

Physiological Signals and Their Use in Augmenting Emotion Recognition for Human–Machine Interaction

R. Benjamin Knapp, Jonghwa Kim, and Elisabeth André

Abstract In this chapter we introduce the concept of using physiological signals as an indicator of emotional state. We review the ambulatory techniques for physiological measurement of the autonomic and central nervous system as they might be used in human–machine interaction. A brief history of using human physiology in HCI leads to a discussion of the state of the art of multimodal pattern recognition of physiological signals. The overarching question of whether results obtained in a laboratory can be applied to ecological HCI remains unanswered.

1 Introduction

The three indicators of emotional state currently used in emotion research have been “evaluative reports, overt actions, and physiological responses” (Bradley and Lang, 2007). Evaluative reports (e.g., questionnaires and verbal anecdotes) and overt actions (e.g., facial gestures, vocal utterances, and body position) have been discussed in other chapters within this handbook. Informed by many diverse fields of science including neuro-physiology, psychophysiology, and human–computer interaction (HCI), we will explore this chapter incorporating the use of physiological signals in emotion recognition and the use of this information as part of human–machine interaction.

Emotions are neuro-physiological processes (e.g., Cacioppo and Gardner, 1999), and attempting to decipher emotional state in the absence of direct measures of physiological changes is indeed ignoring a wealth of relevant and sometimes vital information. In spite of this, the focus of much of the recent research on using emotion as a component of HCI has been on facial, gestural, and speech recognition. (In the major conference on affective computing in 2007, papers using physiological

R.B. Knapp
School of Music and Sonic Arts, Queen’s University Belfast, Belfast, Northern Ireland, UK
e-mail: B.Knapp@qub.ac.uk

measures to detect affect were outnumbered – in a ratio of about 4:1 – by papers trying to detect affect using facial expression, body gestures, and speech analysis.) One argument for this research bias is based on the incorrect assumption that humans cannot detect physiological changes unless they are revealed by overt actions. However, as has been stated many times, “it is a mistake to think of physiology as something that people do not naturally recognize. A stranger shaking your hand can feel its clamminess (related to skin conductivity); a friend leaning next to you may sense your heart pounding; students can hear changes in a professor’s respiration that give clues to stress; ultimately, it is muscle tension in the face that gives rise to facial expressions” (Picard et al., 2001). It should also be noted that revealing changes that are not detectable by human contact indeed might be the most important contribution of physiological measures to emotion recognition (Cowie et al., 2001; Hudlicka, 2003).

In addition to arguments based on the neuro-physiological underpinnings of emotion and the recognition of emotion based on physiological state in everyday interaction, there is yet another strong reason for studying physiological signals when analyzing emotional state. As has been pointed out in publications too numerous to cite, each and every single modality used for emotion recognition has its limitations. Four examples among the different measurement modalities (including physiological measurement) are as follows:

- (1) Self-reporting of emotional state can be erroneous or incomplete due to various psychological factors.
- (2) Speech and facial patterns are strongly linked to social interaction, and so they tend to provide very little information in the case of a person who is not interacting socially.
- (3) Facial or body gestures can be muted or altered in an attempt to conceal emotional state.
- (4) Physiological response correlates of emotion (even within one individual) are highly context dependent (see Bradley and Lang, 2007 for an overview).

Combining multiple modalities, where any single modality can be “noisy” or ambiguous (including physiological signals), has the potential to show significant improvement in recognizing patterns that may provide clues to emotional state (e.g., Kim and André, 2006).

1.1 A Note on the Name “Affective Computing” and Physiological Measurement

Physiological measurement has an interesting connection with discipline names. Many psychologists use the term “affect” to describe a global phenomenon with rather direct links to physiology. For instance, Russell’s group uses the term “core affect” to describe a primitive sense of positiveness or negativeness; energy or sloth that they hypothesize underpins emotion and other complex phenomena. They also

propose that it has links to certain relatively gross kinds of brain activity (Russell, 2003, see also chapter “Editorial: ‘Theories and Models’ of Emotion”). It is natural to think that what body sensors measure may be linked to affect in that sense rather than the subtler phenomena that philosophers call emotions. Hence some researchers who focus on physiological measurement as part of human–computer interaction also find it natural to call their work “affective computing.” However, as has been mentioned and will be discussed further in this chapter, there are indeed unequivocal correlations between physiological changes and specific emotional states and using the term “affective” might be inappropriately limiting.

2 Definitions of “HCI-Appropriate,” Emotion-Related Physiological Signals

Before discussing the history and usage of human physiology as part of a repertoire of emotion measurement, it would make sense to define the physiological signals to be used and briefly explain how they are currently measured. There are a great many textbooks and articles that define these signals and explore their usage in fields ranging from medicine and psychology to bioengineering and instrumentation. Two classic books are *The Handbook of Psychophysiology* (Cacioppo et al., 2007), which defines each physiological signal and summarizes its history and usage within the broad field of psychophysiology, and *Medical Instrumentation, Application and Design* (Webster, 1998), a classic engineering handbook that reviews physiological signals in the context of the medical instrumentation needed to capture each signal.

Among the many methods for measuring changes in human physiology, this section will focus on briefly defining those physiological parameters or signals which are currently being used in the fields of psychophysiology and emotion *and* are measurable in a relatively unobtrusive way (and thus could be envisaged to be part of a human–machine interface). It should be emphasized that many physiological signals have only recently met these criteria due to advances in measurement technologies. Thus, one can expect that the introduction of new technologies will serve to expand this list over the coming years (or months!). A good review (although not entirely complete) of so-called ambulatory monitoring systems, systems that can be used in mobile environments and are relatively unobtrusive, can be found in Ebner-Priemer and Kubiak (2007). In defining these key physiological signals, we will categorize them into physiological signals that originate from the autonomic and somatic components of the peripheral nervous system and physiological signals that originate from the central nervous system (see Fig. 1 for a taxonomy of the nervous system).

2.1 Physiological Data Acquisition

In collecting data from physiological sensors, the signals must ultimately be captured by a data acquisition system connected to a digital system ranging from an

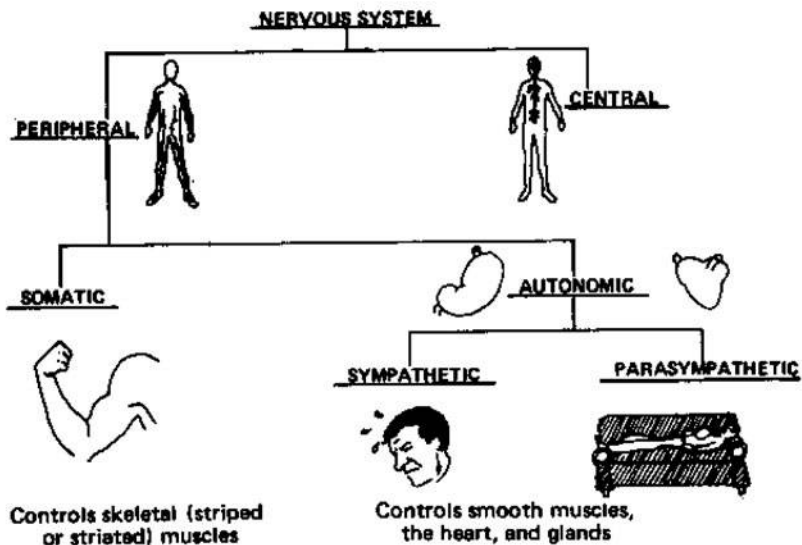


Fig. 1 Taxonomy of the nervous system (from online.sfsu.edu/~psych200/unit5/u5m5.gif)

embedded processor to a super-computer. The choice of available data acquisition systems is vast. Some of the important factors in narrowing this range of choices include the following:

- *Signal isolation*: In order to reduce the chance of electrical shock, *the data acquisition system must electrically isolate the individual from “wall” or “mains” power as well as from the computer itself.* This can be achieved by using a battery-powered wireless data acquisition system or by using a system specifically designed for physiological data acquisition that already incorporates optical or magnetic isolation.
- *Obtrusiveness*: There is a considerable trade-off in the number and type of physiological signals chosen and the obtrusiveness of the measurement. At the limits, either not enough data is being measured or the individual cannot participate in the interaction in any way that is not interfered with by the measurement equipment. While this is one of the most difficult choices in using physiological measures for emotion measurement, there is little to no literature on the impact of the ecology of the measurement on the emotional estimation “accuracy.”

