similar conceptions for different discrete emotions; a number of these researchers have built thesauri of emotional terms. For example, Wordnet is a lexical database of English terms and is widely used in computational linguistics research [141]. Strapparava and Valituti [142] extended Wordnet with information on affective terms.

The Affective Norm for English Words (ANEWs) [143], [144] is one of several projects to develop sets of normative emotional ratings for collections of emotion elicitation objects, in this case English words. This initiative complements others such as the International Affective Picture System (IAPS), a collection of photographs. These collections provide values for valence, arousal, and dominance for each item, averaged over a large number of subjects who rated the items using the Self-Assessment Manikin (SAM) introduced by Lang and colleagues. Finally, affective norms for English Text (ANET) [145] provides a set of normative emotional ratings for a large number of brief texts in the English language.

There are also some text-based affect detection systems that go a step beyond simple word matching by performing a semantic analysis of the text. For example, Gill et al. [146] analyzed 200 blogs and reported that texts judged by humans as expressing fear and joy were semantically similar to emotional concept words (e.g., phobia, terror for fear and delight, and bliss for joy). They used Latent Semantic Analysis (LSA) [147] and the Hyperspace Analogue to Language (HAL) [148] to automatically compute the semantic similarity between the texts and emotion keywords (e.g., fear, joy, and anger). Although this method of semantically aligning text to emotional concept words showed some promise for fear and joy texts, it failed for texts conveying six other emotions, such as anger, disgust, and sadness. So, it is an open question whether semantic alignment of texts to emotional concept terms is a useful method for emotion detection.

Perhaps the most complex approach to textual affect sensing involves systems that construct affective models

from large corpora of world knowledge and applying these models to identify the affective tone in texts [14], [149], [150], [151], [152]. For example, the word "accident" is typically associated with an undesirable event. Hence, the presence of "accident" will increase the assigned negative valence of the sentence "I was late to work because of an accident on the freeway." This approach is sometimes called sentiment analysis, opinion extraction, or subjectivity analysis because it focuses on the valence of a textual sample (i.e., positive or negative; bad or good) rather than assigning the text to a particular emotion category (e.g., angry and sad). Sentiment and opinion analysis is gaining traction in the computational linguistics community and is extensively discussed in a recent review [14].

### 3.7 Multimodality

Views that advocate emotions as expressions and embodiments propose that multiple physiological and behavioral response systems are activated during an emotional episode. For example, anger is expected to be manifested via particular facial, vocal, and bodily expressions, changes in physiology such as increased heart rate, and is accompanied by instrumental action. Responses from multiple systems that are bound in space and time during an emotional episode are essentially the hallmark of basic emotion theory [67]. Hence, it is somewhat surprising that affect detection systems that integrate information from different modalities have been widely advocated but rarely implemented [153]. This is mainly due to the inherent challenges with unisensory affect detection, which increase in multisensory environments. Nevertheless, the advantage of multimodal human-computer interaction systems has been recognized and seen as the next step for the mostly unimodal approaches currently used [154].

There are three methods to fuse signals from different sensors, each depending on when information from the different sensors is combined [154].

**Data Fusion** is performed on the raw data for each signal and can only be applied when the signals have the same

TABLE 5  
Selected Work on Multimodal Approaches for Affect Recognition

Study	Signal	Description	Evaluation & Stimulus
[178]	Face, voice and body	10 emotions in 5 categories. Used Factor Analysis to measure emotion differentiation and modality-specific skills in humans. Results suggest that visual and auditory recognition are two separate abilities.	Self elicitation, actors
[179]	Face, voice	3 categories. Used PCA and SVM for feature and decision level fusion. Recognition with facial features (85%) is better than with speech (71%), but best is fused classifier (84-89%).	Self elicitation
[175]	Face, posture, activity logs	3 categories. Decision level fusion where SVM is used for each modality. Combined classifier produced 86% (higher than with individual modalities).	Aimed at identifying interest/disinterest in children using an educational puzzle. Evaluations by teacher.
[180]	Speech, Text	7 Categories. 33 acoustic features reduced using PCA. Text features produced from emotional lexicon. SVM is used as classifier for each modality. Decision level fusion used.	Broadcast drama. Chinese, ad-hoc thesaurus
[181]	Face, voice	5 categories + 'other', ANOVA. Studies perceptual ability of subjects with individual and combined modality. No automatic classification.	Italian movie clips
[10]	Face, voice	Survey and meta-study. Analyses previously published bi-modal techniques.	

temporal resolution. It can be used for integrating physiological signals coming from the same recording equipment as is commonly the case with physiological signals, but it could not be used to integrate a video signal with a text transcript. It is also not commonly used because of its sensitivity to noise produced by the malfunction or misalignment of the different sensors.

**Feature Fusion** is performed on the set of features extracted from each signal. This approach is more commonly used in multimodal HCI and has been used in affective computing, for example, in the Augsburg Biosignal Toolbox [113]. Features for each signal (EKG, EMG, etc.) are primarily the mean, median, standard deviation, maxima, and minima, together with some unique features from each sensor. These are individually computed for each sensor and then combined across sensors. Table 5 lists some of the features extracted in different studies.

**Decision Fusion** is performed by merging the output of the classifier for each signal. Hence, the affective states would first be classified from each sensor and would then be integrated to obtain a global view across the various sensors. It is the most commonly used approach for multimodal HCI [10], [154].

As indicated above, there have been very few systems that have explored multimodal affect detection. These primarily include amalgamations of physiological sensors and combinations of audio-visual features [112], [163], [164], [165], [166], [167]. These have been reviewed in [12] with some detail. Another approach involves a combination of acoustic-prosodic, lexical, and discourse features for affect detection [84], [168], [169]; however, the research incorporating this approach is too sparse to warrant a meaningful meta-analysis.

Of greater interest is the handful of research efforts that have attempted to monitor three or more modalities. For example, Scherer and Ellgring [170] considered the possibility of combining facial, vocal features, and body movements (posture and gesture) to discriminate among 14 emotions

(e.g., hot anger, shame, etc.; Base rate =  $1/14 = 7.14\%$ ). Single-channel classification accuracies from 21 facial features and 16 acoustic parameters were 52.2 and 52.5 percent, respectively (accuracy rates for gesture and body movements were not provided). A combined 37-channel model yielded a classification accuracy of 79 percent, but the accuracy dropped to 54 percent when only 10 of the most diagnostic features were included in the model. It is difficult to assess whether the combined models led to enhanced or equivalent classification scores compared to the single-channel models because the number of features between single and multi-channel models was not equivalent.

More recently, Castellano et al. considered the possibility of detecting eight emotions (some basic emotions plus irritation, despair, etc.) by monitoring facial features, speech contours, and gestures [171]. Classification accuracy rates for the single-channel classifiers were 48.3, 67.1, and 57.1 percent, for the face, gestures, and speech, respectively. The multimodal classifier achieved an accuracy of 78.3 percent, representing a 17 percent improvement over the best single-channel system.

Although Scherer's and Castellano's systems demonstrate that there are some advantages to multichannel affect detection, the systems were trained and validated on context-free, acted, and emotional expressions. However, practical applications require the detection of naturalistic emotional expressions that are grounded in the context of the interaction. There are also some important differences between real and posed affective expressions [79], [172], [173], [174]; hence, insights gleaned from studies on acted expressions might not generalize to real contexts.

On the more naturalistic front, Kapoor and Picard developed a contextually grounded probabilistic system to infer a child's interest level on the basis of upper and lower facial feature tracking, posture patterns (current posture and level of activity), and some contextual information (difficulty level and state of the game) [175]. The combination of these modalities yielded a recognition accuracy of 86 percent,

which was quantitatively greater than that achieved from the facial features (67 percent upper face and 53 percent lower face) and contextual information (57 percent). However, the posture features alone yielded an accuracy of 82 percent which would indicate that the other channels are redundant with posture.

