Affective Computing

Rosalind W. Picard

The MIT Press

From The MIT Press

MITCogNet

First MIT Press paperback edition, 2000

This book was set in Stone Serif and Stone Sans by Windfall Software using ZzTEX and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Picard, Rosalind W.
Affective computing / Rosalind W. Picard.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-16170-1 (hc.: alk. paper)—978-0-262-66115-7 (pb.: alk. paper)
1. Human–computer interaction. 2. User interfaces (Computer systems). I. Title.
QA76.9.H85P53 1997
004'.01'9—dc21 97-33285
CIP

10 9 8 7

5 Affective Signals and Systems

This part of the book addresses technical issues involved in creating affective computers, specifically, how to build systems with the ability to recognize, express, and "have" emotions. This chapter and the two that follow will propose several building blocks that can be used to start filling in the framework of affective computing. I will also show how several examples from the literature can be woven into this new framework.

There are several ways to organize the building blocks we will be using. The first way is by level of representation: low-level for signals, medium-level for patterns, and high-level for concepts. These all come together in any complete system for recognizing, expressing, or having emotions. In this chapter I will illustrate some low-level signal representations for emotions, for moods, and for human physiological signals that carry affective information. The next two chapters look at mid-level and higher-level representations, respectively, although there are some aspects of all levels in each chapter.

Another way to organize the building blocks is by their uses: representing input and internal signals, recognizing patterns of signals, synthesizing expressions, generating states, analyzing situations, influencing cognition and perception, and so on. The representation issues in all these uses overlap, and many different approaches have been tried for each. This chapter will focus primarily on things that are representing well with signals: internal emotions and moods, and physiological data gathered during recognition of human emotions. This chapter suggests methods for applying tools from linear systems theory and digital signal processing to modeling of affective signals and systems. Chapter 6 focuses on recognition and expression of emotions, primarily using tools from pattern recognition and analysis. Chapter 7 is dominated by symbolic rule-based programming and connectionist models, and focuses on situation analysis, generation of emotions, and mechanisms through which emotions can influence other processes such as memory.

I have tried to keep each chapter in the rest of the book self-contained to make it easy for the reader to skip around to topics of greatest interest.

Modeling an Affective System

Consider what happens when you try to recognize somebody's emotion. First, your senses detect low-level signals—motion around their mouth and eyes, perhaps a hand gesture, a pitch change in their voice and, of course, verbal cues such as the words they are using. Signals are any detectable changes that carry information or a message. Sounds, gestures, and facial expressions are signals that are observable by natural human senses, while blood pressure, hormone levels, and neurotransmitter levels require special sensing equipment. Second, patterns of signals can be combined to provide more reliable recognition. A combination of clenched hands and raised arm movements may be an angry gesture; a particular pattern of features extracted from an electromyogram, a skin conductivity sensor, and an acoustic pitch waveform, may indicate a state of distress. This medium-level representation of patterns can often be used to make a decision about what emotion is present. At no point, however, do you directly observe the underlying emotional state. All that can be observed is a complex pattern of voluntary and involuntary signals, in physical and behavioral forms.

Not only do you perceive expressive signals from a person, but you also perceive non-expressive signals from the environment which indicate where you are, who this is, how comfortable the weather is, and so forth. These signals indicate the context, such as the fact that people are in an office setting, or that it is final exam season. The observer may notice that the weather is oppressive and reason that it could impact moods. Or, the context might be recognized as a situation where a person is expecting some exciting news. With contextual information, the observer proceeds not only to analyze low-level signals and patterns from the environment and from the person who is expressing an emotion, but also to reason in a high-level way regarding what behavior is typical of this situation, and what higher-level goals are at work.

The process of trying to recognize an emotion is usually thought to involve a transformation from signal to symbol, from low-level physical phenomena to high-level abstract concepts. However, because reasoning about the situation can modify the kinds of observations that are made, information can be considered to flow not just from the low-level inputs to the high-level concepts, but also from the high level to the low level. Suppose that in reasoning about a situation you expect that somebody will be in a bad mood; in that case, your high-level expectation can cause your low-level perception to be biased in a negative way, so that you are more likely to perceive a weak

or ambiguous expression as being negative. The recognition of emotions is therefore not merely "bottom-up," from signals to symbols, but also "top-down," in that higher level symbols can influence the way that signals are processed.

High-level reasoning and low-level signals also cooperate in the generation of emotional expression. Suppose that an actor wishes to portray a character who feels hatred. He might begin by thinking, "I want to show hatred" and then proceed to synthesize low-level signals that communicate hatred, changing his posture, behavior, voice, and face, to reflect this emotional state. The whole process has started as a symbol—a cognitive goal to show hatred—and has ended with the generation of expressive signals, so that the audience can recognize his character's hatred. The process of trying to express an emotion is usually considered to involve a transformation from symbol to signal, from high-level concepts to low-level modulation of expressions and behaviors.

However, I have left one important piece out of both of the above descriptions: the emotional state of the system which is either recognizing or expressing the emotion. In humans, this distinction is blurred because all humans have emotional states that automatically influence recognition and expression. But in computers, this distinction needs to be made explicit because a computer can be built with only a subset of these abilities. To recognize an emotion involves perception. But, as I described earlier, we know that human perception is biased by human emotion: an observer's own emotions influence both his low-level perceptual processes and his high-level cognitive processes. An observer will tend to perceive an ambiguous stimulus as being positive or negative, whichever is congruent with his mood. The emotional state of a human also influences her emotional expression. If the actress thinks, "Show hatred," then she may also begin to feel hatred. As I described earlier, it can also be the case that simply posturing her muscles to accurately communicate expressions of hatred can provide bodily feedback to cause her to actually feel the emotion she is expressing. In these cases, the emotional state, if there is one, interacts with both the recognition and expression of an emotion, with both cognitive and physical processes, and with both high-level reasoning and low-level signal processing. More commonly, a person will find that an emotional state simply arises in response to perceiving or reasoning about some events, and expression of that state occurs mostly involuntarily. Figure 5.1 summarizes these interactions.

