INTERNATIONAL JOURNAL OF OCCUPATIONAL SAFETY AND ERGONOMICS (*JOSE*) 2003, VOL. 9, NO. 4, 539–551

# Affective Computing—A Rationale for Measuring Mood With Mouse and Keyboard

**Philippe Zimmermann**
**Sissel Guttormsen**
**Brigitta Danuser**
**Patrick Gomez**

Swiss Federal Institute of Technology, Zürich, Switzerland

Emotions are an increasingly important factor in Human-Computer Interaction (HCI). Up to the present, emotion recognition in HCI implies the use of explicit or intrusive methods, for example, video cameras or physiological measurements. We are developing and evaluating a method for the measurement of affective states through motor-behavioral parameters from standard input devices (mouse and keyboard).

affective computing    mood    emotions    HCI

## 1. INTRODUCTION

The ability to recognize, interpret, and express emotions plays a key role in human communication and increasingly in Human-Computer Interaction (HCI). Recent research could demonstrate (Reeves & Nass, 1996) that humans have an inherent tendency to interact with computers in a natural and social way, mirroring interactions between humans in social situations. Computer systems are currently not able to recognize or respond to these complexities of a natural social interaction.

There has been increasing interest in HCI research in building emotionally intelligent interactive systems that can express and respond to human emotions

Correspondence and requests for offprints should be sent to Philippe Zimmermann, Swiss Federal Institute of Technology, Institute for Hygiene and Applied Physiology, Clausiusstrasse 25, 8092 Zürich, Switzerland. E-mail: <zimmermann@iha.bepr.ethz.ch >.

(Picard, 1997). One of the challenges in building emotionally intelligent systems is the automatic recognition of affective states. Humans use different sources of information to assess a person's emotions, including causal information context and individual traits, as well as information on the person's recognizable bodily reactions. For a computer system, some of this information is difficult to access, for example, the person's traits, it is even more difficult to assess all the multifaceted information and to integrate it into a complete image of the user's affective state. Nonetheless, there has been extensive research in the field of affect measurement that can be roughly grouped into three areas: physiological, behavioral, and psychological approaches.

In this paper we have described a feasibility study of a new behavioral method for measuring user affect with standard computer input devices, mouse and keyboard. We are developing an unobtrusive, noninvasive measurement method, which is able to assess user affect in parallel with task processing, using only inexpensive standard computer devices (in contrast to, e.g., questionnaires, which are completed before or after the task). The method is based on the effects of affect on motor-behavior. It extracts motor-behavioral parameters from log-files of mouse and keyboard actions, which can be used to analyze correlations with affective state.

## 2. AFFECT, EMOTION, MOOD

Although there is no commonly accepted definition of emotion, most authors agree on the fact that emotion is a multifaceted phenomenon that encompasses a diversity of processes, such as appraisal, facial expressions, bodily responses, feeling states, action tendencies, or coping strategies (e.g., Frijda, 1986). Usually two defining aspects are emphasized: somatic reactions and affective feelings.

The terms affect, emotion, and mood are often used interchangeably without clear definition, which leads to difficulties when comparing different research results and methods. We use the term *affect* as the most generalized of the three terms. It may be used to refer to both emotions and moods (Forgas, 1995). An *emotion* has the properties of a reaction. It often has a specific cause, a stimulus or preceding thought, it is usually an intense experience of short duration—seconds to minutes—and the person is typically well aware of it. On the other hand, a *mood* tends to be subtler, longer lasting, less intensive, more in the background, giving the affective state of a person a tendency in a positive or negative direction. Moods tend to be nonspecific compared to emotions, which are usually specific and focused on an identifiable person, object, or event. In psychological research, it has been shown that mood affects memory, assessment, search strategy (e-commerce), judgment, expectations, opinions, and motor behavior (Derbaix & Pecheux, 1999).

In contrast to emotions, people may not be aware of their mood until their attention is drawn to it. Moods tend to bias which emotions are experienced, lowering the activation thresholds for mood-related emotions or serve as an "affective filter." Emotions, on the other hand, often cause moods (Brave & Nass, 2003). Hence it is important to consider the biasing effect of moods, for example, in usability studies: Participants in a good mood are likely to experience positive emotions, participants in a bad mood experience more likely negative emotions. An affective computer can take advantage of this biasing effect of mood by presenting stimuli that sustain the desired moods or, alternatively, counteract undesired mood states. Frustrated users, for example, could be guided to a different task, focus on a different aspect of the current work, or simply be advised to take a break.