Kapoor et al. have extended their system to include a skin conductance sensor and a pressure-sensitive mouse in addition to the face, posture, and context [176]. This system predicts self-reported frustration while children engaged in a Towers of Hanoi problem solving task. The system yielded an accuracy score of 79 percent, which is a substantial improvement over the 58.3 percent accuracy base line. An analysis of the discriminability of the 14 affect predictors indicated that mouth fidgets (i.e., general mouth movements), velocity of the head, and ratio of postures were the most diagnostic features. Unfortunately, the Kapoor et al. (2007) study did not report single-channel classification accuracy; hence, it is difficult to assess the specific benefits of considering multiple channels.

In a recent study, Arroyo et al. [35] considered a combination of context, facial features, seat pressure, galvanic skin conductance, and pressure exerted on a mouse to detect levels of confidence, frustration, excitement, and interest of students in naturalistic school settings. Their results indicated that two parameter (face + context) models explained 52, 29, and 69 percent of the variance for confidence, interest, and excited, respectively. A combination of seat pressure and context yielded the most accurate model (46 percent of variance) for predicting frustration. Their results support the conclusion that in most cases, monitoring facial plus contextual features yielded the best fitting models, and the other channels did not provide any additional advantages.

D'Mello and Graesser [177] considered a combination of facial features, gross body language, and conversational cues for detecting some of the learning-centered affective states. Classification results supported a channel  $\times$  judgment type interaction, where the face was the most diagnostic channel for spontaneous affect judgments (i.e., at any time in the tutorial session), while conversational cues were superior for fixed judgments (i.e., every 20 seconds in the session). The analyses also indicated that the accuracy of the multichannel model (face, dialogue, and posture) was statistically higher than the best single-channel model for the fixed but not spontaneous affect expressions. However, multichannel models reduced the discrepancy (i.e., variance in the precision of the different emotions) of the discriminant models for both judgment types. The results also indicated that the combination of channels yielded enhanced effects for some states but not for others.

## 4 GENERAL DISCUSSION

Despite significant progress, AC is still finding its own voice as a new interdisciplinary field that encompasses research in computer science, engineering, cognitive science, affective science, psychology, learning science, and artifact design. Engineers and computer scientists use machine learning techniques for automatic affect classification from video, voice, text, and physiology. Psychologists

use their long tradition of emotion research with their own discourse, models, and methods, some of which have been incorporated into affective computing research. This review has assumed that affective computing, which focuses on developing practical applications that are responsive to user affect, is inextricably bound to the affective sciences that attempt to understand human emotions. Simply put, affective computing cannot be divorced from the century-long psychological research on emotion. Our emphasis on the multidisciplinary landscape that is typical for AC applications sets this review apart from previous survey papers on affect detection.

The remainder of the discussion is organized as follows: First, although we have advocated a tight coupling between affect theory and affect applications, there are situations where research in the affective sciences and affective computing diverge; these are discussed in Section 4.1. There are also some areas where more convergence between affective computing researchers and emotion theorists is needed; these are considered in Section 4.2. Section 4.3 proposes some open questions and unresolved issues as possible areas of future research for the AC and affective science communities. Finally, Section 4.4 concludes this review.

### 4.1 Points of Divergence between the Affective Sciences and Affective Computing

As noted above, although emotion theories (described in Section 2) have informed the work on affect detection (described in Section 3), AC's focus on computational approaches and applications has produced notable differences, including:

**Broadening of the mental states studied.** The six "basic" emotions proposed by Ekman (anger, disgust, fear, joy, sadness, and surprise) and other emotion taxonomies commonly used in the psychological literature might not be useful for developing most of the computer applications that affective computing researchers are interested in. For example, there is considerable research that the basic emotions have minimal relevance to learning sessions that span 30 minutes to 2 hours [182], [183], [184], [185]. Hence, much of the research on basic emotions is of little relevance to the developers of computer learning environments that aspire to detect and respond to students' emotions. Research on some of the learning-centered emotions such as confusion, frustration, boredom, flow, curiosity, and anxiety is more applicable to affect-sensitive learning environments. These states are also expected to be prominent in many HCI contexts and are particularly relevant to AC. In general, the more comprehensive field of affective science that studies phenomena such as feelings, moods, attitudes, affective styles, and temperament is very relevant to affective computing.

**Flexible epistemological grounding.** While setting the context for AC research, Picard [7] left aside the debate on some of the questions considered important to emotion researchers. For example, there is the question of whether emotions should be represented as a point in a valence-arousal dimensional space or with a predefined label such as anger, fear, etc. Instead of perseverating on such questions, Picard advocated a focus on the issues that are

necessary for the development of affect-sensitive computer applications. This “pragmatist” approach has been common to much affective computing research.

In the short term, the approach paid off by allowing prototypes to be built rapidly [186], [187] and increased research devoted to AC. In particular, there are growing technical communities in the areas of Artificial Intelligence in Education (AIED), Intelligent Tutoring Systems (ITSs) [183], [188], [189], [190], [191], gaming and entertainment, general affect-sensitive HCI systems, and emotion prosthetics (i.e., the development of affective computing applications to support individuals suffering broad autism spectrum disorder [117], [192]).

#### 4.2 Points for Increased Convergence between the Affective Sciences and Affective Computing

Most affect detection research has been influenced by emotion theories; however, perspectives that view emotions as expressions, embodiments, and products of cognitive appraisal have been dominant. Other schools of psychology, particularly social perspectives on emotion, have been on the sidelines of AC research. This is an unfortunate consequence, as highlighted by Parkinson [193] in his criticism of three areas of emotion research. These included the individual, interpersonal, and the representational components of emotion, as individually discussed below.

Parkinson contends that psychologists (and we add AC researchers) have made three problematic assumptions while studying *individual emotions*. First, most AC researchers assume that emotions are *instantaneous* (i.e., they are on or off at any particular point in time). This assumption is reflected in emotion annotation tasks where raters label a user’s response to a stimulus, either instantaneous (e.g., a photo) or over time (e.g., a video), as a category or a point in a dimensional space. Parkinson and others argue that the boundaries are not as clear and the phenomena needs to be better contextualized.

The second limitation pertains to the assumption that emotions are *passive*. This assumption is frequently made by AC researchers and psychologists who disregard the effects of emotion regulation when studying emotional processes. The research protocols often assume a simple stimulus-response paradigm and rarely take into account people’s drives to regulate their affective states, an area that is gaining considerable research attention [194], [195]. For example, while studying the affective response during interactions with learning environments, boredom (a negative low arousal state) might be a state that the learner tries to alleviate by searching for interesting features in the environment. The effect of an affect-inducing intervention would be confounded with the effect of the emotion regulation in this situation.

The third and most important criticism associated with individual emotions is what Parkinson refers to as *impermeable emotions*. This criticism is prominent when emotions are investigated outside of social contexts, as is the case with most AC applications. This is particularly problematic because some emotions are directed at other people and arise from interactions with them (e.g., social emotions such as pride, jealousy, and shame).

Parkinson argues that researchers have made two experimental simplifications while studying *interpersonal*

*emotions* [193]. Interpersonal emotions refer to affective responses that arise while two or more people interact. One limitation is associated with experimental protocols that forgo synchronous interactions for asynchronous communication (usually via prerecorded messages from a sender to a receiver). Real-world interactions are more consistent with face-to-face synchronous communication paradigms (with the exception of e-mails, text messaging, and other forms of asynchronous digital communication).

Another limitation occurs when the interpersonal emotions arise from communications with strangers, as is the case with many laboratory studies. This is inconsistent with interactions in the real world, where prior interpersonal relationships frequently drive emotional reactions. These limitations associated with interpersonal emotions are specifically relevant to AC systems where the goal of affect sensitivity is to provide a richer interface between two conversing users (e.g., an enhanced chat program).

There are also limitations associated with how emotion theorists and AC researchers *represent emotions*. This is an important limitation because it has been argued that the experience of emotion cannot be separated from its representation [193]. Subjects in any AC study use internal representations that may or may not be consciously accessible [196], [197]. Complications arise when representations used by researchers (e.g., Ekman’s six basic emotions) do not match what users experience. Often, researchers try to “impose” an emotional representation (e.g., valence/arousal dimensions) that might be less or more clear, or understood in different ways by different subjects.