When a computer tries to represent emotions and their expression, it may use convenient levels of abstraction—from a low-level representation of a signal such as a waveform of heart rate or a motion sequence of muscular movements, to a high-level interpretation such as the sentence "He looks

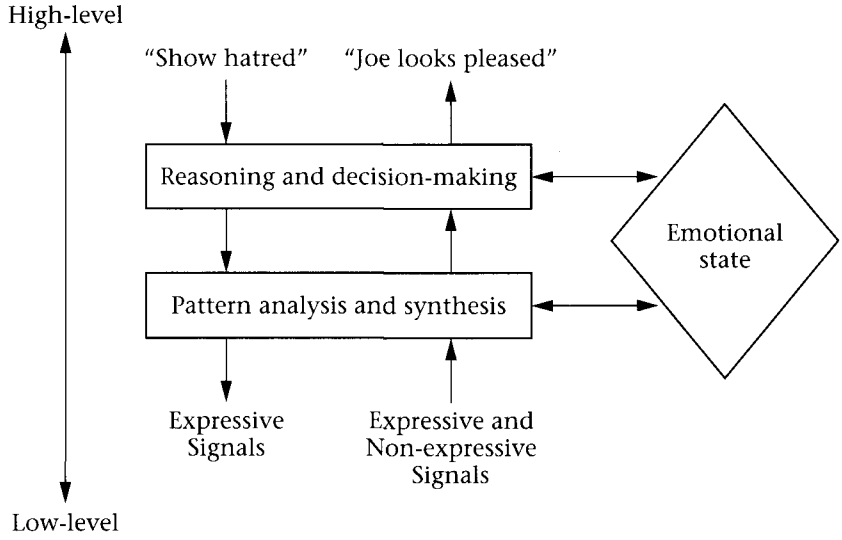

Figure 5.1
Information flows from both high to low and from low to high in a system that can express, recognize, and "have" emotions.

sad." At no point in this process does the computer have to use the same mechanisms used by humans; it might go about recognition, expression, and synthesis of emotions in an entirely different way. However, to the extent that we can understand the way humans do these things, we will have a better idea how to give these abilities to computers. Furthermore, to the extent that we imitate human mechanisms in computers, we have a better chance of debugging the computers when they behave in a peculiar way, and we stand to benefit because the ways in which they behave are likely to be close to ways that humans behave, making it easier for us to interact with them.

A Signal Representation for Emotions and Moods

Now let us consider the use of low-level signal representations for describing some of the pieces of an affective system, starting with an example:

Bruno was known to be short-tempered, but also a very caring person, with a wife and two children whom he deeply loved and desired to support. His ability to provide for his family meant a lot to him, and he tried to work hard to meet their needs. Yesterday his boss got upset over a trifle and fired Bruno from his job. Bruno felt this was unfair and responded with red-hot rage—he punched his fist through the wall, yelled at his boss, and stormed out of the office, ranting about his boss' decision. He drove aggressively across town to the local bar, his anger escalating as he thought about how upset his wife would be and how this would hurt his family. Bruno was furious.

Theorists tell us that emotions usually last for less than a minute or two, while moods can last much longer. However, a typical person might say: "It was days before he stopped being angry," as if emotions could last much longer. How do we account for these differences, and in a way that a computer can represent? How can we represent emotions, moods, influences of temperament, and other affective phenomena in a computer? In the future, physical changes indicative of emotions and moods—levels of hormones, neurotransmitters, and nervous system activity—may be measured and made into a quantitative, physically-based model. However, there is currently no reliable measure of when an emotion begins or ends, or of how intense it is. Nonetheless, let me propose a way in which we can computationally model emotions and moods, accounting for properties of their behavior, at least in a qualitative sense.

Ringing a Bell

To help picture emotions, moods, temperament, and some of their underlying properties, consider a completely different situation, one devoid of emotions but having similar behavior: the striking of a bell. When a bell is struck once, it emits a sound that is loud at first, and then decays in intensity. The intensity of the sound can be modeled by the signal shown in Fig. 5.2. It builds up quickly to a peak, becoming very loud, then gradually fades until it is too faint to be heard.[1]

An interesting thing happens if the bell is struck with exactly the same force, repeatedly, so that each strike occurs before the sound of the previous strike has subsided. In this case, the bell does not ring with the same loudness each time, but sounds louder and louder. The reason for the increase can be seen by adding the sound intensity signals over time to obtain a cumulative sound signal as in Fig. 5.2. The sum grows, despite the fact that the individual strikes are the same. Let us now look at this and several other properties of emotion.

Property of Response Decay

Several analogies can be made between the bell ringing and Bruno's situation. In one sense the stimulus of being fired was like the stimulus of striking the bell; both initiated a response—a sound from the bell, and fury from Bruno, and both responses had a fast rise time followed by a more gradual decay. If there had been no other strikes of the bell then its sound would have subsided; similarly, if Bruno had not ruminated about what the loss of his job meant, what his wife would think, what it would do to his family, and so forth, and if his body had not tensed up with the feelings of anger, then his anger might have subsided quickly. Emotions, like sounds, decay naturally over time unless they are re-stimulated.