Positive mood also decreases risk-taking, in accordance with an evolutionary view of preserving the positive mood. An e-commerce website could take advantage of this fact by predicting that a low-risk purchase is more likely during a good mood whereas a high-risk purchase may be more likely in a neutral or negative mood (Brave & Nass, 2003).

## 3. STRUCTURE AND LABELING OF AFFECT

Neuropsychology differentiates three brain regions concerned with affect: the thalamus, the limbic system, and the cortex. All external sensory input is received by the thalamus, which sends information simultaneously to the cortex for higher level processing, and directly to the limbic system (LeDoux, 1995). The direct thalamic-limbic pathway accounts for the more primitive emotions, such as startle-based fear or aversions or attractions. Objects that appear or move suddenly (e.g., pop-up windows) or loud sounds trigger startle-based fear.

Secondary, more complex emotions—for example, pride or satisfaction—result from activation of the limbic system after processing of stimulus information in the cortex, with diverse theories arguing about the amount of involvement of the limbic system (e.g., Ekman, 1992; Wierzbicka, 1992). Most of the emotions important in an HCI context fall into this category.

Evolutionary theorists argue that all emotions—including complex emotions—are innate, evolved because of specific environmental impact. On the other hand, many emotion theorists argue that emotions are almost entirely learned social constructs, emphasizing the role of cortical processing, and reducing the impact of the limbic system to influences along several dimensions of affect or even a simple on-and-off manner.

Empirical work (Lang, Bradley, & Cuthbert, 1990; Mehrabian, 1970; Russell, 1980) has repeatedly confirmed that differences in affective meaning among stimuli—words, objects, events—can succinctly be described by three basic emotional dimensions: a dimension of affective valence, also called pleasure, ranging from positive (*pleasant*) to negative (*unpleasant*), one of arousal, ranging from *calm* (low-arousal) to *excited* (high-arousal), and one called dominance or control, ranging from *controlled* to *in control*. The valence and arousal dimensions are primary, and they typically account for most of the variance in emotional judgments (Bradley & Lang, 1994).

Between the theories arguing for innate emotions and the theories defending the view of emotions being learned constructs, lie those who believe that there are basic emotions. There is a set of emotions that is innate and shared by all humans, more complex emotions are seen as combinations of these basic emotions. Diverse writers have proposed that there are from 2 to 20 basic emotions (e.g., Plutchik, 1980), such as joy, fear, love, surprise, sadness, and so forth. It is important to say that the dimensional and specific state view of emotion are considered complementary approaches to emphasize different factors influencing emotional responses.

## 4. MOOD MEASUREMENT METHODS

### 4.1. Physiological

There is evidence in literature suggesting that physiological signals have characteristic patterns for specific affective states (e.g., Ekman, Levenson, & Friesen, 1983). Several studies have even provided evidence for a correlation between physiological variables and the affective dimensions of valence and arousal (e.g., Gomez & Danuser, 2002; Lang, Greenwald, Bradley, & Hamm, 1993), thus suggesting that emotion is fundamentally organized by these two parameters. Physiological signals such as skin conductance, heart rate, blood pressure, respiration, pupillary dilation, electroencephalography (EEG), or muscle action potentials can provide information regarding the intensity and quality of an individual's internal affect experience.

In our experiment, we used a combination of physiological signals as a primary source to verify self-assessment results of affect, also because it was possible to get data simultaneously with the task performed.

## 4.2. Psychological

Self-reports are widely used and still serve as a primary method for ascertaining emotion and mood. There exist literally dozens of affect inventories: verbal descriptions of an emotion or emotional state, rating scales, standardized checklists, questionnaires, semantic and graphical differentials, or projective methods. Many of these methods are based on the dimensional model of affect. The Semantic Differential Scale devised by Mehrabian and Russell (1974) consists of a set of 18 bipolar adjective pairs that generate scores on the valence, arousal, and dominance scales. We used a similar questionnaire with bipolar adjective pairs in German, adapted from the Mehrdimensionaler Befindlichkeitsfragebogen (MDBF; multi-dimensional mood questionnaire; Steyer, Schwenkmezger, Notz, & Eid, 1997).