In summary, there is considerable merit to the argument that some affective phenomena cannot be understood at the level of the individual and must always be studied as a social process [198]. Theories that highlight the social function of emotions must be integrated with perspectives that view emotions as expressions, embodiments, and products of cognitive appraisal.

#### 4.3 Problematic Assumptions, Open Questions, and Unresolved Issues

We propose some important items that have not been adequately addressed by the AC community (including our own research). Emotion theorists have considered these items at some length; however, no clear resolution to these problems has emerged.

**Experience versus expression.** The existence of a one-to-one correspondence between the experience and the expression of emotion is perhaps the most widespread and potentially problematic assumption. Affect detection systems inherently infer that a user is angry (the experience) because the system has detected an angry face (the expression). In reality, however, the story is considerably more complex. For example, although facial expressions are considered to be ubiquitous with affective expression and guide entire theories of emotion [67], meta-analyses on correlations between facial expressions and self-reported emotions have yielded small to medium effects [199], [200]. The situation is similarly murky for the link between paralinguistic features of speech and emotions, as illustrated in a recent synthesis of the psychological literature [20]. It may be tempting to conclude that emotion

experience and expression are two sides of the same coin; however, emerging research suggests that this is a tenuous conclusion at best. Of course, physiological and bodily channels do convey some information about a person's emotional states. However, the exact informational value of these channels as emotional indicators is still undetermined.

**Coherence among multiple components of emotion.** Another common assumption made by most AC researchers is that the experience of an emotion is manifested with a sophisticated synchronized response that incorporates peripheral physiology, facial expression, speech, modulations of posture, affective speech, and instrumental action. A corollary of this assumption is that affect detection accuracy will increase when multiple channels are considered over any one channel alone. However, with the exception of intense prototypical emotional experiences or acted expressions [171], [201], converging research points to low correlations between the different components of an emotional episode [20], [21]. Identifying instances where emotional components are coherent versus episodes when they are loosely coupled is an important requirement for affect detection systems.

**Emotions in a generally nonemotional world.** One relatively unspoken, but nonetheless ubiquitous assumption in the AC community is that interactions with computers are generally affect-free, and emotional episodes occasionally intervene in this typically unemotional world. Hence, affective computing is reduced to a mere matter of intercepting and responding to infrequent "bouts of emotion." This view is in stark contrast with contemporary theories that maintain that cognitive processes such as memory encoding and retrieval, causal reasoning, deliberation, and goal appraisal operate throughout the experience of emotion [38], [43], [202], [203], [204]. Some emotional episodes are more intense than others, and these episodes seem to be at the forefront of most AC research. However, affect is always actively influencing cognition and behavior and the challenge is to model these perennially present, but somewhat subtle, manifestations of emotion.

**Context-free affect detection.** Another limitation of affect detection systems is that they are trained on data sets where the affective expressions are collected in a context-free environment. But the nature of affective expressions is not context-free. Instead, it is highly situation-dependent [19], [122], [205], [206] and context is critical because it helps to disambiguate between various exemplars of an emotion category [19]. Since affective expressions can convey different meanings in different contexts, training affect detection systems in context-free environments is unlikely to produce systems that will generalize to new contexts. Hence, a coupling of top-down contextually driven predictive models of affect with bottom-up diagnostic models is essential for affect detection [177], [207], and considerable research is needed along this front.

**Socially divorced detection contexts.** As highlighted in Section 4.2, the problem of impermeable emotions is widespread in AC research. More specifically, since some emotions serve social functions, there is the question of how they are expressed in the absence of social contexts. One complication with applying sociologically inspired emotion theories to AC is that these theories are aimed at interactions among people, but users in AC applications

are more often dealing with objects, rather than people. Hence, it is important to understand the emotional impact that artifacts (such as computer applications) have on their users' emotions [208].

**Categories or dimensions.** Of critical importance to AC researchers is the privileged status awarded to "labeled" states such as anger and happiness. The concept, the value, and even the existence of such "labeled" states is still a matter of debate [19]. Although most AC applications seem to require these categorizations, some have proposed *dimensional* models, and some researchers in HCI have argued that other approaches might be more appropriate for building computer systems. Identifying the appropriate level of representation for practical AC applications is still an unresolved question.

**Evaluating affect detection systems.** Finally, there is the question of how to characterize the performance of an affect detector. In assessing the reliability of a coding scheme, or measurement instrument, kappa scores (agreement after correcting for chance [209]) ranging from 0.4-0.6 are typically considered to be fair, 0.6-0.75 are good, and scores greater than 0.75 are excellent [210]. On the basis of this categorization, the kappa scores obtained by affect detection systems in naturalistic contexts range from poor to fair. It is important to note, however, that these bounds on kappa scores address problems when the decisions are clear-cut and decidable. Emotion detection is an entirely different story because computer algorithms are inferring a complex mental state and emotions are notoriously fuzzy, ill-defined, and possibly indeterminate. Obtaining useful bounds on the performance of affect detection is an important challenge because it is unlikely that perfect accuracy will ever be achieved and there is no objective gold standard.

#### 4.4 Concluding Remarks

This survey integrated some of the key emotion theories with a review of emerging affect detection systems. Perspectives that conceptualize emotions as expressions, embodiments, cognitive appraisals, social constructs, products of neural activity, and psychological interpretation of core affect were contextualized within descriptions of affect detection techniques from several modalities, including facial expressions, voice, body language, physiology, brain imaging, text, and approaches that encompassed multiple modalities. The need to bring ideas from different disciplines was highlighted by important debates pertaining to the universality of basic emotions, cross-cultural innate facial expressions for these emotions, and unresolved issues related to the appraisal process. These and other issues, such as self-regulation in affective processes, the lack of coherence in affective signals, or even the view of emotions as a process or as an emergent property of a social interaction, have yet to be thoroughly considered by affective researchers. These unanswered questions and uncharted territories (such as affective neuroscience) present challenges and opportunities that can sustain AC research for several decades.

#### ACKNOWLEDGMENTS

The authors gratefully acknowledge the editor, Jonathan Gratch, the associated editor, Brian Parkinson, and three

anonymous reviewers for their valuable comments and suggestions. Sidney D'Mello was supported by the US National Science Foundation (NSF) (ITR 0325428, HCC 0834847) and Institute of Education Sciences, US Department of Education (R305A080594). Any opinions, findings, and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF and US Department of Energy (DoE).