It should be pointed out that while so-called wireless data acquisition systems are usually superior to wired systems in terms of their obtrusiveness, they might indeed include many wires connecting the physiological sensors to the wireless transmitter. There are very few commercially available systems that combine the sensor with the transmitter to eliminate all wires and are most commonly found in the consumer sports arena from companies such as Polar, Suunto, or Nike.

- *Sampling rate and resolution:* These will be based on the bandwidth and dynamic range of the physiological signal acquired. The range of sampling rates is considerably reduced for wireless data acquisition systems.
- *Synchronization of multiple data streams:* In many cases, more than one channel of physiological data is analyzed simultaneously (or combined with video or audio streams). The data acquisition system – both hardware and software – must have the capability of synchronizing multiple streams of data with multiple sampling rates.
- *Shielding and differential amplification:* The amplitude of the voltage of physiological signals can be as small as a fraction of a microvolt. In order to avoid contamination from other signals, so-called noise signals, electrical shielding of the signal wires from sensor to data acquisition system is extremely important. Also, data acquisition systems using a technique known as differential amplification, amplifying only the difference between two sensor signals, should be used if possible. By amplifying only the difference between two sensor signals, any noise that is common to both sensors will be considerably reduced.

2.2 Autonomic and Somatic Nervous System

The somatic component of the peripheral nervous system is concerned with sensing information that happens outside the body and is responsible for the voluntary control of our skeletal muscles to interact with this external environment. Signals that measure this voluntary control of the muscles are measuring aspects of the somatic nervous system. The autonomic (ANS) component of the peripheral nervous system is responsible for sensing what happens within the body and regulating involuntary responses including those of the heart and smooth muscles (muscles that control such things as the constriction of the blood vessels, the respiratory tract, and the gastrointestinal tract). There are two components of the ANS, the parasympathetic and the sympathetic (see Figs. 2 and 3). The parasympathetic component is responsible for slowing the heart rate and relaxing the smooth muscles. The sympathetic component of the ANS is responsible for the opposite, i.e., raising the heart rate and constricting the blood vessels which causes, among other effects, an increase in blood pressure. It is also responsible for changes in skin conductivity. The sympathetic response is slower and longer lasting than the parasympathetic response and is associated with the so-called flight or fight reaction. Physiological signals that measure the involuntary responses of the peripheral nervous system are measuring aspects of the autonomic nervous system.

2.2.1 Electrodermal System

Measurement of the electrodermal activity or EDA is one of the most frequently used techniques to capture the affective state of users, especially for exploring

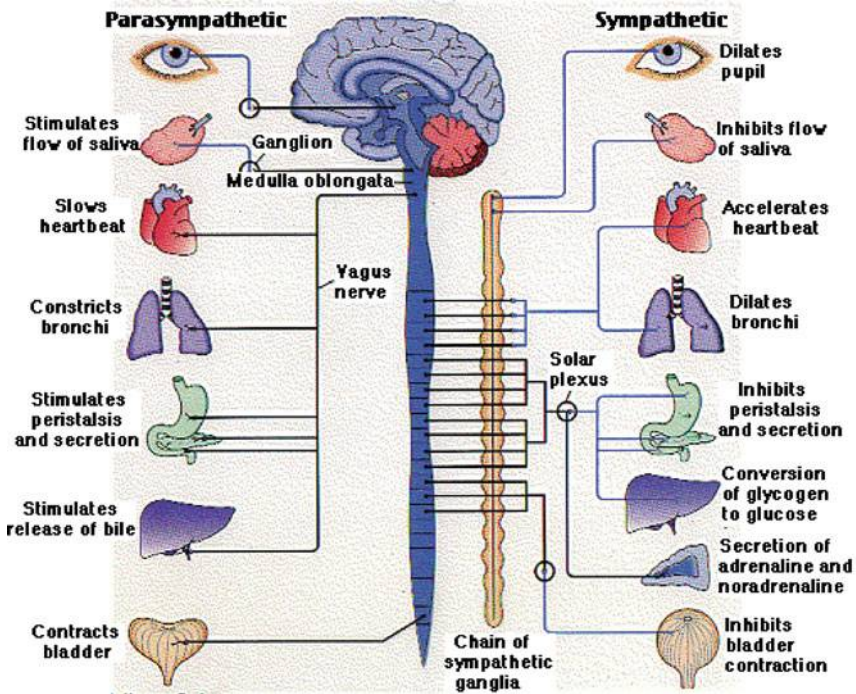


Fig. 2 Autonomic nervous system (from <http://www.wickersham.us/anne/images/autonomic.gif>)

attention and varying arousal in emotion. It is controlled by the sympathetic nervous system. EDA sensors (electrodes) measure the ability of the skin to conduct electricity. A small fixed voltage is applied to the skin through the electrodes and the skin's current conduction or resistance is measured (this is preferred over applying a current and measuring the voltage produced). The value of this conductivity is usually in the range of 2–20 μS (500 $\text{k}\Omega$ –50 $\text{k}\Omega$). The skin conductivity consists of two separate components. There is a slow-moving tonic component or skin conductance level (SCL) that indicates a general activity of the perspiratory glands from temperature or other influences and a faster phasic component or skin conductance response (SCR – also known as galvanic skin response or GSR) that is influenced by emotions and the level of arousal. For example, when a subject is startled or experiences anxiety, there will be an increase in the skin conductance due to increased quantity of sweat in the sweat ducts of the glands.

EDA is most significant on the palm of the hands and the bottom of the feet. The two most common measurement techniques are thus

1. to place an electrode on each of two fingers (usually the thumb and the index finger) and
2. to place two electrodes across the palm.