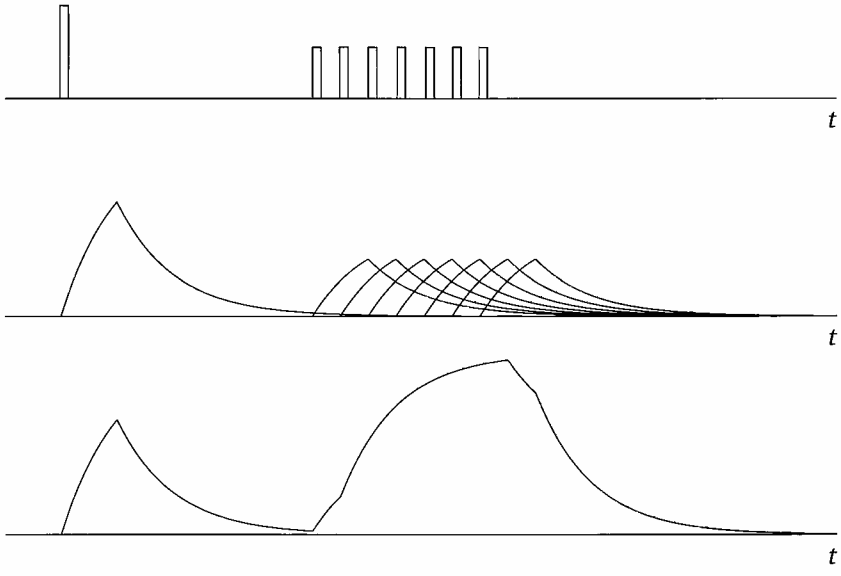

Figure 5.2
Top: The pattern and intensity of strikes applied to a bell. These represent the input signals. Middle: The bell's response to each strike. Bottom: Sum of responses, representing the net intensity of the sound the bell produces. These are the system's output signals. Notice that rapidly repeated small strikes produce a higher overall sound than does one large strike.

Although it seems natural to speak of the intensity of an emotion increasing and decaying, it is not clear which signals could be measured to give rise to a physical measure of emotion intensity, unlike the bell, where sound intensity can be measured from the acoustic waveform. Although I will give an example below to illustrate the use of signals for representing emotion intensity, it must be kept in mind that this is presently an abstract representation, not a representation of a known measurable quantity. Researchers are working on ways to measure changes associated with each emotion, including measurements of autonomic nervous system activity, and relative neurotransmitter and hormonal concentrations, and eventually we can expect quantitative measures that may fit signal representations like I propose. However, for now, the representations I describe of emotional intensity are more qualitative, intended to capture the intensity that a person might assign to their felt emotion.

Property of Repeated Strikes

For Bruno, each new thought was as another strike of the bell. Each strike gave rise to an emotion of anger; the repeated strikes had the effect of escalating the overall intensity of his anger. It did not matter if the intensity

of Bruno's thoughts was the same over time; as long as the thoughts kept coming, Bruno's anger could escalate higher than it was at the initial emotion-producing event, much like a bell rings louder with repeated strikes.

A similar analogy holds for stressful events in general: a single stressful event may be so small as not to call attention to itself. However, a lot of repeated "little" stress-producing events can lead to a greater level of stress than one major stress-producing event. Cumulative little hurts can cause greater pain than a single instantaneous painful event. This signal representation provides a convenient way to visualize how emotional intensity can accrue, even though emotions may be of short duration.

Property of Temperamental Influence

Emotional response is influenced by one's personality. Personality is believed to be largely a function of environmental influences; however, there is evidence that part of personality, one's temperament, may be largely set before birth. Fifteen percent of infants have significantly more active nervous systems than most babies, and are more easily excited or stressed (Kagan, 1994). Throughout their early lives, these highly reactive babies tend to grow into inhibited, more easily fearful children. Although children can grow out of these early extreme temperamental differences, highly reactive babies tend to become at least mildly inhibited adolescents and adults, and tend not to develop extroverted personalities.

Although it is not entirely clear how temperament exerts its influences, it is apparently an attribute of our nervous system. There is a case of conjoined twins that supports this hypothesis. These two girls share one body from the chest down, giving the appearance of one person with two heads. They are completely connected biochemically: when one takes medicine, it helps the other. It might therefore seem surprising that their friends and family describe the girls as having two different temperaments and two different personalities. However, the twins have two hearts, two stomachs, and, interestingly, two communicating but distinct nervous systems, which can account for two temperaments in the same body (Wewerka, Miller, and Doman, 1996) .

In the bell example, temperament is analogous to the physical characteristics of the bell. Two bells of different shapes and material properties will not emit sounds of the same form, even when struck with the exact same stimulus. One bell may have to be struck much harder than the other before it will emit any sound. It may not ring properly if the strikes are in too quick succession. The innate physical properties of the bell are analogous to the innate neurochemical mechanisms of temperament; both influence the response to a stimulus. The shape of the response curve is determined by the bell's physical characteristics. Similarly, the shape and timing of the human emotional

response curves are influenced by temperament. For example, the response of the bell system to an instantaneous strike is described mathematically as:

$$y = a e^{-bt}$$

where y is the output sound intensity at time t, and the parameters a and b control the height of the response, and how fast it decays, respectively. The analogy for an emotion is that there is a rapid rise in emotional intensity upon a triggering event, followed by a natural decay in intensity of the emotion.

Figure 5.3 shows galvanic skin conductivity responses from two different people who were startled by acoustic tones. Both show peaks when they are startled, corresponding to greater arousal. Each peak also shows a natural decay, which can be modeled with a function such as the decaying exponential used above. The skin conductivity of the person who rated high in extroversion is lower on average than that of the person who rated low in extroversion. Studies have shown that extroverts are chronically less aroused than introverts (Kahneman, 1973).