## REFERENCES

- [1] S. D'Mello, S. Craig, B. Gholson, S. Franklin, R. Picard, and A. Graesser, "Integrating Affect Sensors in an Intelligent Tutoring System," *Proc. Computer in the Affective Loop Workshop at 2005 Int'l Conf. Intelligent User Interfaces*, pp. 7-13, 2005.
- [2] C. Darwin, *Expression of the Emotions in Man and Animals*. Oxford Univ. Press, Inc., 2002.
- [3] C. Darwin, *The Expression of the Emotions in Man and Animals*. John Murray, 1872.
- [4] W. James, "What Is an Emotion?" *Mind*, vol. 9, pp. 188-205, 1884.
- [5] S. Turkle, *The Second Self: Computers and the Human Spirit*. Simon & Schuster, 1984.
- [6] K.R. Scherer, "Studying the Emotion-Antecedent Appraisal Process: An Expert System Approach," *Cognition and Emotion*, vol. 7, pp. 325-355, 1993.
- [7] R.W. Picard, *Affective Computing*. The MIT Press, 1997.
- [8] K. Boehner, R. DePaula, P. Dourish, and P. Sengers, "Affect: From Information to Interaction," *Proc. Fourth Decennial Conf. Critical Computing: Between Sense and Sensibility*, pp. 59-68, 2005.
- [9] J. Tao and T. Tan, "Affective Computing: A Review," *Proc. First Int'l Conf. Affective Computing and Intelligent Interaction*, J. Tao, T. Tan, and R.W. Picard, eds., pp. 981-995, 2005.
- [10] M. Pantic and L. Rothkrantz, "Toward an Affect-Sensitive Multimodal Human-Computer Interaction," *Proc. IEEE*, vol. 91, no. 9, pp. 1370-1390, Sept. 2003.
- [11] N. Sebe, I. Cohen, and T.S. Huang, "Multimodal Emotion Recognition," *Handbook of Pattern Recognition and Computer Vision*, World Scientific, 2005.
- [12] Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39-58, Jan. 2009.
- [13] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor, "Emotion Recognition in Human-Computer Interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32-80, 2001.
- [14] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, pp. 1-135, 2008.
- [15] J. Gratch and S. Marsella, *Social Emotions in Nature and Artifact: Emotions in Human and Human-Computer Interaction*. Oxford Univ. Press, to be published.
- [16] K. Scherer, T. Bänziger, and R. EB, *Blueprint for Affective Computing: A Source Book*. Oxford Univ. Press, to be published.
- [17] D. Gkay and G. Yildirim, *Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives*. Information Science Publishing, 2010.
- [18] *New Perspectives on Affect and Learning Technologies*, R.A. Calvo and S. D'Mello, eds., Springer, to be published.
- [19] J.A. Russell, "Core Affect and the Psychological Construction of Emotion," *Psychological Rev.*, vol. 110, pp. 145-172, 2003.
- [20] J.A. Russell, J.A. Bachorowski, and J.M. Fernandez-Dols, "Facial and Vocal Expressions of Emotion," *Ann. Rev. of Psychology*, vol. 54, pp. 329-349, 2003.
- [21] L. Barrett, "Are Emotions Natural Kinds?" *Perspectives on Psychological Science*, vol. 1, pp. 28-58, 2006.
- [22] L. Barrett, B. Mesquita, K. Ochsner, and J. Gross, "The Experience of Emotion," *Ann. Rev. of Psychology*, vol. 58, pp. 373-403, 2007.
- [23] T. Dalgleish, B. Dunn, and D. Mobbs, "Affective Neuroscience: Past, Present, and Future," *Emotion Rev.*, vol. 1, pp. 355-368, 2009.
- [24] T. Dalgleish and M. Power, *Handbook of Cognition and Emotion*. John Wiley & Sons, Ltd., 1999.
- [25] M. Lewis, J. Haviland-Jones, and L. Barrett, *Handbook of Emotions*, third ed. Guilford Press, 2008.
- [26] R.J. Davidson, K.R. Scherer, and H.H. Goldsmith, *Handbook of Affective Sciences*. Oxford Univ. Press, 2003.
- [27] A. Ortony and T. Turner, "What's Basic about Basic Emotions," *Psychological Rev.*, vol. 97, pp. 315-331, July 1990.
- [28] J. Russell, "Is There Universal Recognition of Emotion from Facial Expression?: A Review of the Cross-Cultural Studies," *Psychological Bull.*, vol. 115, pp. 102-141, 1994.
- [29] P. Ekman, *Universals and Cultural Differences in Facial Expressions of Emotion*. Univ. of Nebraska Press, 1971.
- [30] P. Ekman and W.V. Friesen, *Unmasking the Face*. Malor Books, 2003.
- [31] S.S. Tomkins, *Affect Imagery Consciousness: Volume I, The Positive Affects*. Tavistock, 1962.
- [32] C. Izard, *The Face of Emotion*. Appleton-Century-Crofts, 1971.
- [33] C. Izard, "Innate and Universal Facial Expressions: Evidence from Developmental and Cross-Cultural Research," *Psychological Bull.*, vol. 115, pp. 288-299, 1994.
- [34] N. Frijda, "Emotion, Cognitive Structure, and Action Tendency," *Cognition and Emotion*, vol. 1, pp. 115-143, 1987.
- [35] I. Arroyo, D.G. Cooper, W. Burleson, B.P. Woolf, K. Muldner, and R. Christopherson, "Emotion Sensors Go to School," *Proc. 14th Conf. Artificial Intelligence in Education*, pp. 17-24, 2009.
- [36] M.B. Arnold, *Emotion and Personality Vol. 1 & 2*. Columbia Univ. Press, 1960.
- [37] C. Smith and P. Ellsworth, "Patterns of Cognitive Appraisal in Emotion," *J. Personality and Social Psychology*, vol. 48, pp. 813-838, 1985.
- [38] K. Scherer, A. Schorr, and T. Johnstone, *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford Univ. Press, 2001.
- [39] I.J. Roseman, M.S. Spindel, and J.P.E., "Appraisals of Emotion Eliciting Events: Testing a Theory of Discrete Emotions," *J. Personality and Social Psychology*, vol. 59, pp. 899-915, 1990.
- [40] S. Schachter and J.E. Singer, "Cognitive, Social, and Physiological Determinants of Emotional State," *Psychological Rev.*, vol. 69, pp. 379-399, 1962.
- [41] R. Lazarus, *Emotion and Adaptation*. Oxford Univ. Press, 1991.
- [42] I.J. Roseman, "Cognitive Determinants of Emotion," *Emotions, Relationships and Health*, pp. 11-36, Beverly Hills Publishing, 1984.
- [43] A. Ortony, G. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge Univ. Press, 1988.
- [44] J. Ledoux, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Simon & Schuster, 1998.
- [45] R. Zajonc, "On the Primacy of Affect," *Am. Psychologist*, vol. 39, pp. 117-123, 1984.
- [46] R. Zajonc, "Feeling and Thinking: Preferences Need No Inferences," *Am. Psychologist*, vol. 33, pp. 151-175, 1980.
- [47] C. Conati, "Probabilistic Assessment of User's Emotions during the Interaction with Educational Games," *J. Applied Artificial Intelligence*, special issue on merging cognition and affect in HCI, vol. 16, pp. 555-575, 2002.
- [48] J.R. Averill, "A Constructivist View of Emotion," *Emotion: Theory, Research and Experience*, pp. 305-339, Academic Press, 1980.
- [49] A. Glenberg, D. Havas, R. Becker, and M. Rinck, "Grounding Language in Bodily States: The Case for Emotion," *The Grounding of Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, R.A.Z.D. Pecher, ed., pp. 115-128, Cambridge Univ. Press, 2005.
- [50] J. Stets and J. Turner, "The Sociology of Emotions," *Handbook of Emotions*, M. Lewis, J. Haviland-Jones, and L. Barrett, eds., third ed., pp. 32-46, The Guilford Press, 2008.
- [51] P. Salovey, "Introduction: Emotion and Social Processes," *Handbook of Affective Sciences*, R.J. Davidson, K.R. Scherer, and H.H. Goldsmith, eds., Oxford Univ. Press, 2003.
- [52] J.Y. Chiao, T. Iidaka, H.L. Gordon, J. Nogawa, M. Bar, E. Aminoff, N. Sadato, and N. Ambady, "Cultural Specificity in Amygdala Response to Fear Faces," *J. Cognitive Neuroscience*, vol. 20, pp. 2167-2174, Dec. 2008.
- [53] G. Peterson, "Cultural Theory and Emotions," *Handbook of the Sociology of Emotions*, J. Stets and J. Turner, eds., pp. 114-134, Springer, 2006.
- [54] D. Matsumoto, "Cultural Similarities and Differences in Display Rules," *Motivation and Emotion*, vol. 14, pp. 195-214, 1990.
- [55] P. Ekman and W. Friesen, *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions*. Prentice-Hall, 1975.
- [56] T.D. Kemper, "Predicting Emotions from Social-Relations," *Social Psychology Quarterly*, vol. 54, pp. 330-342, Dec. 1991.