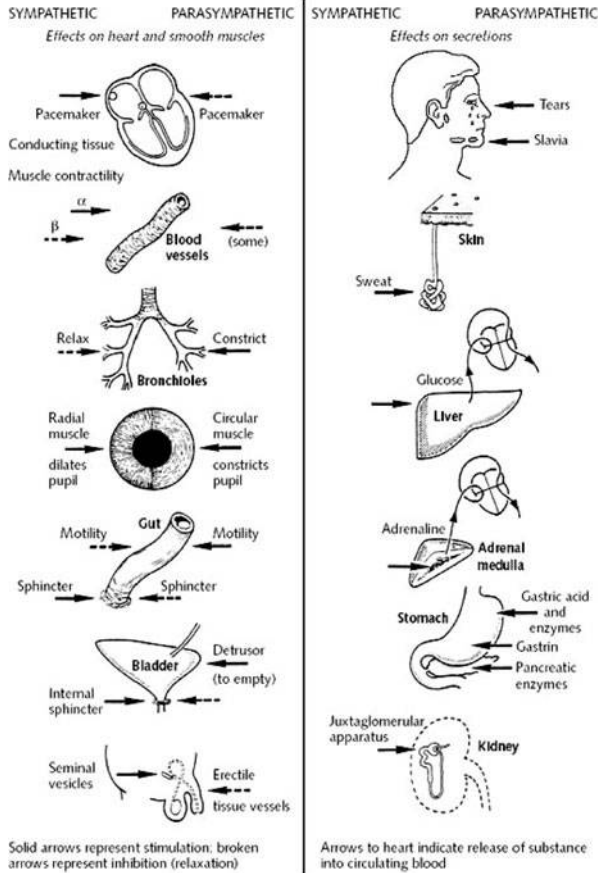


Fig. 3 Autonomic nervous system (after Jennett (1989) Human Physiology)

For the most accurate recordings, the electrodes consist of silver–silver chloride cup electrodes. However, in using the measurement of EDA as part of an HCI system, the EDA is commonly measured using standard conductive metal plates (Fig. 4).

Changes in physical activity, environmental conditions such as temperature and humidity, movement of the electrodes on the skin, and changes in pressure on the electrodes all serve to confound the measurement of EDA and must be mitigated. Physical activity and ambient conditions primarily affect the SCL and should be measured using other modalities (motion sensors, temperature and humidity sensors, etc.) in an attempt to limit the artifact.

In addition to the amplitude of the SCR and SCL, temporal parameters of the SCR such as latency, rise and decay time are all important features of the EDA used in determining attention and emotional state.

Fig. 4 Mouse used to measure EDA (from Thought Technology)



2.2.2 Cardiovascular System

Heart Rate and Heart Rate Variability

Another important physiological correlate of emotion is the frequency or the period of contraction of the heart muscle. As shown in Fig. 3, the heart rate (HR) is controlled by both the parasympathetic response (decreasing HR) and sympathetic response (increasing HR). The heart rate or period can be derived from many measurement techniques as will be discussed below. The higher frequency changes (0.15–0.4 Hz) are heavily influenced by breathing [respiratory sinus arrhythmia (RSA)], especially with younger and more physically fit individuals. The lower frequency changes (0.05–0.15 Hz) are not influenced by the RSA and can reveal other aspects of the ANS.

Electrocardiography (ECG): Electrocardiography measures electrical changes associated with the muscular contraction of the heart. More specifically, the ECG results from the sino-atrial (SA) node and atrio-ventricular (AV) node of the heart electrically activating the first of two small heart chambers, the atria, and then the two larger heart chambers, the ventricles. Particularly, the contraction of the ventricles produces the specific waveform known as the QRS complex as shown in Fig. 5. The heart rate is most commonly derived from the ECG by measuring the time between R components of the QRS complex – the so-called R–R interval.

The ECG is measured at the body surface across the axis of the heart. Electrodes are placed on the skin; they transduce the electric field caused by the previously mentioned electrical activity of the heart to electron flow in the measurement lead. This signal is then high-pass filtered to remove long-term offset and amplified by a factor of around 1,000. In medicine, the ECG is most commonly measured using

Fig. 5 ECG wave

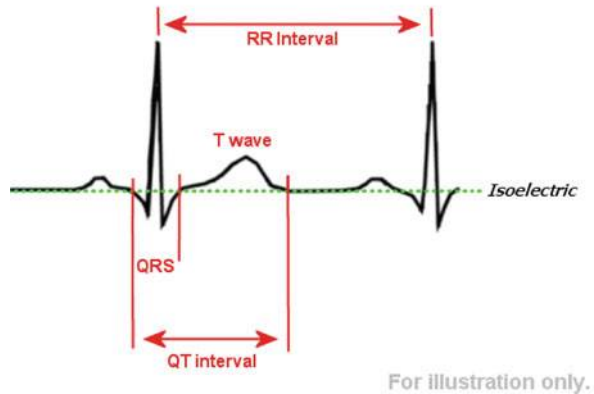


Fig. 6 Watch with wireless ECG-derived heart rate monitor (www.suunto.com)



a standard 12-lead (electrode) configuration; however, when only the heart rate is being measured, this can be reduced as low as a 2-lead configuration as shown in Fig. 6. Indeed, as is done with exercise equipment, heart rate measurements can even be recorded simply by measuring the voltage potential from hand to hand (although not nearly as accurate as chest leads).

In addition to measuring heart rate, one other aspect of the ECG wave that is important to emotion research is the amplitude of the T wave shown in Fig. 5. This is because the T wave amplitude has been found to be an indicator of activity of the sympathetic nervous system.

Photoplethysmography and Blood Volume Pulse (BVP): An alternative to measuring heart rate directly from the electrical activity of the heart is to measure the pulsation of blood flow through the vascular system. The most common technique for achieving this is to shine light from an infra-red LED into the skin and measure either the amount of light transmission through the skin or the amount of light reflection from the skin (or both). As the heart beats, the perfusion of the blood vessels

Fig. 7 Photoplethysmography (from <http://www.medis-de.com>)



underneath the light source ebbs and flows and thus the absorption characteristics of the light pulses with the heart beat. As shown in Fig. 7, the finger tip is the most common location for measurement of the BVP, although it can also be measured in other places on the extremities of the arm, feet, or even earlobe. Photoplethysmography becomes quite inaccurate due to even minor motion of the body and is often coupled with the use of accelerometers to detect motion and to attempt to compensate (e.g., Morris et al., 2008).

It is important to note that two other physiological parameters associated with changes in emotional state can be measured with photoplethysmography: peripheral blood perfusion (e.g., Kunzman and Gruhn 2005) and blood oxygen saturation (SpO₂) levels (e.g., Karekla et al., 2004). The effect of the change in magnitude of the BVP because of changes in peripheral blood perfusion can also cause the measurement of HR and HRV to be somewhat less accurate due to missed beats.

Impedance plethysmography works similar to photoplethysmography except that instead of measuring the change in blood perfusion by measuring the change in light reflectance/transmission, impedance plethysmography measures the change in the skin's capability to conduct electricity. This is most often measured by applying a small AC current across the chest and measuring the change in voltage. This technique has enabled the creation of clothing that can measure HR and HRV.

In addition to modulation of the heart rate, emotional changes can influence the contractility of the blood vessels. This change can affect blood pressure and perfusion as well as skin temperature.

As mentioned previously, peripheral perfusion can be quite ergonomically measured by photoplethysmography. Just as simply, skin temperature can be measured by placement of a thermistor on the skin surface (although this is confounded considerably by ambient temperature). Ecological measurement of blood pressure,

Fig. 8 Portapress ambulatory blood pressure monitor (www.finepress.com)



however, has still not been achieved. This is because the most common (and most accurate) non-invasive techniques for measuring blood pressure require that a cuff is placed on the arm and is inflated to cut off arterial blood flow. The acoustical changes that appear as this flow is returning to normal [first the appearance of a heart rhythm (k) sound and then the disappearance of the k sound] as the cuff is deflated indicate the systolic and diastolic blood pressure, respectively. This inflation and deflation of the cuff make it difficult to measure continuous blood pressure changes. Although expensive and large, one of the few ambulatory devices available is the Portapress monitor shown in Fig. 8.