Property of Linearity and Time-Invariance

There is signal processing theory that simplifies the analysis of emotion producing systems when they obey certain properties, especially the properties of *linearity* and *time invariance*. A *linear system* has an input and an output, and if you double the amplitude of the input, then the amplitude of the output doubles.[2] If you put nothing in, you get nothing out. If the response of the system to input A is B, and the response to input C is D, then if you make a new input by summing two others, $A + C$, then the new output will be the sum of the previous outputs, $B + D$. A system, linear or nonlinear, can be *time invariant*. Suppose you put A into the system at time t_i, and the system outputs B at time t_j. Now you put A in at a later time, $t_i + 3$. If the system is time-invariant, then the output B will occur at time $t_j + 3$. In other words, the system's behavior is not influenced by time.

The bell can be modeled with a linear time-invariant system under certain conditions. The input to the system is the striking of the bell, usually modeled by a short pulse, with amplitude proportional to the intensity of the strike. The output signal is the loudness of the bell's sound. The stronger the input pulse, the stronger the output sound; however, this linear relationship does not hold for all strengths of input. If you hit the bell too softly, it will not ring; if you hit it too hard, crushing it, the resulting sound will not be a proper ring, nor will the bell ring again. In other words, its response is linear only over a certain range of inputs. Similar limitations apply to the property of time-invariance. In general, it does not matter what time you strike the bell,

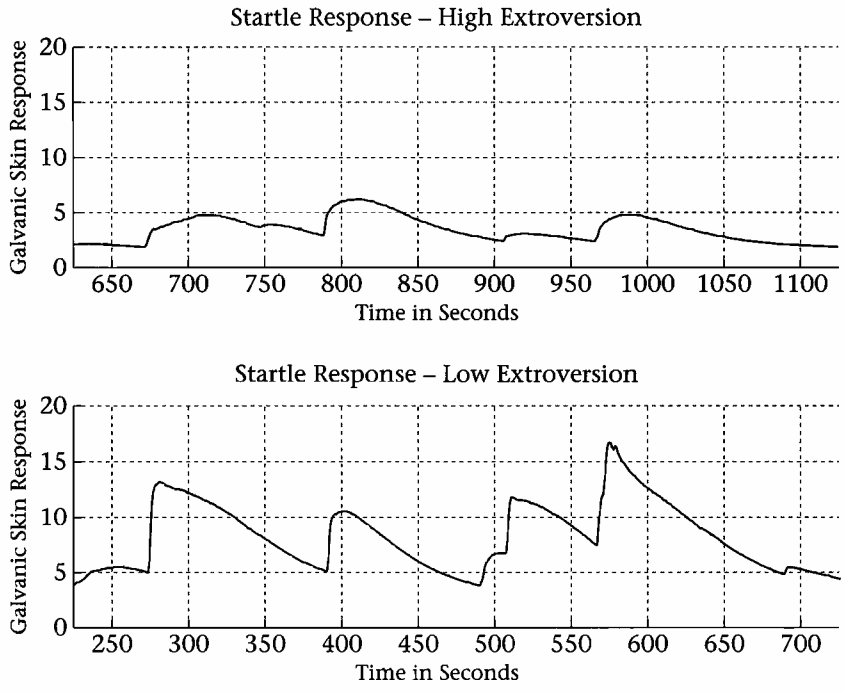

Figure 5.3
Startle tones of the same intensity were played one after another to two subjects, resulting in the galvanic skin conductivity responses shown here (in microSiemens.) Higher values tend to correspond to higher arousal in the person. Each tone resulted in a peak, but because they were all played within a few minutes of each other, habituation effects cause the responses to vary in amplitude. The signal on the top is a sample response for an extrovert; below it is a sample response for an introvert.

it will behave the same. However, if you strike it repeatedly, without giving it time to respond between strikes, or with such a rhythm as to interfere with its clapper moving and causing it to ring, then timing does matter. Most real-world systems are only linear and time-invariant under certain conditions.

When a system is linear and time-invariant, then it is much easier to understand how it will behave. All that is needed to completely characterize the behavior of the system is its response to a single special input, one that can be used to build all other possible inputs by scaling and summing them—that is, by linear combination. Once the response to this fundamental input is known, then the responses to all other possible inputs are known. We can completely characterize what the bell will do simply by ringing it once; this is true as long as we stay within the range in which the bell system is linear and time-invariant.

To some extent people are time-invariant systems. You can be startled today, and respond with a small jump. You can be startled similarly tomorrow and you will probably respond in the same way. However, like the bell that is struck too quickly, if you are startled too many times in a short period, then you will habituate to the signal, and your response to it will decline. With an emotional system, the situation is significantly more complicated than with the bell. Emotions are in part a function of novelty; consequently, the exact same input will generally not produce the same response over time. However, we can expect a similar input with the same level of novelty to produce a similar response in somebody over time. However, it should be kept in mind that there are hidden and uncontrollable factors, especially biochemical, that influence a person's response. It is hard to observe all the different variables at work in a person.

If people were always linear and time-invariant, then we could predict their responses to any input simply by characterizing how they respond to a few special inputs. However, these properties depend on how the system is defined, what are its inputs and outputs. There are usually many ways to choose inputs and outputs, which thereby change the definition of the system. For example, if the system is a person, and the input is a piece of music played for the person to hear, then the output might be the person's expression of happiness, given that she likes that piece of music. Suppose we choose the system input to be the intensity of the music waveform, and the system output to be the person's subjective rating of happiness. We can test if the system is linear by playing the same piece of music twice as loud, and seeing if the person's rating of happiness doubles. Alternatively, the input might be completely different: the number of pieces of music we play. If we play two pieces she likes, we ask if it doubles her subjective report of happiness. Of course there are also other possible outputs, such as the amount of curvature in her smile, or how much her heart rate and skin conductivity change.