- [57] J. Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford Univ. Press, 2004.
- [58] A. Damasio, *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Harcourt, Inc., 2003.
- [59] A. Ohman and J.J.F. Soares, "Emotional Conditioning to Masked Stimuli: Expectancies for Aversive Outcomes Following Nonrecognized Fear-Relevant Stimuli," *J. Experimental Psychology: General*, vol. 127, pp. 69-82, 1998.
- [60] M.G. Calvo and L. Nummenmaa, "Processing of Unattended Emotional Visual Scenes," *J. Experimental Psychology: General*, vol. 136, pp. 347-369, 2007.
- [61] M.E. Dawson, A.J. Rissling, A.M. Schell, and R. Wilcox, "Under What Conditions Can Human Affective Conditioning Occur without Contingency Awareness?: Test of the Evaluative Conditioning Paradigm," *Emotion*, vol. 7, pp. 755-766, 2007.
- [62] R.J. Davidson, P. Ekman, C.D. Saron, J.A. Senulis, and W.V. Friesen, "Approach-Withdrawal and Cerebral Asymmetry: Emotional Expression and Brain Physiology I," *J. Personality and Social Psychology*, vol. 58, pp. 330-341, 1990.
- [63] P.C. Petrantonis and L.J. Hadjileontiadis, "Emotion Recognition from EEG Using Higher Order Crossings," *IEEE Trans. Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186-194, Mar. 2010.
- [64] M.D. Lewis, "Bridging Emotion Theory and Neurobiology through Dynamic Systems Modeling," *Behavioral and Brain Sciences*, vol. 28, pp. 169-194, 2005.
- [65] J. Coan, "Emergent Ghosts of the Emotion Machine," *Emotion Rev.*, to appear.
- [66] R.A. Calvo, "Latent and Emergent Models in Affective Computing," *Emotion Rev.*, to appear.
- [67] P. Ekman, "An Argument for Basic Emotions," *Cognition and Emotion*, vol. 6, pp. 169-200, 1992.
- [68] P. Ekman, "Expression and the Nature of Emotion," *Approaches to Emotion*, K. Scherer and P. Ekman, eds., pp. 319-344, Erlbaum, 1984.
- [69] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement: Investigator's Guide 2 Parts*. Consulting Psychologists Press, 1978.
- [70] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Classifying Facial Actions," *IEEE Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, Oct. 1999.
- [71] A. Asthana, J. Saragih, M. Wagner, and R. Goecke, "Evaluating AAM Fitting Methods for Facial Expression Recognition," *Proc. 2009 Int'l Conf. Affective Computing and Intelligent Interaction*, 2009.
- [72] T. Brick, M. Hunter, and J. Cohn, "Get the FACS Fast: Automated FACS Face Analysis Benefits from the Addition of Velocity," *Proc. 2009 Int'l Conf. Affective Computing and Intelligent Interaction*, 2009.
- [73] M.E. Hoque, R. el Kaliouby, and R.W. Picard, "When Human Coders (and Machines) Disagree on the Meaning of Facial Affect in Spontaneous Videos," *Proc. Ninth Int'l Conf. Intelligent Virtual Agents*, 2009.
- [74] R. El Kaliouby and P. Robinson, "Generalization of a Vision-Based Computational Model of Mind-Reading," *Proc. First Int'l Conf. Affective Computing and Intelligent Interaction*, pp. 582-589, 2005.
- [75] R. El Kaliouby and P. Robinson, "Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, vol. 3, p. 154, 2004.
- [76] B. McDaniel, S. D'Mello, B. King, P. Chipman, K. Tapp, and A. Graesser, "Facial Features for Affective State Detection in Learning Environments," *Proc. 29th Ann. Meeting of the Cognitive Science Soc.*, 2007.
- [77] H. Aviezer, R. Hassin, J. Ryan, C. Grady, J. Susskind, A. Anderson, M. Moscovitch, and S. Bentin, "Angry, Disgusted, or Afraid? Studies on the Malleability of Emotion Perception," *Psychological Science*, vol. 19, pp. 724-732, 2008.
- [78] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Fully Automatic Facial Action Recognition in Spontaneous Behaviour," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 223-230, 2006.
- [79] M. Pantic and I. Patras, "Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments from Face Profile Image Sequences," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 2, pp. 433-449, Apr. 2006.
- [80] H. Gunes and M. Piccardi, "Bi-Modal Emotion Recognition from Expressive Face and Body Gestures," *J. Network and Computer Applications*, vol. 30, pp. 1334-1345, 2007.
- [81] P.N. Juslin and K.R. Scherer, "Vocal Expression of Affect," *The New Handbook of Methods in Nonverbal Behavior Research*, Oxford Univ. Press, 2005.
- [82] T. Johnstone and K. Scherer, "Vocal Communication of Emotion," *Handbook of Emotions*, pp. 220-235, Guilford Press, 2000.
- [83] R. Banse and K.R. Scherer, "Acoustic profiles in Vocal Emotion Expression," *J. Personality and Social Psychology*, vol. 70, pp. 614-636, 1996.
- [84] C.M. Lee and S.S. Narayanan, "Toward Detecting Emotions in Spoken Dialogs," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 2, pp. 293-303, Mar. 2005.
- [85] L. Devillers, L. Vidrascu, and L. Lamel, "Challenges in Real-Life Emotion Annotation and Machine Learning Based Detection," *Neural Networks*, vol. 18, pp. 407-422, 2005.
- [86] L. Devillers and L. Vidrascu, "Real-Life Emotions Detection with Lexical and Paralinguistic Cues on Human-Human Call Center Dialogs," *Proc. Ninth Int'l Conf. Spoken Language Processing*, 2006.
- [87] B. Schuller, J. Stadermann, and G. Rigoll, "Affect-Robust Speech Recognition by Dynamic Emotional Adaptation," *Proc. Speech Prosody*, 2006.
- [88] D. Litman and K. Forbes-Riley, "Predicting Student Emotions in Computer-Human Tutoring Dialogues," *Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics*, 2004.
- [89] B. Schuller, R.J. Villar, G. Rigoll, and M. Lang, "Meta-Classifiers in Acoustic and Linguistic Feature Fusion-Based Affect Recognition," *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing*, 2005.
- [90] R. Fernandez and R.W. Picard, "Modeling Drivers' Speech under Stress," *Speech Comm.*, vol. 40, pp. 145-159, 2003.
- [91] P. Bull, *Posture and Gesture*. Pergamon Press, 1987.
- [92] M. Demeijer, "The Contribution of General Features of Body Movement to the Attribution of Emotions," *J. Nonverbal Behavior*, vol. 13, pp. 247-268, 1989.
- [93] A. Mehrabian, *Nonverbal Communication*. Aldine-Atherton, 1972.
- [94] N. Bernstein, *The Co-Ordination and Regulation of Movement*. Pergamon Press, 1967.
- [95] M. Coulson, "Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence," *J. Nonverbal Behavior*, vol. 28, pp. 117-139, 2004.
- [96] J. Montepare, E. Koff, D. Zaitchik, and M. Albert, "The Use of Body Movements and Gestures as Cues to Emotions in Younger and Older Adults," *J. Nonverbal Behavior*, vol. 23, pp. 133-152, 1999.
- [97] R. Walk and K. Walters, "Perception of the Smile and Other Emotions of the Body and Face at Different Distances," *Bull. Psychonomic Soc.*, vol. 26, p. 510, 1988.
- [98] P. Ekman and W. Friesen, "Nonverbal Leakage and Clues to Deception," *Psychiatry*, vol. 32, pp. 88-106, 1969.
- [99] S. Mota and R. Picard, "Automated Posture Analysis for Detecting Learner's Interest Level," *Proc. Computer Vision and Pattern Recognition Workshop*, vol. 5, p. 49, 2003.
- [100] S. D'Mello and A. Graesser, "Automatic Detection of Learner's Affect from Gross Body Language," *Applied Artificial Intelligence*, vol. 23, pp. 123-150, 2009.
- [101] N. Bianchi-Berthouze and C. Lisetti, "Modeling Multimodal Expression of User's Affective Subjective Experience," *User Modeling and User-Adapted Interaction*, vol. 12, pp. 49-84, Feb. 2002.
- [102] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer, "Automated Analysis of Body Movement in Emotionally Expressive Piano Performances," *Music Perception*, vol. 26, pp. 103-119, Dec. 2008.
- [103] J.P. Pines, *Biopsychology*. Allyn and Bacon, 2004.
- [104] J.L. Andreassi, *Human Behaviour and Physiological Response*. Taylor & Francis, 2007.
- [105] R. Yerkes and J. Dodson, "The Relation of Strength of Stimulus to Rapidity of Habit-Formation," *J. Comparative Neurology and Psychology*, vol. 18, pp. 459-482, 1908.
- [106] R. Malm, "Activation," *Experimental Foundations of Clinical Psychology*, pp. 386-422, Basic Books, 1962.
- [107] E. Duffy, "Activation," *Handbook of Psychophysiology*, pp. 577-622, Holt, Rinehart and Winston, 1972.
- [108] P. Ekman, R. Levenson, and W. Friesen, "Autonomic Nervous System Activity Distinguishes among Emotions," *Science*, vol. 221, pp. 1208-1210, 1983.
- [109] R. Croyle and J. Cooper, "Dissonance Arousal: Physiological Evidence," *J. Personality and Social Psychology*, vol. 45, pp. 782-791, 1983.