It is argued that pulse wave velocity, the measurement of the time it takes for a pulse to move down the arm, is proportional to changes in blood pressure. See Harata et al. (2006) for a review. The pulse arrival time can be measured at two locations using photoplethysmography and then the pulse transit time (and thus pulse wave velocity) can be calculated. An example of a current attempt at using the ecological measurement of pulse wave velocity for emotion tracking can be seen in Fig. 9. There is relatively general agreement that this measurement cannot be used for measuring the absolute value of blood pressure due to the affects of the contractility of the arm's vascular system.

2.2.3 Respiratory System

The respiratory system is one of several systems of the body that are under both autonomic and voluntary (somatic nervous system) control. Thus respiration rate and depth as an indicator of emotional state must be used with caution and always viewed in context. In controlled environments, it has been found that variation of respiration rate generally decreases with relaxation. Startle events and tense situations may result in momentary respiration cessation and negative emotions generally cause irregularities in the respiration pattern.

Fig. 9 Pulse wave velocity measurement of blood pressure (www.exmocare.com)



The respiration signal (breathing rate and intensity) is commonly acquired by using a strain gauge or a piezo sensor embedded in an elastic band worn around the chest. The sensor measures the expansion and contraction of the band, which is proportional to the respiration rate. Two bands are often used, one placed on the upper chest and another around the lower abdomen, to measure the depth of inspiration and exhalation.

Another technique for measuring respiration is impedance plethysmography – using the fact that the impedance of the chest cavity varies with respiration. This is highly advantageous because both heart rate and respiration can be determined from one chest strap or article of clothing.

One other ecological technique of measuring respiration rate is to analyze the respiratory sinus arrhythmia (RSA) from the ECG waveform (as discussed above). As with impedance plethysmography, both heart rate and respiration rate can be determined using a single chest band. Although not nearly as accurate as piezo or strain-gauge bands (due to baseline variation and noise in the ECG signal), with improving signal processing techniques, the so-called ECG-derived respiration (EDR) rate has become an increasingly common method of determining respiration rate (e.g., Yeon et al., 2007).

Blood oxygen saturation levels (SpO_2) as well as the pCO_2 levels can be used as another parameter in quantifying emotional response. The cause of changes in the oxygenation of the blood is multi-factorial. That is, it is not simply a function of respiration or cardiovascular activity or any one other system, but a combination of many systems. Photoplethysmography, as discussed previously, is the most ecological technique for measuring blood oxygenation.

2.2.4 Visible (Overt) Effects of Autonomic Physiological Changes: Tears, Eye Blinks, Pupil Dilation, and “Goose Bumps”

There are several overt changes that can directly indicate activity of the autonomic nervous system. While this chapter focuses on physiological changes, since these overt changes can be a direct (uncognitively mediated) function of the underlying physiology, they are worth mentioning. Overt properties of the eye that fall into this category include tears (tear volume and ocular hydration), eye blinks, and pupil dilation, which can be measured with various visual recognition systems that can co-exist with HCI. All have been associated with various changes in emotional state. Visual recognition systems have also been used in attempting to quantify “goose bumps” or “goose flesh” or the pilomotor reflex. The anecdotal reporting of the pilomotor reflex is commonly mentioned in descriptions of emotional response.

2.2.5 Muscle Activity

As has been discussed in other chapters of this book, it is well known that overt facial gestures are a well-studied indicator of emotional state. However, before visible movement occurs on the face, activation of the underlying musculature must occur. Indeed, there are many circumstances where there are measureable changes in the activation of the facial muscles and no visible facial gesture. This can be due to “rapid, suppressed, or aborted” expressions (Cacioppo et al., 1992) or due to the actual attachment of the muscular structure to the skin. Thus, the ability to measure muscular activation in the face is another point in which physiological measurement can supplement measurement of overt changes. Changes in muscular tension in other areas of the body such as the arms may also indicate an overall level of stress or, as with the face, indicate the presence of overt gestures that cannot be viewed with visual observation.

Measurement of muscle tension is commonly achieved using surface electromyography (sEMG or just EMG). Surface electromyography measures muscle activity by detecting the electrical potential that occurs on the skin when a muscle is flexed. This electrical potential is created by motor neurons depolarizing or “firing” causing the muscle fibers to contract. The rate (frequency) of depolarization is proportional to the amount of contraction (until the individual motor neuron begins to saturate). At the same time, more and larger motor neurons are recruited and begin to fire simultaneously. Thus, as shown in Fig. 10, an increase in muscle contraction is seen as an increase in amplitude of the EMG as well as a modulation of the frequency spectrum. The structure of motor neurons and the muscle fibers they innervate is called a motor unit and the potential measured by the surface EMG is commonly referred to as a motor unit action potential or MUAP. As with measurement of the ECG, EMG measurement involves the use of electrodes which are placed on the skin surface (Fig. 11).

Fig. 10 EMG waveform. Note the four increasingly large contraction events indicated by an increase in amplitude and spectral complexity of the EMG

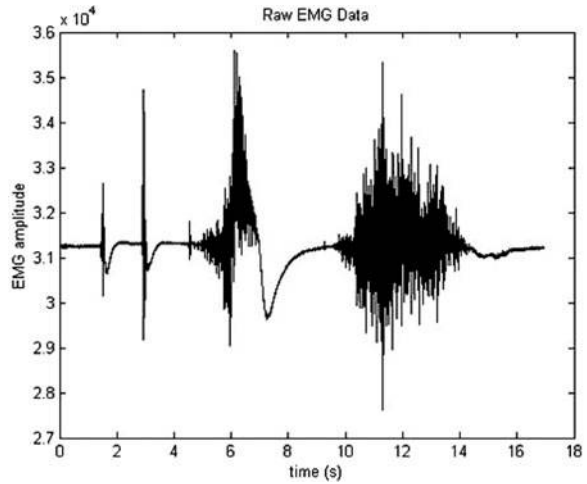


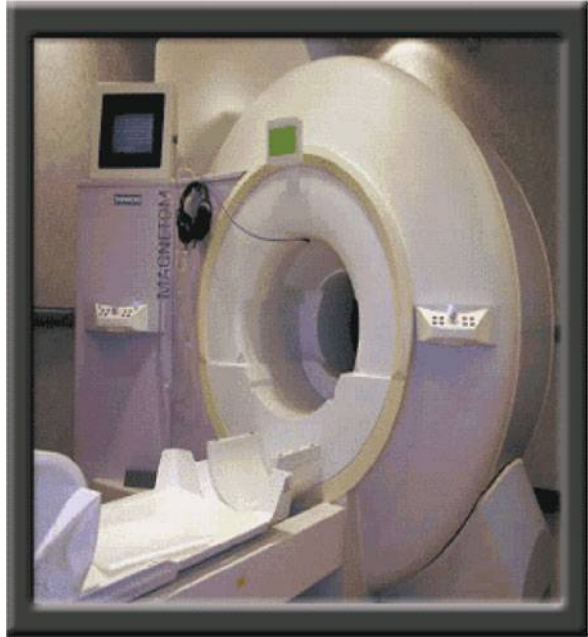
Fig. 11 An ecological EMG band for wearing on arm or leg (www.infusionsystems.com)



2.3 Central Nervous System

The central nervous system (CNS) is composed of the brain and the spinal cord and is responsible for processing information and controlling the activity of the peripheral nervous system. Over the past decade, functional imaging of the brain (imaging of the dynamic function rather than the static condition of the brain) during emotion-inducing activities has been yielding an increasing body of knowledge of how and where emotions are processed in the brain (Bradley and Lang, 2007). The capability to compare and correlate CNS activity as measured with functional imaging with the activities of the ANS as measured using the techniques discussed previously has the potential to reveal mapping functions between

Fig. 12 fMRI machine



emotional stimulation and physiological response. Unfortunately, because of the enormous size, cost, and operational/environmental constraints, the most important functional imaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) are unlikely to be used as part of a human-machine interaction scenario for several years to come (see Fig. 12).