Property of Activation and Saturation

We can already see problems with these linear systems in practice—two pieces may make someone twice as happy as one piece, and three may make them happier still, but eventually the effect saturates. The same property holds for the physiological components of emotion. Something that causes your heart rate to accelerate cannot do so indefinitely; the heart can only beat so fast. The same is true for neurotransmitter and hormone levels, and for all other bodily changes. Feelings can only reach certain heights, or depths. Consequently, linear systems only approximate human behavior under certain restrictions. In reality, human behavior is nonlinear. An important open research problem

involves characterizing how a person responds to different events under different conditions—analogous to characterizing the shape of different bells' responses to different situations. Responses might be both person-dependent and emotion-dependent. For example, an emotion like anger might have a more rapid response time than an emotion like grief, especially in a person prone to anger.

Temperament, mood, and cognitive expectation can influence emotion activation. For example, most people can tolerate some level of anger-producing stimuli before they actually feel angry; however, if they are in a bad mood, their tolerance may be lowered. Alternatively, someone with a cheerful disposition may have a higher innate tolerance. Also, certain personalities are more reactive or arousable than others, influencing emotional responsiveness, as seen above in Fig. 5.3. Cognitive expectation is also important. Suppose that you are watching a tennis tournament and your favorite player is expected to win easily. If she wins, then you are apt to feel happy, but probably not as happy as in the case where she is expected to be crushed by her opponent, and surprises everyone by emerging the victor.

How can all the influences I described be accounted for in a signal representation? Most of them are caused by a mixture of interacting physical and cognitive systems, with a potentially very complex set of interactions. A true physically based model is likely to be a tangle of parameters with nonlinear relationships, which may make it intractable for any practical uses. For example, there is no single input in the human system, unlike in the bell system. Instead, the input is a complex function of cognitive and physical events. Nonetheless, let me propose that these influences can be represented by the use of a simple nonlinear function applied to the inputs of an emotional system. This will result in the influences we have seen of differing activation and saturation levels, as well as providing an intermediate range of behavior that is approximately linear.

The proposed function is a "sigmoidal nonlinearity" described by the equation:

$$y = \frac{g}{1 + e^{-(x-x_0)/s}} + y_0.$$

This function is special in that it describes a large variety of natural phenomena. In this equation, x is the input, which may represent many possible stimuli, originating both inside and outside the person. The output is y, the height of the curve. In the bell analogy, the value of x is the strength of the actual strike, and the value of y is the effective value of the strike that is input into the linear system modeling the bell. All tiny values of x have the same effect: they make no sound. All very large values have the same effect: they

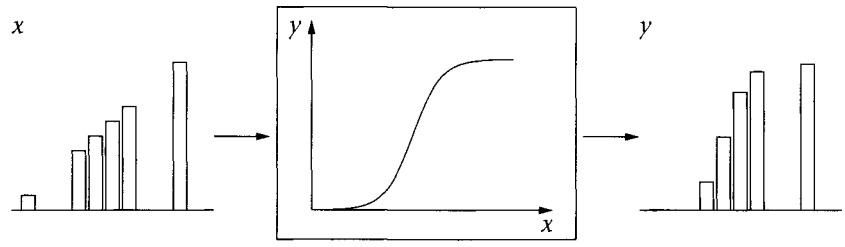

Figure 5.4
A sigmoid is applied to each input, x to convert it to a new value, y. If x is very small, y is zero. As x increases without bound, y clips to a maximum value. The middle of the transition region is approximately linear; values of x in this region pass through the sigmoid relatively unchanged.

make the maximum sound. In between lies the more interesting behavior. For inputs near the center of the curve, the response is approximately linear. Figure 5.4 shows the sigmoid applied to six inputs of different intensity. Only the medium values pass through relatively unchanged; the smallest value is ignored, and the biggest values are saturated.

In the equation above, the parameter s controls the steepness of the slope, representing how fast the output y changes with the input x. Smaller values of s make the sigmoid steeper, and more responsive. The steepness can be set according to personality. The behavior of a person who moves quickly from mild anger to losing their temper would be modeled with a steep sigmoid, compared to someone who endured a much greater range of intense events before losing their temper. The parameter x_o shifts the sigmoid left or right. When it is far to the right, then a stronger input is required to activate an output. The sigmoid might be shifted left or right according to a person's mood. A good mood can allow smaller inputs to activate positive emotions, accomplished by shifting the sigmoid to the left, as in Fig. 5.5. The parameter g controls the gain applied by the sigmoid. It is the same for all three examples in Fig. 5.5, but can be increased or decreased to control the overall amplitude of the sigmoid. This value might be coupled to the arousal level of a person; someone highly aroused might be capable of experiencing a greater intensity of emotion. Finally, the parameter y_o, shifts the entire curve up or down. This parameter might be controlled by cognitive expectation. For example, the expectation of a win in the tennis example above could dampen the joy of victory and accentuate the pain of defeat simply by lowering the sigmoids applied to positive and negative inputs. If the player wins as expected, the positive sigmoid tones down the positive input; if the player loses, unexpectedly, the negative sigmoid amplifies the negative input. The parameters of the sigmoid provide a rich set of controls for adjusting inputs before they proceed to activate emotions.