- [110] F. Nasoz, K. Alvarez, C.L. Lisetti, and N. Finkelstein, "Emotion Recognition from Physiological Signals Using Wireless Sensors for Presence Technologies," *Cognition, Technology and Work*, vol. 6, pp. 4-14, 2004.
- [111] O. Villon and C. Lisetti, "A User-Modeling Approach to Build User's Psycho-Physiological Maps of Emotions Using Bio-Sensors," *Proc. IEEE RO-MAN 2006, 15th IEEE Int'l Symp. Robot and Human Interactive Comm., Session Emotional Cues in Human-Robot Interaction*, pp. 269-276, 2006.
- [112] R.W. Picard, E. Vyzas, and J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175-1191, Oct. 2001.
- [113] J. Wagner, N.J. Kim, and E. Andre, "From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification," *Proc. IEEE Int'l Conf. Multimedia and Expo*, pp. 940-943, 2005.
- [114] K. Kim, S. Bang, and S. Kim, "Emotion Recognition System Using Short-Term Monitoring of Physiological Signals," *Medical and Biological Eng. and Computing*, vol. 42, pp. 419-427, May 2004.
- [115] R.A. Calvo, I. Brown, and S. Scheduling, "Effect of Experimental Factors on the Recognition of Affective Mental States through Physiological Measures," *Proc. 22nd Australasian Joint Conf. Artificial Intelligence*, 2009.
- [116] J.N. Bailenson, E.D. Pontikakis, I.B. Mauss, J.J. Gross, M.E. Jabon, C.A.C. Hutcherson, C. Nass, and O. John, "Real-Time Classification of Evoked Emotions Using Facial Feature Tracking and Physiological Responses," *Int'l J. Human-Computer Studies*, vol. 66, pp. 303-317, 2008.
- [117] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Physiology-Based Affect Recognition for Computer-Assisted Intervention of Children with Autism Spectrum Disorder," *Int'l J. Human-Computer Studies*, vol. 66, pp. 662-677, 2008.
- [118] E. Vyzas and R.W. Picard, "Affective Pattern Classification," *Proc. AAAI Fall Symp. Series: Emotional and Intelligent: The Tangled Knot of Cognition*, pp. 176-182, 1998.
- [119] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion Recognition Using Bio-Sensors: First Steps towards an Automatic System," *Affective Dialogue Systems*, pp. 36-48, Springer, 2004.
- [120] O. AlZoubi, R.A. Calvo, and R.H. Stevens, "Classification of EEG for Emotion Recognition: An Adaptive Approach," *Proc. 22nd Australasian Joint Conf. Artificial Intelligence*, pp. 52-61, 2009.
- [121] A. Heraz and C. Frasson, "Predicting the Three Major Dimensions of the Learner's Emotions from Brainwaves," *World Academy of Science, Eng. and Technology*, vol. 25, pp. 323-329, 2007.
- [122] J. Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford Univ. Press, 1998.
- [123] J. Panksepp, "Emotions as Natural Kinds within the Mammalian Brain," *Handbook of Emotions*, pp. 137-156, Guilford, 2000.
- [124] M.H. Immordino-Yang and A. Damasio, "We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education," *Mind, Brain, and Education*, vol. 1, pp. 3-10, 2007.
- [125] J.K. Olofsson, S. Nordin, H. Sequeira, and J. Polich, "Affective Picture Processing: An Integrative Review of ERP Findings," *Biological Psychology*, vol. 77, pp. 247-265, Mar. 2008.
- [126] A. Bashashati, M. Fatourechi, R.K. Ward, and G.E. Birch, "A Survey of Signal Processing Algorithms in Brain-Computer Interfaces Based on Electrical Brain Signals," *J. Neural Eng.*, vol. 4, pp. R32-R57, June 2007.
- [127] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A Review of Classification Algorithms for EEG-Based Brain-Computer Interfaces," *J. Neural Eng.*, vol. 4, pp. R1-R13, 2007.
- [128] C.E. Osgood, W.H. May, and M.S. Miron, *Cross-Cultural Universals of Affective Meaning*. Univ. of Illinois Press, 1975.
- [129] C. Lutz and G. White, "The Anthropology of Emotions," *Ann. Rev. Anthropology*, vol. 15, pp. 405-436, 1986.
- [130] C.E. Osgood, *Language, Meaning, and Culture: The Selected Papers of C.E. Osgood*. Praeger Publishers, 1990.
- [131] A.V. Samsonovich and G.A. Ascoli, "Cognitive Map Dimensions of the Human Value System Extracted from Natural Language," *Proc. AGI Workshop Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms*, pp. 111-124, 2006.
- [132] M.A. Cohn, M.R. Mehl, and J.W. Pennebaker, "Linguistic Markers of Psychological Change Surrounding September 11, 2001," *Psychological Science*, vol. 15, pp. 687-693, Oct. 2004.
- [133] J. Pennebaker, M. Mehl, and K. Niederhoffer, "Psychological Aspects of Natural Language Use: Our Words, Our Selves," *Ann. Rev. Psychology*, vol. 54, pp. 547-577, 2003.
- [134] J. Kahn, R. Tobin, A. Massey, and J. Anderson, "Measuring Emotional Expression with the Linguistic Inquiry and Word Count," *Am. J. Psychology*, vol. 120, pp. 263-286, Summer 2007.
- [135] C.G. Shields, R.M. Epstein, P. Franks, K. Fiscella, P. Duberstein, S.H. McDaniel, and S. Meldrum, "Emotion Language in Primary Care Encounters: Reliability and Validity of an Emotion Word Count Coding System," *Patient Education and Counseling*, vol. 57, pp. 232-238, May 2005.
- [136] Y. Bestgen, "Can Emotional Valence in Stories Be Determined from Words," *Cognition and Emotion*, vol. 8, pp. 21-36, Jan. 1994.
- [137] J. Hancock, C. Landrigan, and C. Silver, "Expressing Emotion in Text-Based Communication," *Proc. SIGCHI*, 2007.
- [138] J. Pennebaker, M. Francis, and R. Booth, *Linguistic Inquiry and Word Count (LIWC): A Computerized Text Analysis Program*. Erlbaum Publishers, 2001.
- [139] C. Chung and J. Pennebaker, "The Psychological Functions of Function Words," *Social Comm.*, K. Fielder, ed., pp. 343-359, Psychology Press, 2007.
- [140] W. Weintraub, *Verbal Behavior in Everyday Life*. Springer, 1989.
- [141] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller, "Introduction to Wordnet: An On-Line Lexical Database," *J. Lexicography*, vol. 3, pp. 235-244, 1990.
- [142] C. Strapparava and A. Valitutti, "WordNet-Affect: An Affective Extension of WordNet," *Proc. Int'l Conf. Language Resources and Evaluation*, vol. 4, pp. 1083-1086, 2004.
- [143] M.M. Bradley and P.J. Lang, "Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings," Technical Report C1, The Center for Research in Psychophysiology, Univ. of Florida, 1999.
- [144] R.A. Stevenson, J.A. Mikels, and T.W. James, "Characterization of the Affective Norms for English Words by Discrete Emotional Categories," *Behavior Research Methods*, vol. 39, pp. 1020-1024, 2007.
- [145] M.M. Bradley and P.J. Lang, "Affective Norms for English Text (ANET): Affective Ratings of Text and Instruction Manual," Technical Report No. D-1, Univ. of Florida, 2007.
- [146] A. Gill, R. French, D. Gergle, and J. Oberlander, "Identifying Emotional Characteristics from Short Blog Texts," *Proc. 30th Ann. Conf. Cognitive Science Soc.*, B.C. Love, K. McRae, and V.M. Sloutsky, eds., pp. 2237-2242, 2008.
- [147] T. Landauer, D. McNamara, S. Dennis, and W. Kintsch, *Handbook of Latent Semantic Analysis*. Erlbaum, 2007.
- [148] K. Lund and C. Burgess, "Producing High-Dimensional Semantic Spaces from Lexical Co-Occurrence," *Behavior Research Methods Instruments and Computers*, vol. 28, pp. 203-208, 1996.
- [149] E. Breck, Y. Choi, and C. Cardie, "Identifying Expressions of Opinion in Context," *Proc. 20th Int'l Joint Conf. Artificial Intelligence*, 2007.
- [150] H. Liu, H. Lieberman, and S. Selker, "A Model of Textual Affect Sensing Using Real-World Knowledge," *Proc. Eighth Int'l Conf. Intelligent User Interfaces*, pp. 125-132, 2003.
- [151] M. Shaikh, H. Prendinger, and M. Ishizuka, "Sentiment Assessment of Text by Analyzing Linguistic Features and Contextual Valence Assignment," *Applied Artificial Intelligence*, vol. 22, pp. 558-601, 2008.
- [152] C. Akkaya, J. Wiebe, and R. Mihalcea, "Subjectivity Word Sense Disambiguation," *Proc. 2009 Conf. Empirical Methods in Natural Language Processing*, 2009.
- [153] A. Jaimes and N. Sebe, "Multimodal Human-Computer Interaction: A Survey," *Computer Vision and Image Understanding*, vol. 108, pp. 116-134, 2007.
- [154] R. Sharma, V.I. Pavlovic, and T.S. Huang, "Toward Multimodal Human-Computer Interface," *Proc. IEEE*, vol. 86, no. 5, pp. 853-869, May. 1998.
- [155] C.O. Alm, D. Roth, and R. Sproat, "Emotions from Text: Machine Learning for Text-Based Emotion Prediction," *Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 347-354, 2005.
- [156] T. Danisman and A. Alpkocak, "Feeler: Emotion Classification of Text Using Vector Space Model," *Proc. AISB 2008 Convention, Comm., Interaction and Social Intelligence*, 2008.
- [157] C. Strapparava and R. Mihalcea, "Learning to Identify Emotions in Text," *Proc. 2008 ACM Symp. Applied Computing*, pp. 1556-1560, 2008.