Electroencephalography (EEG) measures the electrical activity of the brain as it appears at the surface of the skull. As the millions of neurons within the brain “fire,” the electric field generated by the electrochemical process can be measured using surface recording electrodes that function in a similar manner as the electrodes used for EMG and ECG. However, unlike ECG signals which are typically measured using up to 12 electrodes, EEG recordings can use anywhere from 3 to 256 (and even more!) electrodes (see Fig. 13). The choice of electrode quantity will depend on the desired number of locations on the head to be measured and the desired spatial resolution within any given location. The measurement of the EEG is also considerably more difficult than either the EMG or the ECG because many of the important features of the EEG signal are more than 1,000 times smaller than the EMG or ECG signal. Measurement of signals below 1 μV is common and can require that more attention is paid to cleaning the electrode site and applying electrolyte. Additionally, considerable signal averaging (comb filtering) and spectral and spatial filtering are required to remove non-EEG signals, “noise,” from the EEG signal. These “noise”



Fig. 13 Three examples of the range of electrode quantity found in EEG interfaces. From *left to right*: the emotivEpoc (www.emotiv.com), a standard 10–20 EEG array (www.biopac.com), and a 256 lead array (www.biosemi.com)

signals include EMG and movements of the eyes causing baseline potential shifts (EOG).

There are three general areas of research into the use of EEG measurement as an indicator of emotion:

1. *Spatial location and distribution of EEG signals*: While emotions are not located in any one particular location of the brain (Phan et al., 2004), amplitude and temporal asymmetries in EEG response patterns can yield insight into emotional state (e.g., Costa et al., 2006).
2. *The temporal response patterns of the EEG to stimuli – evoked response potentials (ERPs)*: The pattern of the time, location, and amplitude of the EEG response to auditory, visual, or emotional imagery can yield clues to emotional state and processing. While the ERP is studied at locations across the skull, the time response is limited to a narrow range from 100 mS (fast) to 1,000 mS (slow). This is one of the largest areas of research in the physiological correlates of emotional state – especially in the area of the correlation of ERPs and emotionally evocative visual stimuli (e.g., Holmes et al., 2003; Schupp et al., 2003).
3. *The frequency structure of the EEG signal – “brain waves”*: Synchronization of neural activity underneath the recording electrodes gives rise to oscillatory behavior in a collection of frequency bands. These band ranges include the following:
 - Delta band (1–4 Hz)
 - Theta band (4–8 Hz)
 - Alpha band (8–13 Hz)
 - Beta band (13–30 Hz)
 - Gamma band (36–44 Hz)

Correlations between the presence, timing, and location of these frequencies on the skull have been found to be related to several aspects of attention, vigilance, and emotional state (e.g., Sebastiani et al., 2003).

3 A Very Brief History of Physiological Measurement and Emotion

The history of using physiological signals as a means for recognizing emotion is a relatively short one. The use of this information as a means for augmenting human-machine interaction has only just begun.

3.1 Measuring the Correlations Between Physiology and Emotion

History is worth knowing in the field of emotion and physiology, because it is an area where popular conceptions of science are deeply attached to old ideas. William James (1890) captured the public imagination when he claimed (in 1884) that the essence of emotion was awareness of visceral changes that occur in response to extreme situations (such as meeting a bear in the woods) (Cannon, 1927). If that were so, then measuring the visceral changes directly would allow artificial systems to detect a person's emotions as well as the person him- or herself – or possibly better. In reality, even contemporary experts who are sympathetic to James accept that his idea captures only part of the truth. Injuries that prevent visceral feedback do not nullify emotion; visceral states are not as sharply distinguished as emotions are; experienced emotion can be changed by manipulating cognitive state, but not visceral states; and so on.

Not long after James released his idea, prominent figures (including Jung) took up the idea that skin conductivity could reveal otherwise hidden psychological events. Lie detection quickly became a high-profile application for it and related techniques. The idea captured the public imagination to the extent that employers were held liable for failing to use lie detection before hiring employees who went on to commit crimes. However, by 1959, it was becoming clear that standard forms of lie detection had questionable scientific validity (Lykken, 1998). It was not the physiological measures that were the focus of the uncertainty, but the methodology of the protocol and specifically the questions that were being asked. In studying emotion and, most importantly, physiology and emotion, it was becoming clear that context and induction techniques were critical.

Over the ensuing decades, psychophysiologicalists have built up an enormous body of evidence on relationships between physiological measures and emotion. Two good summaries are in Cacioppo et al. (2000) and Bradley and Lang (2007). This research has shown that if other variables are meticulously controlled, there are associations between emotional states and physiological variables. The phrase “meticulous control” is key. Physiological signals are subject to multiple influences. They are affected not only by emotion but also by almost any kind of effort, mental or physical. Traditional experiments dealt with these problems by creating situations that prevented irrelevant variables from intruding. Even with meticulous control, recent literature surveys such as found in Kreibig et al. (2007) show that correlations between physiological changes and changes in emotional state are not always consistent.

3.2 *Physiology and Human–Machine Interaction*

The question then becomes, “If, to find correlates of emotional states, physiological signals must be measured in meticulously controlled environments, how then can they be used as part of an ecological human–computer interaction paradigm?” To see how researchers have answered this, we must examine what has occurred in three separate fields.

3.2.1 *Physiological Control of Music*

The first endeavors to use physiological signals to control machines occurred well before mice and GUIs and even PCs existed. One of the most interesting examples, as shown in Fig. 14, was Alvin Lucier’s piece in 1965 entitled *Music for Solo Performer*. In this performance he sonified the alpha activity of his EEG. While the idea was to listen to changes of cognitive state, i.e., there was no direct intention of quantifying emotional state, the synchronization of alpha activity is proportional to relaxation and so there was probably more than a little self-induction of a low activation emotion.

In 1978, Dick Raaijmaker used EMG, EDA, ECG, and acoustic measurement of respiration to sonify the level of stress of an individual dismounting a bicycle over the course of 30 min. While a large component of the changes in physiological state was caused by physical exertion, changes in emotional state throughout the course of the piece clearly influenced the sound.

The increasing use of physiological state (and by possible unintentional consequence, emotional state) as a tool for artistic expression led to the development by Knapp and Lusted in 1987 of the BioMuse (Knapp and Lusted, 1990). This was one of the first commercially available systems which enabled musical performers to use physiological state (in this case EMG, ECG, and EEG) to control consumer electronic musical instruments and introduced the capability to control the newly ubiquitous PC.