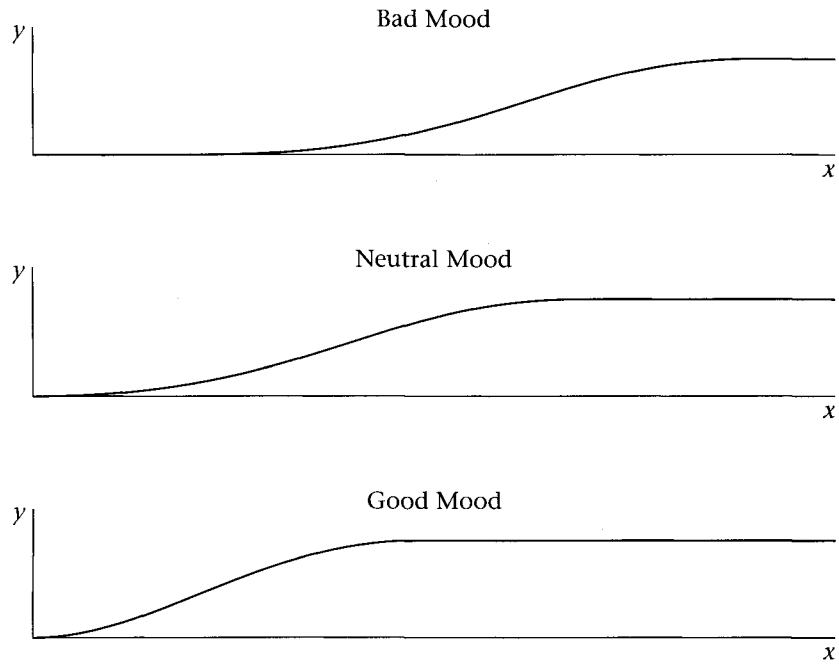

Figure 5.5
The parameters of a sigmoid are influenced by personality, as well as by cognitive and physical events. Here the sigmoids are shown shifted as an influence of mood.

The sigmoidal nonlinearity can also account for abrupt changes in an emotional response. As the parameter s approaches zero, the transition region of the curve becomes vertical, indicating that a tiny change in a certain range of input values can lead to a significant change in the output. The behavior is analogous to a physical phase transition—such as when water becomes ice. We might say that a person suddenly "snapped" or suddenly "went over the edge." We might expect this transition to be steeper for negative than for positive emotions, since one often finds that people hold back on negative emotions and then experience a catastrophic release, marked by a sudden burst of tears, or burst of anger, "This is the LAST straw!" as they reach their breaking point. In contrast, positive emotions can transition relatively smoothly. Finally, just like water undergoes two phase transitions: ice to water, and water to steam, an emotional response curve might have multiple transition regions, not just one as shown in these sigmoids. Someone may go through multiple stages of anger, each with its own identifiable region of behavior and discontinuity of transition. A signal representation is easily adapted to accommodate these cases.

Many models of emotion synthesis, which will be described in Chapter 7 employ activation thresholds for each emotion, but overlook the need for

representing saturation. The proposed signal representation accounts for both kinds of phenomenon, while also being able to represent other influences, such as those of cognitive expectation, temperament, and mood. However, as with the models presented later, there is no easy way to find the quantitative values of the emotion response functions—the parameters g, x_0, s and y_0 for the sigmoid, or the parameters a and b for the basic response of the system to an input. However, given that these parameters describe qualitative behavior, we can argue that their exact values are not as important as their relative values. For a person who is quick to fly into a rage and slow to be joyful, the slope for anger should be steep compared to the slope for joy, while the exact values may not be as important.

I have been deliberately vague about the nature of the inputs to the sigmoids, describing an input value x qualitatively as an event that might provoke an emotion. The figure illustrating the sigmoids gives the appearance that a single value of event intensity exists, which is input into the sigmoidal nonlinearity to generate the actual input to an emotion generation system. However, no single intensity measure has been found yet. The same is true of the Yerkes-Dodson inverted-U curve I showed earlier, which plots performance as a function of arousal. The Yerkes-Dodson law leaves unspecified how one would measure arousal and performance. It is understood that each axis represents a complex function of many variables. For example, there are multiple systems that contribute to arousal in the brain, each of which has a specific chemical identity. One group makes serotonin, another noradrenaline, another acetylcholine, and another dopamine (LeDoux, 1996). Arousal can also be observed in physiological changes such as pupilary dilation and galvanic skin conductivity. Furthermore, not all of these changes will happen in the same way with every arousing stimulus. To date, there is no single measurement that corresponds to arousal. Nevertheless, describing a complex concept with a single value provides a useful shorthand for many applications.

Property of Cognitive and Physical Feedback

A human emotional system can receive a so-called strike not only from an external event, but also from an internal event generated by a previous strike. In other words, the human system contains a feedback loop. If you sense the ceiling falling around you, your initial reaction may be to jump up and run out of the room. Often, such a response is immediately followed by cognitive feedback, such as "Oh, that's not the ceiling, it's a poster that came untacked." The cognitive feedback in this case tempers your bodily response. Your heart slows back to its normal pace and you can once again concentrate on what you were thinking previously. Alternatively the feedback

could have reinforced the response—for example, "Oh no, it's not just the ceiling, there go the bookshelves too!"—and caused your arousal level to climb even higher.

There are paths in the brain, especially between limbic system structures and the cortex, which are capable of carrying feedback. There is also physical feedback from the body, for example if you feel sad and you let your shoulders droop and your head hang, then this tends to reinforce the sad feeling. Alternatively you might *think* you should appear happy, adjust your posture and facial expression accordingly and seek out jovial events, which consequently mitigate your sad state. We have seen that your mood also influences which thoughts are retrieved—bad moods bias retrieval toward negative thoughts. Feedback can be physical or cognitive, and it can decrease or increase the intensity of an affective state.