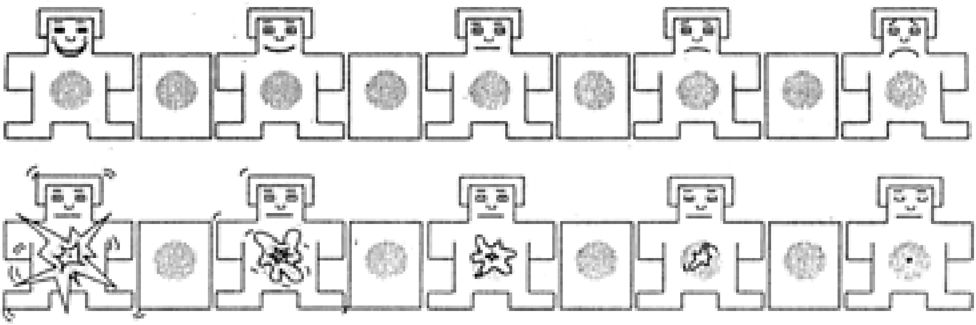

**Figure 1. The scales valence (top) and arousal (bottom) of the Self-Assessment Manikin (SAM).**

The Self-Assessment Manikin (SAM), devised by Lang (1980), is designed to assess the dimensions valence, arousal, and dominance directly by means of three sets of graphical manikins (see Figure 1 for valence and arousal dimensions). It has been extensively tested in conjunction with the International Affective Picture System (Center for the Study of Emotion and Attention, 1999). This graphical version takes only a very short time to complete (5 to 10 s) and there is little chance of confusion with terms as in the verbal version.

## 4.3. Behavioral

There exists a broad field of behavioral methods for the measurement of affect: facial expressions, voice modulation, gestures, posture, cognitive performance, cognitive strategy, motor behavior (e.g., hand muscles, head movement), and so forth. Behavioral measurement methods are based on the fact that the body usually responds physically to an emotion (e.g., changes in muscle tension, coordination, strength, frequency) and that the motor system acts as a carrier for communicating affective state. Especially promising for these methods is that humans also use many of these signals in everyday life to judge the affective state of other people. In contrast to physiological methods can behavioral methods be applied in a noninvasive way (although video cameras used in face recognition may be considered obtrusive).

The two most prominent of these methods are face recognition and voice intonation analysis, which both have been investigated in many research projects (Cowie et al., 2001). There are also a few existing projects dealing with motor behavior in HCI, for example, the analysis of mouse clicking behavior after frustrating events during a computer task (Scheirer, Fernandez, Klein, & Picard, 2002), where four distinct patterns of mouse clicking could be found or the visual comparison of mouse movement patterns on an e-commerce site for user modeling (Lockerd & Mueller, 2001).

## 5. MOOD INDUCTION METHODS

There are many techniques for the induction of moods. The main methods include imagination, hypnosis, music, social interaction, imitation of a facial expression, or memories for positive and negative life events (Schneider, Gur, Gur, & Muenz, 1994). To increase the effectiveness of such procedures, some authors combined different types of methods.

One of the most widely used approaches to induce a given mood, is the Velten procedure (Velten, 1968). In this procedure participants are given a number of statements to induce a specific mood. But aside from issues of what participants actually experience when the Velten induction is used, research has indicated that the effects of the Velten induction do not last more than a very short time, approximately 10 min, and the effects can be dissipated easily by intervening tasks.

A method for inducing longer lasting moods is the use of film clips including sound. The use of films to manipulate mood has been tried and tested both in laboratory and field research and has been found to produce salient and enduring moods (Forgas, & Moylan, 1987). Films have a relatively high degree of ecological validity, in so far as emotions are often evoked by dynamic visual and auditory stimuli that are external to the individual (Gross & Levenson, 1995). One important limitation of the use of films in this context is, however, that there are no widely accepted sets of emotion eliciting film stimuli, as is the case with picture sets.

## 6. DISCUSSION

Existing methods for measuring affect all have a number of drawbacks or are not applicable in the field of HCI. Physiological signals, for example, are measured with a wide variety of instruments and sensors. Unfortunately, using physiological signals necessitates specialized and frequently expensive equipment and technical expertise to run the equipment, which makes this method suitable for lab experiments but not for applied use. Sensors have to be attached directly to the body, which can be considered obtrusive or even invasive by many participants. Furthermore, it can be quite difficult to separate confounding factors influencing physiological reactions in order to attribute significant changes to the experimental variable (Kramer, 1991). The extent to which emotions can be distinguished on the basis of physiological activity alone remains a debated issue.