Fig. 14 From Alvin Lucier’s *Music for Solo Performer* – controlling sound with EEG

3.2.2 Physiological Control of Computers (Without Emotional Assessment)

At this time, research on using physiological signals to augment human–computer interaction became established with work on EMG (Putnam and Knapp, 1993), EOG (LaCourse and Hludik, 1990), and the new field of brain–computer interfaces (BCIs) (Wolpaw et al., 1998) using EEG. Much of this work was targeted at improving interaction for those with disabilities (Lusted and Knapp, 1996). For example, ERPs from the EEG were combined with EOG to augment mouse control (Patmore and Knapp, 1995).

Research on physiologically augmented interfaces has continued to expand into many arenas including assistive living and computer gaming. From these investigations, new, more ecological interfaces are being created (Knapp and Lusted, 2005) from which forms the foundation for applying physiological measurement of emotion to HCI.

3.2.3 Physiological Control of Computers with Emotional State Assessment

Combining the research on physiological interfaces with the ongoing research on psychophysiology and emotion, Picard and her colleagues at the MIT Media Lab began developing what she termed “affective interfaces” and “affective clothing” (Picard and Healey, 1997). Interfaces ranging from jewelry to gloves (see Fig. 15) were being used to investigate whether correlates of emotional state could be found with ecological interfaces.

Many new physiologically based systems are currently being created in research centers and from commercial enterprises. Some of these can be seen in Figs. 4, 11,



Fig. 15 The galvactuator – mapping EDA to light

and 13. In the field of music, some researchers are beginning to look at using emotion as part of conducting (Nakra, 2000) and performing (Knapp and Cook, 2005). It still remains to be seen, however, whether measurements using any of these new interfaces, even in conjunction with other measures such as facial recognition or speech recognition and operating in constrained environments, can accurately assess emotional state. This is the ultimate question of the current research, “Can an automatic recognition system be developed that can process physiological signals and other indicators of emotional state and come to any reasonably consistent result?”

4 The Present: Current Studies on the Automatic Recognition of Emotional States Using Physiological Signals

Current work on pattern recognition of physiological state for HCI is now being investigated by many research groups. The results are beginning to show that again, in constrained environments, it might be possible to use machine learning algorithms such as Bayesian classifiers, HMM, neural networks, and fuzzy systems to achieve statistically significant identification.

4.1 Emotion Recognition Using Only Physiological Signals

In 2001, using only one subject, Picard showed that certain affective states may be recognized by using physiological measures including heart rate, skin conductivity, temperature, muscle activity, and respiration velocity (Picard et al., 2001). Eight emotions deliberately elicited from a subject in multiple weeks were classified with an overall accuracy of 81%. Nasoz et al. (2003) used movie clips to elicit target emotions from 29 subjects and achieved the best recognition accuracy (83%) by applying the Marquardt backpropagation algorithm.

In other work (Haag et al., 2004), the IAPS photoset (IAPS, 1995) is used to elicit target emotions with positive and negative valence and variable arousal level from a single subject. They classified arousal intensity and valence of the emotions separately using neural network classifier and achieved recognition accuracy of 96.6 and 89.9%, respectively. More recently, an interesting user-independent emotion recognition system was reported by Kim et al. (2004). They developed a set of recording protocols using multimodal stimuli (audio, visual, and cognitive) to evoke targeted emotions (sadness, stress, anger, and surprise) from 175 children aged from 5 to 8 years. Classification ratio of 78.43% for three emotions (sadness, stress, and anger) and 61.76% for four emotions (sadness, stress, anger, and surprise) has been achieved by adopting support vector machine as pattern classifier. Particularly, analysis steps in the system are fitted to handle relative short length of input signal (segmented in 50 s) compared with the other previous works that require longer signal length of about 2–6 min. Wagner and colleagues (2005) presented an

approach to the recognition of emotions elicited by music using four-channel biosignals which were recorded while the subject was listening to music songs and reached an overall recognition accuracy of 92% for a four-class problem.

Wagner and colleagues also applied statistical methods to find out which features are significant for a specific emotion class. It turned out that joy was characterized by a high SC and EMG level, deep and slow breathing, and an increased heart rate. In contrast, anger was accompanied by a flat and fast breathing. The SC and EMG level was high as well. Pleasure and sadness are well identified by a low SC and EMG signal, but pleasure has a faster heart rate. The results were, however, highly user dependent. When applying the same method to a data set collected by MIT, they could observe perceptible differences in the physiological responses of the recorded subjects. While in their own data set, positive emotions were characterized by a low SC level, the MIT data set showed a high level of SC. Nevertheless, it turned out that a high SC and EMG level was also a good indicator in general for high arousal. They could also well correlate a higher breathing rate with the emotions in negative valence.

Physiological data sets used in most of the works are obtained by using visual elicitation methods in a lab setting where subjects intentionally express desired emotion types while looking at selected photos or watching movie clips. A recognition accuracy of over 80% on the average seems to be acceptable for realistic applications. However, it can be clearly observed that the accuracy strongly depends on the data sets (which were obtained in laboratory conditions). That is, the results were achieved for specific users in specific contexts. In view of a generally applicable recognition system for realistic online applications, it is desirable to automatically select the most significant features and tune specific classifiers to manifold data sets obtained from different natural contexts.

4.2 Relation, Dependencies, and Correlation of Physiological Signals with Other Modalities

The integration of multiple modalities for emotion recognition arouses an intuitive expectation of better recognition accuracy rate compared to unimodal analysis. On the other side, however, it requires to solve more complex analysis and classification problems. Furthermore, there are some well-known interactions which prevent any simple interpretation of physiological signals as “emotion channels.” Because respiration is closely linked to cardiac function, a deep breath can affect other measures, for example, EMG and SC measurements. Particularly, in an experiment by [Kim06], such an irregularity could be observed when the subject is talking. However, it is also possible to imagine ways of turning a multimodal analysis to advantage. For instance, the respiration wave shows particular abrupt changes corresponding to certain facial muscle activity. It is also possible that EMG sensors on the jaws or forehead could improve accuracy in computation of FAP variations in facial emotion analysis. In psychophysiology, EMG was often used to find the correlation between cognitive emotion and physiological reactions. In (Sloan, 2004),

for example, EMG has been positioned on the face (jaw) to distinguish “smile” and “frown” by measuring activity of zygomatic major and corrugator supercilii. EMG sensors could also be positioned on the body to measure muscle contraction intensity for gesture recognition.

Kim and colleagues (2005) investigated how far the robustness of an emotion recognition system can be increased by integrating both vocal and physiological cues. They evaluated several fusion methods as well as a hybrid recognition scheme and compared them with the unimodal recognition methods. The best results were obtained by feature-level fusion in combination with feature selection (78%). They did not achieve the same high gains that were achieved for video–audio analysis, which seems to indicate that speech and physiological data contain less complementary information. Furthermore, they investigated the subjects’ emotional response in a naturalistic scenario where it cannot be excluded that the subjects are inconsistent in their emotional expression. Inconsistencies are less likely to occur in scenarios where actors are asked to deliberately express emotions via speech and mimics, which explains why fusion algorithms lead to a greater increase of the recognition rate in the latter case.

Experiments by Chanel and colleagues (2006) indicate that an integrated analysis of EEG and physiological data might improve robustness of emotion assessment since some participants were more expressive in their physiological response, while for others better results were obtained with EEG data.