Imagine if the bell was struck each time not only by somebody hitting it, but also by a force double the maximum intensity of its previous sound. As the sound increases, so does the force with which the bell is struck. If the bell could sound arbitrarily loud, then what we would have is a feedback system with an output that grows without bound.[3] Without some attenuating force to interrupt this feedback loop, the bell would quickly sound its death knell. In the human emotional system there is feedback. However, something keeps the emotional responses from growing arbitrarily large. The proposed sigmoid nonlinearity can be used to limit the results of feedback by bounding the output values of the system. Bounding these variables can keep the output from growing too large.

Representing Mood

Mood operates over relatively long time scales compared to emotions. Mood can be thought of as a background process that is always there, while emotions tend to come and go. Moods can predispose or bias a person toward certain emotions. A bad mood can make it easier for a negative-valenced emotion to be activated, while a good mood makes this more difficult. Although we usually think of moods as good, bad, or neutral, they also come with other distinctions. A bad mood due to anger has a high level of arousal; a bad mood due to immense sadness is marked by a depressed state, of low arousal. A peaceful good mood is low in arousal, contrasted with the good mood that accompanies an exciting new romance. The dimensions of valence and arousal provide a useful description of most moods.

Mood can exert its influence by adjusting the sigmoidal nonlinearites applied to inputs. A highly aroused bad mood can shift the sigmoid for negatively valenced events so that even a slightly negative event can pass into the system and activate responses. The high level of arousal can increase the

gain on the sigmoid. A good mood can shift the same sigmoid in the other direction, so that trivial negative events are ignored by the emotion-producing system. In this way mood can influence the generation of emotions.

But what generates moods? Unfortunately, scientists do not have an answer yet. However, we can build a flexible representation that can accommodate several possible generators of mood. For example, we might let any event influence mood, even if its valence is so insignificant that it lies below the activation region. For example, body chemistry—especially dietary influences, medication, and changes in hormones—can affect mood without necessarily eliciting an emotion. Subtle changes can be accumulated over a window of time, so that even if the system receives lots of inputs, each of which is too small to activate an emotion, they will nonetheless influence the mood. The mood can be constructed in this way: summing a function of recent positively valenced inputs and subtracting a function of recent negatively valenced inputs. Like emotions, moods cannot be of unbounded intensity. Physiological limits are imposed at some point. These limits can also be built into the computer model, either as hard limits, or as a saturating function on the outputs, as proposed above for emotions.

Example: Rafe

Let's illustrate how the above representations come together in a real situation with both physical and cognitive inputs: the scenario where Rafe gets hit by the out-of-control woman in a wheelchair (from Chapter 1). Rafe's emotional system has inputs that can be external or internal events. Some inputs are:

1. The oppressive heat and humidity (all time)
2. Watching the top pros play ($t = 1, 2, 3$)
3. Victory of favorite player ($t = 3$)
4. Pain of wheelchair slamming into Rafe ($t = 7$)
5. Bodily feedback ($t = 8$)
6. Recognizing Rebecca's accident and embarrassment. ($t = 9$)
7. Opportunity to help Rebecca ($t > 9$)

We are told that Rafe has a happy disposition. This is encoded with the sigmoid shown at the top right of Fig. 5.6, which lets almost all positive inputs pass through to influence emotions. The positive inputs described above are all of sufficient intensity to pass through Rafe's positive-input sigmoid. Consequently, each gives rise to positive emotions. I represent this in Fig. 5.6 as a sequence of small positive pulses giving rise to a sequence of small bell-

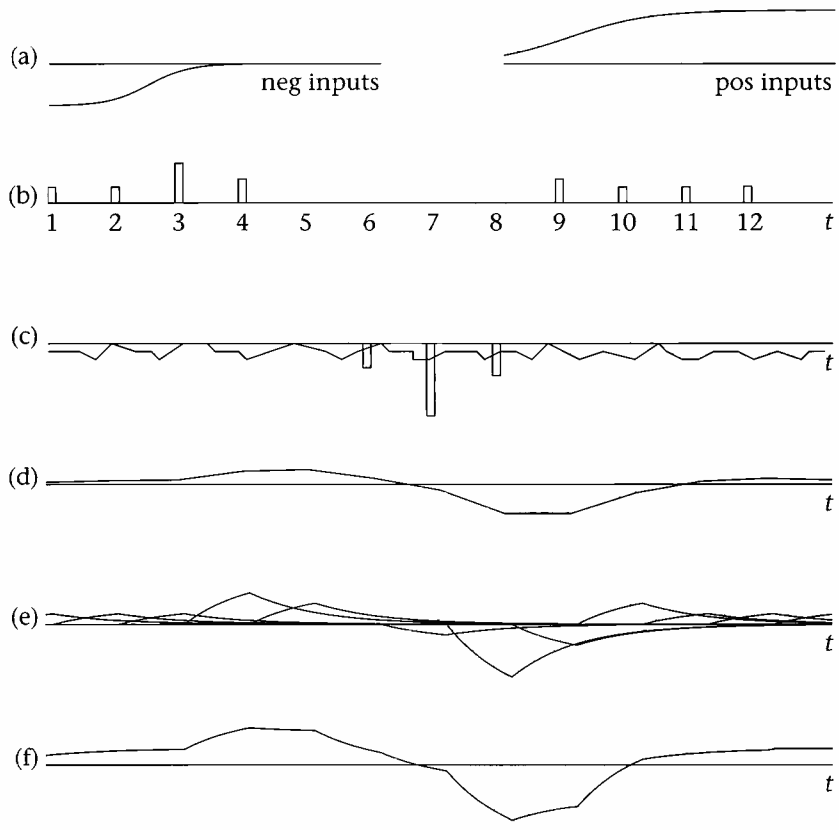

Figure 5.6
Signal representations for Rafe scenario. (a) Sigmoid for negative events (left) is less easily activated than the sigmoid for positive events (right). (b) Positive inputs in time. (c) Negative inputs in time. (d) Mood is a slow-varying function of all inputs. (e) Emotional responses arise only from the inputs that are of large enough intensity to pass through the sigmoids. (f) Accumulated positive and negative emotional responses.