Self-reports are still the primary method for assessing affect. They also pose a number of problems. People can feel pressed to give wrong answers, for example, for social desirability reasons; self-reports are either retrospective and events in the past are subject to distortions or they are concurrent, interrupting the user during the task; questionnaires can only assess the conscious experience of emotion and mood, but much of the affective experience is nonconscious; finally, questions about affect are potentially influenced by when they are asked, because of the eventually different present mood and memory degradation.

Behavioral methods like face and voice recognition have high recognition accuracy, in some systems up to 98% on a small set of emotions. However, these methods are tested almost exclusively on “produced” affect expressions, for example, from actors, rather than on actual emotions. It can be expected that recognition accuracy would drop heavily in natural situations. In addition, people can consider recording devices like video cameras obtrusive.

In our project, we are developing a behavioral measurement method analyzing data from keyboard and mouse usage. We believe that patterns in different parameters like mouse click counts, mouse speed, or keystroke speed correlate with the scores on the valence and arousal dimensions (e.g., Clark, 1983). The advantages of this method are striking: an unobtrusive, noninvasive way of measuring affect directly on the computer without any additional devices, concurrent with the task performed.

# 7. METHOD

This experiment was designed to investigate the influence of induced affects on motor-behavior parameters while completing a computer task. Film clips were used as affect elicitors. The task was an online-shopping task, which required participants to shop on an e-commerce website for office supplies. Ninety-six students (46 female, 50 male, aged between 17 and 38) participated in this experiment.

## 7.1. Design

The experiment applied a 5 × 1 mixed design. The 5 different mood states, PVHA, PVLA, NVHA, NVLA, and nVnA (P—positive, N—negative, H—high, L—low, n—neutral, V—valence, A—arousal) serve as a between group factor (independent variable). The control run (neutral mood induction) as compared to the induction run (5 different mood inductions) is the within group factor.

## 7.2. Mood Induction

**TABLE 1. Content of the Film Clips Used in the Experiment**

| Mood | Content |
|---|---|
| Neutral (control) | Educational movie about the characteristics of different materials (e.g., wood, rock, concrete) |
| Neutral | Documentary about the architect Louis Kahn and his work |
| PV/HA | Clips of different sports (e.g., surfing, skiing, climbing) with rock and pop music |
| PV/LA | Takes of landscapes and animals with classical music |
| NV/HA | Extract from “Deer Hunter” (Cimino, 1978), depicting captives in Vietnam |
| NV/LA | Documentary about the Borinage (Jean, 1999), an old mining area and now a slum in Belgium |

*Notes.* PV—Positive Valence, HA—High Arousal, NV—Negative Valence, LA—Low Arousal.

We used film clips to induce in the first step a neutral mood, in the second step one of 5 different moods. The six clips (see Table 1) were selected in a preliminary study for their ability to (a) induce moods effectively and reliably and (b) stay within ethical guidelines. The film clips were between 7 and 11 min long.

## 7.3. Questionnaires

Affective state was quantified with the rating scales of the graphical Self-Assessment Manikin (Lang, 1980) and a semantic differential with 6 bipolar adjective pairs in German. Participants were asked to first rate their momentary state on the 9-point graphical scale on the dimensions valence and arousal (see Figure 1), afterwards on the 9-point semantic scale. All questionnaires were completed electronically on the computer.

## 7.4. Task

Participants had to shop on an e-commerce website for office supplies. The task was selected because of its applied, real-world nature with little impact on the induced mood. Each task was divided into 8 subtasks telling the participant to buy one of the products from the website (e.g., “Buy 6000 sheet of fanfold paper, 70 g/$m^2$.”) or—as a last task—to write a predefined message to the shop operator.

## 7.5. Physiological Measurements

Several physiological parameters were measured concurrent with the task and the film clips. The parameters included respiration, pulse, skin conductance level, and corrugator activity. Respiration was measured noninvasively using a volume calibrated respiratory inductive plethysmograph (Respitrace PLUS, SensorMedics, USA). The heart rate was measured with three sensors on the torso, skin conductance level, and corrugator supercilii EMG were measured with sensors on the forehead and the left hand with the Varioport Measurement System (Becker Meditec, Karlsruhe, Germany).