- [158] S.K. D'Mello, S.D. Craig, J. Sullins, and A.C. Graesser, "Predicting Affective States Expressed through an Emote-Aloud Procedure from AutoTutor's Mixed-Initiative Dialogue," *Int'l J. Artificial Intelligence in Education*, vol. 16, pp. 3-28, 2006.
- [159] S. D'Mello, N. Dowell, and A. Graesser, "Cohesion Relationships in Tutorial Dialogue as Predictors of Affective States," *Proc. 2009 Conf. Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling*, V. Dimitrova, R. Mizoguchi, B. du Boulay, and A.C. Graesser, eds., pp. 9-16, 2009.
- [160] C. Ma, A. Osherenko, H. Prendinger, and M. Ishizuka, "A Chat System Based on Emotion Estimation from Text and Embodied Conversational Messengers," *Proc. 2005 Int'l Conf. Active Media Technology*, pp. 546-548, 2005.
- [161] A. Valitutti, C. Strapparava, and O. Stock, "Lexical Resources and Semantic Similarity for Affective Evaluative Expressions Generation," *Proc. Int'l Conf. Affective Computing and Intelligent Interaction*, pp. 474-481, 2005.
- [162] W.H. Lin, T. Wilson, J. Wiebe, and A. Hauptmann, "Which Side Are You On? Identifying Perspectives at the Document and Sentence Levels," *Proc. 10th Conf. Natural Language Learning*, pp. 109-116, 2006.
- [163] Z. Zeng, Y. Hu, G. Roisman, Z. Wen, Y. Fu, and T. Huang, "Audio-Visual Emotion Recognition in Adult Attachment Interview," *Proc. Int'l Conf. Multimodal Interfaces*, J.Y.F.K.H. Quek, D.W. Massaro, A.A. Alwan, and T.J. Hazen, eds., pp. 139-145, 2006.
- [164] Y. Yoshitomi, K. Sung-Il, T. Kawano, and T. Kilazoe, "Effect of Sensor Fusion for Recognition of Emotional States Using Voice, Face Image and Thermal Image of Face," *Proc. Int'l Workshop Robot and Human Interactive Comm.*, pp. 178-183, 2000.
- [165] L. Chen, T. Huang, T. Miyasato, and R. Nakatsu, "Multimodal Human Emotion/Expression Recognition," *Proc. Third IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 366-371, 1998.
- [166] B. Dasarathy, "Sensor Fusion Potential Exploitation: Innovative Architectures and Illustrative Approaches," *Proc. IEEE*, vol. 85, no. 1, pp. 24-38, Jan. 1997.
- [167] G. Caridakis, L. Malatesta, L. Kessous, N. Amir, A. Paouzaoui, and K. Karpouzis, "Modeling Naturalistic Affective States via Facial and Vocal Expression Recognition," *Proc. Int'l Conf. Multimodal Interfaces*, pp. 146-154, 2006.
- [168] J. Liscombe, G. Riccardi, and D. Hakkani-Tür, "Using Context to Improve Emotion Detection in Spoken Dialog Systems," *Proc. Ninth European Conf. Speech Comm. and Technology*, 2005.
- [169] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke, "Prosody-Based Automatic Detection of Annoyance and Frustration in Human-Computer Dialog," *Proc. Int'l Conf. Spoken Language Processing*, pp. 2037-2039, 2002.
- [170] K. Scherer and H. Ellgring, "Multimodal Expression of Emotion: Affect Programs or Componential Appraisal Patterns?," *Emotion*, vol. 7, pp. 158-171, 2007.
- [171] G. Castellano, L. Kessous, and G. Caridakis, "Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech," *Affect and Emotion in Human-Computer Interaction*, pp. 92-103, Springer, 2008.
- [172] S. Afzal and P. Robinson, "Natural Affect Data: Collection & Annotation in a Learning Context," *Proc. Third Int'l Conf. Affective Computing and Intelligent Interaction and Workshops*, pp. 1-7, 2009.
- [173] J. Cohn and K. Schmidt, "The Timing of Facial Motion in Posed and Spontaneous Smiles," *Int'l J. Wavelets, Multiresolution and Information Processing*, vol. 2, pp. 1-12, 2004.
- [174] P. Ekman, W. Friesen, and R. Davidson, "The Duchenne Smile—Emotional Expression and Brain Physiology .2," *J. Personality and Social Psychology*, vol. 58, pp. 342-353, 1990.
- [175] A. Kapoor and R.W. Picard, "Multimodal Affect Recognition in Learning Environments," *Proc. 13th Ann. ACM Int'l Conf. Multimedia*, pp. 677-682, 2005.
- [176] A. Kapoor, B. Burleson, and R. Picard, "Automatic Prediction of Frustration," *Int'l J. Human-Computer Studies*, vol. 65, pp. 724-736, 2007.
- [177] S. D'Mello and A. Graesser, "Multimodal Semi-Automated Affect Detection from Conversational Cues, Gross Body Language, and Facial Features," *User Modeling and User-Adapted Interaction*, vol. 10, pp. 147-187, 2010.
- [178] T. Baenziger, D. Grandjean, and K.R. Scherer, "Emotion Recognition from Expressions in Face, Voice, and Body. The Multimodal Emotion Recognition Test (MERT)," *Emotion*, vol. 9, pp. 691-704, 2009.
- [179] C. Busso et al., "Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information," *Proc. Int'l Conf. Multimodal Interfaces*, T.D.R. Sharma, M.P. Harper, G. Lazzari, and M. Turk, eds., pp. 205-211, 2004.
- [180] Z.J. Chuang and C.H. Wu, "Multi-Modal Emotion Recognition from Speech and Text," *Int'l J. Computational Linguistics and Chinese Language Processing*, vol. 9, pp. 1-18, 2004.
- [181] A. Esposito, "Affect in Multimodal Information," *Affective Information Processing*, pp. 203-226, Springer, 2009.
- [182] R. Baker, S. D'Mello, M. Rodrigo, and A. Graesser, "Better to Be Frustrated than Bored: The Incidence and Persistence of Affect during Interactions with Three Different Computer-Based Learning Environments," *Int'l J. Human-Computer Studies*, vol. 68, no. 4, pp. 223-241, 2010.
- [183] S. D'Mello, R. Picard, and A. Graesser, "Towards an Affect-Sensitive AutoTutor," *IEEE Intelligent Systems*, vol. 22, no. 4, pp. 53-61, July 2007.
- [184] B. Lehman, M. Matthews, S. D'Mello, and N. Person, "What Are You Feeling? Investigating Student Affective States during Expert Human Tutoring Sessions," *Proc. Ninth Int'l Conf. Intelligent Tutoring Systems*, B. Woolf, E. Aimeur, R. Nkambou, and S. Lajoie, eds., pp. 50-59, 2008.
- [185] B. Lehman, S. D'Mello, and N. Person, "All Alone with Your Emotions: An Analysis of Student Emotions during Effortful Problem Solving Activities," *Proc. Workshop Emotional and Cognitive Issues in ITS at the Ninth Int'l Conf. Intelligent Tutoring Systems*, 2008.
- [186] S. D'Mello, S. Craig, K. Fike, and A. Graesser, "Responding to Learners' Cognitive-Affective States with Supportive and Shakeup Dialogues," *Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction*, pp. 595-604, Springer, 2009.
- [187] W. Burleson and R.W. Picard, "Gender-Specific Approaches to Developing Emotionally Intelligent Learning Companions," *IEEE Intelligent Systems*, vol. 22, no. 4, pp. 62-69, July/Aug. 2007.
- [188] S. D'Mello, B. Lehman, J. Sullins, R. Daigle, R. Combs, K. Vogt, L. Perkins, and A. Graesser, "A Time for Emoting: When Affect-Sensitivity Is and Isn't Effective at Promoting Deep Learning," *Proc. 10th Int'l Conf. Intelligent Tutoring Systems*, J. Kay and V. Aleven, eds., to appear.
- [189] I. Arroyo, B. Woolf, D. Cooper, W. Burleson, K. Muldner, and R. Christopherson, "Emotion Sensors Go to School," *Proc. 14th Int'l Conf. Artificial Intelligence in Education*, V. Dimitrova, R. Mizoguchi, B. Du Boulay, and A. Graesser, eds., pp. 17-24, 2009.
- [190] J. Robison, S. McQuiggan, and J. Lester, "Evaluating the Consequences of Affective Feedback in Intelligent Tutoring Systems," *Proc. Int'l Conf. Affective Computing and Intelligent Interaction*, pp. 37-42, 2009.
- [191] C. Conati and H. Maclaren, "Empirically Building and Evaluating a Probabilistic Model of User Affect," *User Modeling and User-Adapted Interaction*, vol. 19, pp. 267-303, 2009.
- [192] R. El Kaliouby, R. Picard, and S. Baron-Cohen, "Affective Computing and Autism," *Annals New York Academy of Sciences*, vol. 1093, pp. 228-248, 2006.
- [193] B. Parkinson, *Ideas and Realities of Emotion*. Routledge, 1995.
- [194] J. Gross, "Emotion Regulation," *Handbook of Emotions*, M. Lewis, J. Haviland-Jones, and L. Barrett, eds., third ed., pp. 497-512, Guilford, 2008.
- [195] J. Gross, "The Emerging Field of Emotion Regulation: An Integrative Review," *Rev. of General Psychology*, vol. 2, pp. 271-299, 1998.
- [196] A. Ohman and J. Soares, "Unconscious Anxiety—Phobic Responses to Masked Stimuli," *J. Abnormal Psychology*, vol. 103, pp. 231-240, May 1994.
- [197] A. Ohman and J. Soares, "On the Automatic Nature of Phobic Fear—Conditioned Electrodermal Responses to Masked Fear-Relevant Stimuli," *J. Abnormal Psychology*, vol. 102, pp. 121-132, Feb. 1993.
- [198] M.S. Clark and I. Brissette, "Two Types of Relationship Closeness and Their Influence on People's Emotional Lives," *Handbook of Affective Sciences*, K.R.S. Richard, J. Davidson, and H. Hill Goldsmith, eds., pp. 824-835, Oxford Univ. Press, 2003.