As noted above, there are two views on emotions: emotions as the result of bodily reactions and emotions as the result of cognitive process. To analyze emotions by a computer system, a combination of both views has been proven useful. When determining the user’s emotional state, we have evidence from causal factors, such as events during the interaction with a computer system. Apart from that, we also have evidence from the consequences of emotional states, being the physiological response. Conati et al. (2003) successfully employed Bayesian networks within a probabilistic framework to model such a bidirectional derivation of emotional states from its causes as well as its consequences in a game-like tutor system.

4.3 Emotion Recognition in an Ecological Environment

In most of the work that has been cited, data sets used were obtained in a lab setting where subjects intentionally express desired emotion types while looking at selected photos, listening to music, or watching movie clips. A recognition accuracy of over 80% on the average seems to be acceptable for realistic applications. However, it remains unknown what level of information can be derived from biosensors when emotion is part of action and interaction and multiple factors impinge on variables such as heart rate and respiration. The fact that physiological sensors are very sensitive to motion artifacts makes it hard to employ them in everyday situations.

As a first step toward a more natural scenario, Kim and colleagues (2005) used a slightly modified version of the quiz “Who wants to be a millionaire?” Questions

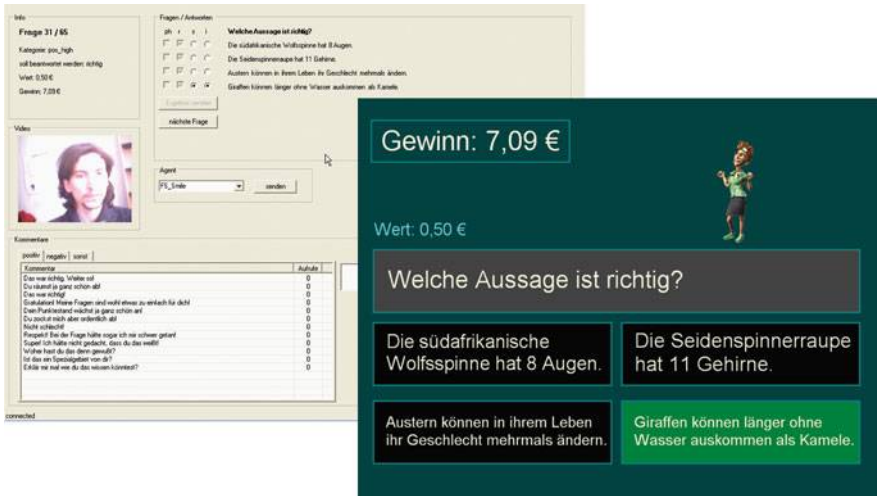


Fig. 16 Interface for the wizard (*left-hand side*) and for the subjects (*right-hand side*)

along with options for answers were presented on a graphical display whose design was inspired by the corresponding quiz shows on German TV. The subjects were equipped with a directed microphone to interact with a virtual quiz master via spoken natural language utterances and six-channel biosensors. In addition to the biosensors mentioned above, blood volume pulse (BVP) and skin temperature (TEMP) were used. The virtual quiz master was represented by a disembodied voice using the AT&T Natural Voices speech synthesizer (see right-hand side of Fig. 16).

While the users interacted with the system, their physiological and speech signals as well as the interaction with the quiz master were recorded. The quiz experiment was designed in a Wizard-of-Oz fashion where the quiz agent who presents the quiz is controlled by a human quiz master who guides the actual course of the quiz, following a working script to evoke situations that lead to a certain emotional response. The wizard was allowed to freely type utterances, but also had access to a set of macros that contained pre-defined questions or comments which made it easier for the human wizard to follow the script and to get reproducible situations (see left-hand side of Fig. 16).

In order to get a sufficient amount of data, it might be necessary to induce certain emotional states in the user. To accomplish this task, the wizard's working script in the experiment described above was divided into four situations which correspond to positions on the axes of a two-dimensional emotion model (see Fig. 17): (1) low arousal, positive valence; (2) high arousal, positive valence; (3) low arousal, negative valence; and (4) high arousal, negative valence.

First, the users were offered a set of very easy questions every user is supposed to know to achieve equal conditions for all of them. This phase was characterized by a slight increase of the score and gentle appraisal of the agent and served to induce an emotional state of positive valence and low arousal in the user. In phase 2, the user was confronted with extremely difficult questions nobody is supposed to

Fig. 17 Phases of the Kim experiment

Phase	Keyword	Arousal	Valence
1	Easy	low	Positive
2	Challenging	high	Positive
3	Boring	low	negative
4	Unfair	high	negative

know. Whatever option the user chose, the agent pretended the user's answer was correct so that the user got the feeling that s/he hits the right option just by chance. In order to evoke high arousal and positive valence, this phase led to a high gain of money. During the third phase, the experimenters tried to stress the user by a mix of solvable and difficult questions that led, however, not to a drastic loss of money. Furthermore, the agent provided boring information related to the topics addressed in the questions. Thus, the phase should lead to negative valence and low arousal. Finally, the user got frustrated by unsolvable questions. Whatever option the user chose, the agent always pretended the answer was wrong, resulting in a high loss of money. Furthermore, the experimenters included simple questions for which they offered similar sounding options. The user was supposed to choose the right option, but the experimenters made him/her believe that the speech recognizer was not working properly and deliberately selected the wrong option. This phase was intended to evoke high arousal and negative valence.

The setting allowed the experimenters to evoke spontaneous physiological reactions under controlled conditions. The emotional states were unconsciously expressed (apart from rare cases where users deliberately tried to conceal their emotions) and the display of emotions was not instructed. Despite the relatively natural scenario, we have to consider that the data were still recorded under laboratory conditions. As long as users are equipped with biosensors, we cannot avoid that they are aware that they are being monitored.

5 Summary

The debate of which emotions can be distinguished on the basis of physiological changes is far from being resolved. Empirical studies done in psychophysiology provide evidence that there is a strong relationship between physiological reactions and emotional/affective states of humans. Nevertheless, emotion assessment from physiological signals still bears a number of challenges. It is very hard to uniquely map physiological patterns onto specific emotion types. Physiological patterns may widely differ from user to user and from situation to situation. On the other hand, physiological signals have considerable advantages for emotion assessment. We can continuously gather information about the users' emotional changes while they are connected to biosensors. Moreover, physiological reactions should be more robust than visual techniques against possible artifacts of human social masking.

Combining physiological signals with other modalities has proven to be a positive next step in recognizing emotional state.

However, it is still open to question whether results that can be obtained in tightly controlled situations can be transformed into ecological techniques that work in real-life situations or whether HCI using physiological sensors to measure emotion will repeat the pattern that was seen with lie detection – high public expectations, which in the long run are not matched by results (National Research Council, 2003).