## 7.6. Behavioral Measurements

During the experiment, all mouse and keyboard actions were recorded to a log-file by software running in the background of the operating system, invisible to the participant. Log-file entries contain exact time and type of action (e.g., mouse button down, mouse position *x* and *y* coordinates, which key pressed).

## 7.7. Procedure

After arrival at the laboratory, sensors and respiratory measurement bands were attached and connected, and then participants were asked to complete participant and health data questionnaires on the computer. All instructions during the experiment were written and given at the appropriate stages on the computer interface. The procedure was automated (see Figure 2).

The participants first familiarized themselves with the online-shopping task and then indicated their mood on the graphical and verbal differentials. Afterwards, the first movie clip, expected to be affectively neutral, was presented (control run). Participants then filled out the mood assessment questionnaires, completed the online-shopping task, and filled out the questionnaires again. Then the second movie clip was randomly chosen, inducing one of the moods, PVHA, PVLA, NVHA, NVLA, nVnA (P—positive, N—negative, H—high, L—low, n—neutral, V—valence, A—arousal). The film was followed by graphical and verbal differential questionnaires, then the task, and the two questionnaires again. The experiment ended after 1.5 to 2 hrs.

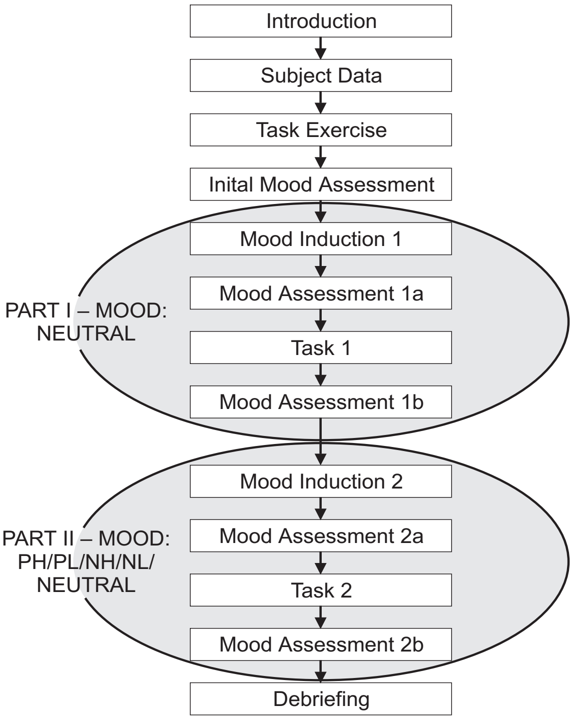


**Figure 2. Procedure of the experiment.** *Notes.* PH—positive valence, high arousal; PL—positive valence, low arousal; NH—negative valence, high arousal; NL—negative valence, low arousal.

## 8. ANALYSIS

The collected data will be analyzed in three steps. In a first step, the independent measure is the mood induction procedure of the 5 different affective states, PVHA, PVLA, NVHA, NVLA, nVnA (P—positive, N—negative, H—high, L—low, n—neutral, V—valence, A–arousal). The mood changes of participants (compared to neutral state) are the dependent measure, quantified with the rating scales of the SAM and MDBF. First preliminary results show that the film clips were effective in inducing the expected mood changes. In the second step, we will check if the data from the self-assessment questionnaires can be validated with the physiological measurements.

In the third step, we will analyze the mouse and keyboard actions from the log-files. The following parameters will be tested for correlations with mood state (the list may be extended): number of mouse clicks per minute, average duration of mouse clicks (from button-down until button-up event), total distance of mouse movements in pixels, average distance of a single mouse movement, number and length of pauses in mouse movement, number of events “heavy mouse movement” (more than 5 changes in direction in 2 s), maximum, minimum and average mouse speed, keystroke rate per second, average duration of keystroke (from key-down until key-up event), and performance.

## 9. CONCLUSIONS

We have described a rationale of a method for measuring mood with mouse and keyboard. Given that the method proves successful, it would make other affect assessing methods like face recognition or online mood assessing questionnaires obsolete, and we would have a simple but effective alternative. If a system knows how its users feel, it can appropriately react to these moods.

Beyond the development of affect recognition methods, several open questions remain: What are the adequate emotional reactions of a system that knows how you feel? Should a system try to change your mood? How could it do that? With which emotions should HCI research be concerned? How accurate does the measurement have to be? These are interesting research questions to be addressed.