- [199] A.J. Fridlund, P. Ekman, and H. Oster, "Facial Expressions of Emotion," *Nonverbal Behavior and Communication*, A.W. Siegman and S. Feldstein, eds., second ed., pp. 143-223, Erlbaum, 1987.
- [200] W. Ruch, "Will the Real Relationship between Facial Expression and Affective Experience Please Stand Up—The Case of Exhilaration," *Cognition and Emotion*, vol. 9, pp. 33-58, Jan. 1995.
- [201] K. Scherer and H. Ellgring, "Are Facial Expressions of Emotion Produced by Categorical Affect Programs or Dynamically Driven by Appraisal?" *Emotion*, vol. 7, pp. 113-130, Feb. 2007.
- [202] G. Bower, "Mood and Memory," *Am. Psychologist*, vol. 36, pp. 129-148, 1981.
- [203] G. Mandler, "Emotion," *Cognitive Science (Handbook of Perception and Cognition)*, B.M. Bly and D.E. Rumelhart, eds., second ed., Academic Press, 1999.
- [204] N. Stein and L. Levine, "Making Sense Out of Emotion," *Memories Thoughts, and Emotions: Essays in Honor of George Mandler*, A.O.W. Kessen and F. Kraik, eds., pp. 295-322, Erlbaum, 1991.
- [205] J. Bachorowski and M. Owren, "Vocal Expression of Emotion—Acoustic Properties of Speech Are Associated with Emotional Intensity and Context," *Psychological Science*, vol. 6, pp. 219-224, July 1995.
- [206] D. Keltner and J. Haidt, "Social Functions of Emotions," *Emotions: Current Issues and Future Directions*, T.J. Mayne and G.A. Bonanno, eds., pp. 192-213, Guilford Press, 2001.
- [207] C. Conati and H. Maclaren, "Modeling User Affect from Causes and Effects," *Proc. First and 17th Int'l Conf. User Modeling, Adaptation and Personalization*, 2009.
- [208] D.A. Norman, *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, 2005.
- [209] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, vol. 20, pp. 37-46, 1960.
- [210] C. Robson, *Real World Research: A Resource for Social Scientist and Practitioner Researchers*. Blackwell, 1993.



**Rafael A. Calvo** received the PhD degree in artificial intelligence applied to automatic document classification and has also worked at Carnegie Mellon University and the Universidad Nacional de Rosario, and as a consultant for projects worldwide. He is a senior lecturer in the School of Electrical and Information Engineering at the University of Sydney, and the director of the Learning and Affect Technologies Engineering (Latte) Research Group. He has authored numerous publications in the areas of affective computing, learning systems, and Web engineering. He is the recipient of three teaching awards. He is a senior member of the IEEE and a member of the IEEE Computer Society. He is an associate editor of the *IEEE Transactions on Affective Computing*.



**Sidney D'Mello** received the BS degree in electrical engineering from Christian Brothers University in 2002 and the MS degree in mathematical sciences and the PhD degree in computer science from the University of Memphis in 2004 and 2009, respectively. He is presently a postdoctoral researcher in the Institute for Intelligent Systems and the Department of Psychology at the University of Memphis. His primary research interests are in the cognitive and affective sciences, artificial intelligence, human-computer interaction, and the learning sciences. More specific interests include affective computing, artificial intelligence in education, speech recognition and natural language understanding, and computational models of human cognition. He has published more than 70 papers in these areas. He is a member of the IEEE Computer Society.