References

- Bradley MM, Lang PJ (2007) “Emotion and motivation” in handbook of psychophysiology. Cambridge University Press, Cambridge, MA, pp 582–607
- Cacioppo JT, Berntson GG, Larsen JT, Poehlmann KM, Ito TA (2000) The psychophysiology of emotion. In: Lewis M, Haviland-Jones JM (eds) Handbook of emotions. Guilford Press, New York, NY, pp 173–191
- Cacioppo JT, Bush LK, Tassinary LG (1992) Microexpressive facial actions as a function of affective stimuli: replication and extension. *Pers Soc Psychol Bull* 18:515–526
- Cacioppo JT, Gardner WL (1999) Emotion. *Annu Rev Psychol* 50:191–214
- Cacioppo JT, Tassinary LG, Bernston GG (eds) (2007) Handbook of psychophysiology, 3rd edn. Cambridge University Press, Cambridge, MA
- Cannon WB (1927) The James–Lange theory of emotions: a critical examination and an alternative theory. *Am J Psychol* 39:106–127
- Center for the Study of Emotion and Attention [CSEA-NIMH] (1995) The international affective picture system: digitized photographs. Center for Research in Psychophysiology, University of Florida, Gainesville, FL
- Chanel G, Kronegg J, Grandjean D, Pun T (2006) Emotion assessment: arousal evaluation using EEG’s and peripheral physiological signals. In: MRCS06, Istanbul, pp 530–537
- Conati C, Chabbal R, Maclaren H (2003) A study on using biometric sensors for detecting user emotions in educational games. In: 3rd workshop on affective and attitude user modeling, Pittsburgh, PA, Jun 2003
- Costaa T, Rognonib E, Galati D (2006) EEG phase synchronization during emotional response to positive and negative film stimuli. *Neurosci Lett* 406(3):159–164
- Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG (2001) Emotion recognition in human–computer interaction. *IEEE Signal Process Mag* 18:32–80
- Ebner-Priemer UW, Kubiak T (2007) Psychological and psychophysiological ambulatory monitoring, a review of hardware and software solutions. *Eur J Psychol Assess* 23(4):214–227
- Haag A, Goronzy S, Schaich P, Williams J (2004) Emotion recognition using bio-sensors: first steps towards an automatic system. In: Affective dialogue systems. Lecture Notes in Computer Science, 2004, vol 3068/2004, pp 36–48, DOI: 10.1007/978-3-540-24842-2_4
- Harata K, Kawakami M, O’Rourke M (2006) Pulse wave analysis and pulse wave velocity: a review of blood pressure interpretation 100 years after Korotkov. *Circ J* 70(10):1231–1239
- Holmes A, Vuilleumier P, Eimer M (2003) The processing of emotional facial expression is gated by spatial attention: evidence from event-related brain potentials. *Cogn Brain Res* 16: 174–184
- Hudlicka E (2003) To feel of not to feel: the role of affect in human–computer interaction. *Int J Hum Comput Stud* 59:1–32
- James W (1890) The principles of psychology. Holt, New York, NY
- Jennett S (1989) Human physiology. Churchill Livingstone, Edinburgh
- Karekla M, Forsyth JP, Kelly MM (2004) Emotional avoidance and panicogenic responding to a biological challenge procedure. *Behav Ther* 35(4):725–746

- Kim J, André E (2006) Emotion recognition using physiological and speech signal in short-term observation. In: Perception and interactive technologies, LNCS, Springer-Verlag, Berlin Heidelberg, vol 4021/2006, pp 53–64
- Kim J, André E, Rehm M, Vogt T, Wagner J (2005) Integrating information from speech and physiological signals to achieve emotional sensitivity. In: Proceedings of the 9th European conference on speech communication and technology, Lisbon, Portugal, 4–8 September 2005
- Kim KH, Bang SW, Kim SR (2004) Emotion recognition system using short-term monitoring of physiological signals. *Med Biol Eng Comput* 42:419–427
- Knapp RB, Cook PR (2005) The integral music controller: introducing a direct emotional interface to gestural control of sound synthesis. In: Proceedings of the international computer music conference (ICMC), Barcelona, Spain, 4–9 Sept 2005
- Knapp RB, Lusted HS (1990) A bioelectric controller for computer music applications. *Comput Music J* 14(1):42–47
- Knapp RB, Lusted HS (2005) Designing a biocontrol interface for commercial and consumer mobile applications: effective control within ergonomic and usability constraints. In: Proceedings of the 11th international conference on human computer interaction, Las Vegas, NV, 22–27 Jul 2005
- Kreibig SD, Wilhelm FH, Roth WT, Gross JJ (2007) Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology* 44:787–806
- Kunzmann U, Gruhn D (2005) Age differences in emotional reactivity: the sample case of sadness. *Psychol Aging* 20:47–59
- LaCourse JR, Hludik FC (1990) An eye movement communication – control system for the disabled. *IEEE Trans Biomed Eng* 37:1215–1220
- Lusted HS, Knapp RB (1996) Controlling computers with neural signals. *Sci Am* 275(4):82–87
- Lykken DT (1998) A tremor in the blood: uses and abuses of the lie detector. McGraw-Hill, New York, NY
- Morris M, Dishongh T, Guilak F (2008) Apparatus for monitoring physiological, activity, and environmental data. USPTO Application #: 2008 0154098.2006
- Nakra TM (2000) Inside the conductor's jacket: analysis, interpretation, and musical synthesis of expressive gesture. MIT Media Laboratory Perceptual Computing Section Technical report no. 518
- Nasoz F, Alvarez K, Lisetti C, Finkelstein N (2003) Emotion recognition from physiological signals for presence technologies. *Int J Cogn Technol Work Spec Issue Presence* 6:1
- National Research Council (2003) The polygraph and lie detection. Committee to review the scientific evidence on the polygraph. Division of Behavioural and Social Science and Education. The National Academic Press, Washington, DC
- Noh YS, Park SJ, Park SB, Yoon HR (2007) A novel approach to classify significant ECG data based on heart instantaneous frequency and ECG-derived respiration using conductive textiles. In: Proceedings of the 29th annual international conference of the IEEE EMBS, 22–26 Aug 2007, pp 1503–1506
- Patmore DW, Knapp RB (1995) A cursor controller using evoked potentials and EOG. In: Proceedings of the RESNA '95 annual conference, Vancouver, 9–14 Jun 1995, pp 702–704
- Phan KL, Wager TD, Taylor SF, Liberzon I (2004) Functional neuroimaging studies of human emotions. *CNS Spectr* 9:258–266
- Picard RW, Healey J (1997) Affective wearables. *Pers Technol* 1(4):231–240
- Picard R, Vyzas E, Healy J (2001) Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Trans Pattern Anal Mach Intell* 23(10):1175–1191
- Putnam WL, Knapp RB (1993) Real-time computer control using pattern recognition of the electromyogram. In: Proceedings of the IEEE international conference on biomedical engineering, San Diego, CA, 27–29 Oct 1993, pp 1236–1237
- Russell JA (2003) Core affect and the psychological construction of emotion. *Psychol Rev* 110(1):145–172

- Sebastiani L, Simoni A, Gemignani A, Ghelarducci B, Santarcangelo EL (2003) Human hypnosis: autonomic and electroencephalographic correlates of a guided multimodal cognitive–emotional imagery. *Neurosci Lett* 338(1):41–44
- Schupp HT, Junghofer M, Weike AI, Hamm AO (2003) The selective processing of briefly presented affective pictures: an ERP analysis. *Psychophysiology* 41:441–449
- Sloan DM (2004 November) Emotion regulation in action: emotional reactivity in experiential avoidance. *Behav Res Therapy* 42(11):1257–1270
- Wagner J, Kim J, André E (2005) From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification. In: ICME'05, Amsterdam, Jul 2005
- Webster JG (eds) (1998) *Medical instrumentation, application and design*, 3rd edn. Wiley, New York, NY
- Wolpaw JR, Ramoser H, McFarland DJ, Pfurtscheller G (1998) EEG-based communication: improved accuracy by response verification. *IEEE Trans Rehabil Eng* 6(3):326–333