# REFERENCES

Bradley, M.M., & Lang, P.J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*, 49–59.

Brave, S., & Nass, C. (2003). Emotion in human-computer-interaction. In J.A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook* (pp. 81–96). Mahvah, NJ, USA: Erlbaum.

Center for the Study of Emotion and Attention. (1999). *The international affective picture system.* Gainesville, FL, USA: Center for Research in Psychophysiology, University of Florida.

Clark, D.M. (1983). On the induction of depressed mood in the laboratory: Evaluation and comparison of the Velten and musical procedures. *Advances in Behavior Research and Theory*, *5*, 27–49.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, et al. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, *18*(1), 32–80.

Derbaix, C., & Pecheux, C. (1999). Mood and children: Proposition of a measurement scale. *Journal of Economic Psychology*, *20,* 571–591.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*, 169–200.

Ekman, P., Levenson, R.W., & Friesen, W.V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221, 1208–1209.

Forgas, J.P. (1995). Mood and judgments: The Affect Infusion Model (AIM). *Psychological Bulletin*, 117, 39–66.

Forgas, J.P., & Moylan, S. (1987). After the movies: The effects of transient mood states on social judgments. *Personality and Social Psychology Bulletin, 13*, 478–489.

Frijda, N.H. (1986). *The emotions*. Cambridge, UK: Cambridge University Press.

Gerrards-Hesse, A., Spies, K., Hesse, F.W. (1994). Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology*, 85, 55–78.

Gomez, P., & Danuser, B. (2002). Breathing responses to music and noises. *International Journal of Psychophysiology, 45*, 72–73.

Gross, J.J., & Levenson, R.W. (1995). Emotion elicitation using films. *Cognition and Emotion*, *9*(1), 87–108.

Kramer, A.F. (1991). Physiological metrics of mental workload: A review of recent progress. In D.L. Damos (Ed.). *Multiple-task-performance* (pp. 329–360). London, UK: Taylor & Francis.

Lang, P.J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J.B. Sidowski, J.H. Johnson, & T.A. Williams (Eds.). *Technology in mental health care delivery systems* (pp. 119–l37). Norwood, NJ, USA: Ablex.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*, 377–398.

Lang, P.J., Greenwald, M.K., Bradley, M.M., & Hamm, A.O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*, 261–273.

LeDoux, J.E. (1995). Emotion: Clues from the brain. *Annual Review of Psychology, 46*, 209–235.

Lockerd, A., & Mueller, F. (2001). Cheese: Tracking mouse movement activity on websites, a tool for user modeling. In *CHI 2001 extended abstracts on human factors in computer systems, Seattle, Washington* (pp. 279–280). New York, NY, USA: ACM Press.

Mehrabian, A. (1970). A semantic space for nonverbal behavior. *Journal of Consulting and Clinical Psychology*, *35*, 248–257.

Mehrabian, A., & Russell, J.A. (1974). *An approach to environmental psychology*. Cambridge, MA, USA: MIT Press.

Picard, R.W. (1997). *Affective computing*. Cambridge, MA, USA: MIT Press.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellermans (Eds.), *Emotion theory, research, and experience: Vol. 1, Theories of emotion* (pp. 3–33). New York, NY, USA: Academic Press.

Reeves, B., & Nass, C. (1996). *The media equation. How people treat computers, television, and new media like real people and places*. New York, NY, USA: Cambridge University Press.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178.

Scheirer, J., Fernandez, R., Klein, J., & Picard, R.W. (2002). Frustrating the user on purpose: A step toward building an affective computer. *Interacting with Computers*, *14*, 93–118.

Schneider, F., Gur, R.C., Gur, R.E., & Muenz, L.R. (1994). Standardized mood induction with happy and sad facial expressions. *Psychiatry Research*, *51*, 19–31.

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997): *Der Mehrdimensionale Befindlichkeitsfragebogen* (*MDBF*) [Multi-dimensional mood questionnaire]. Göttingen, Germany: Hogrefe.

Velten, E. (1968). A laboratory task for induction of mood states. *Behaviour Research and Therapy*, *6*, 473–482.

Wierzbicka, A. (1992). Talking about emotions: Semantics, culture, and cognition. *Cognition and Emotion*, *6*, 285–319.