like emotional responses. The victory of his favorite player is represented as a large positive input pulse, giving rise to a larger bell-like emotional response. After the accident, when he realizes Rebecca's situation and perceives the opportunity to help her, these appraisals elicit additional positive emotions.

In contrast, Rafe's sigmoid at the top left, for negative inputs, ignores small negative inputs, while quickly saturating strong negative inputs. The heat and humidity are always present, but their amplitude is so low that the sigmoid for negative inputs zeros their influence. I show this as a small negative noisy signal that contributes to the overall mood, but that is not significant enough to activate emotions. As these small negative inputs influence Rafe's mood, it gradually shifts his negative sigmoid toward the right, allowing less

significant negative inputs to influence his emotions. By the time he stands in line and appraises the oppressive heat and humidity, a small negative emotion is generated. A few moments later when the wheelchair slams into him, a big negative emotion is generated. An instant later, bodily feedback produces another negative input which passes through the sigmoid to generate a negative emotion.

As mentioned, all the inputs, no matter what their intensity, influence Rafe's mood. His mood is modeled by summing all of the positive and negative inputs, before they enter the sigmoids, and filtering the sum. Filtering forms a linear combination of previous inputs over time, weighted by how recently they occurred, to generate an instantaneous value of mood. The mood therefore tends to change more slowly than the emotions. The filter used to create the mood shown in Fig. 5.6 combined the present and previous three inputs, giving the greatest weight to the present input, and linearly decreasing weights to older inputs. The weighting and summation used here may not be the same as the physical mechanisms in the body that contribute to our feelings; nonetheless, they capture the qualitative changes that are plausibly associated with a real situation.

My illustrations of the Rafe situation only distinguish valence and intensity. For mood, these descriptions are sufficient for most purposes. However, we know that there is more differentiation among emotions than simply valence and intensity. A negative input that contributes to anger may not contribute to sadness, and vice-versa. Although the examples I gave only showed the use of signal representations for valence and intensity, the representations can also model other distinctions. For example, basic emotions such as anger, fear, joy, and sadness could have their own sigmoids to specify their activation and saturation characteristics. The different emotions could have inhibitory and excitatory influences on each other, either directly or via some intermediate mechanism. Velásquez's connectionist model, "Cathexis," is one way to implement such direct influences (Velásquez, 1997). Alternatively, the influence that mood already exerts can be used to regulate interactions between emotions. If an input is negative enough to generate anger, then it also will contribute to a bad mood. The bad mood automatically shifts all the sigmoids for the other emotions to the right, making the negative emotions easier to activate and the positive ones more difficult to activate.

In some cases it might be important to distinguish inputs not just as negative or positive, but also as physical and cognitive, especially since it is possible for one to feel physically bad (e.g., weak and exhausted) but mentally good (e.g., happy something is completed), or vice versa. However, the majority of the time people do not make such distinctions. It is common to hear someone say "I am feeling pretty good" or "I am not feeling so

good," lumping mental and physical components together. Using a signal representation does not require all these details to be specified, but permits a wide range of possibilities to be represented. The specific use of signals I have proposed here is intended to illustrate their ability to account for properties of emotions that we do know something about. The flexibility of the representation is an important advantage since many details of the human emotional system remain to be determined.

Use of a computational signal representation for moods and emotions raises specific questions for theorists. For example, can the effects of temperament and personality be adequately represented in terms of an emotional response function, activation function, and saturation function? What other assessments of inputs, besides positive and negative, are needed to account for the diversity of moods and emotions that can be elicited? The proposed representation includes a small number of parameters which capture particular degrees of freedom—such as the activation region of the sigmoid, or the decay rate of the emotional response; do these parameters control useful behaviors? The representation currently accounts for many properties, summarized below. There may be other properties, as yet undiscovered, to which it may or may not be adaptable. For example, theorists have not articulated what the role of noisy thoughts or other distractions are in terms of influencing the intensity of emotions; however, the signal representation easily accommodates the addition of a noisy signal, if this addition becomes important.

In summary, the proposed use of signals for describing the low-level behavior of emotions provides a flexible representation for moods and emotions which handles physical and cognitive inputs while including influences of temperament and personality. It therefore provides not only a tool for theorists who are trying to model some low-level behavior of emotions, but also a representation that a computer can use in modeling internal emotional signals, especially as part of a subsystem for generating and regulating emotions.

Summary of Properties

The proposed signal representation accounts for the following properties of behavior in an emotion system:

- *Response decay.* An emotional response is of relatively short duration, and will fall below a level of perceptibility unless it is re-activated.
- *Repeated strikes.* Rapid repeated activation of an emotion causes its perceived intensity to increase.
- *Temperament and personality influences.* A person's temperament and personality influence emotion activation and response.

- *Nonlinearity.* The human emotional system is nonlinear, but may be approximated as a linear system for a certain range of inputs and outputs.
- *Time-invariance.* The human emotional system can be modeled as independent of time for certain durations. For short durations, habituation effects occur. For longer durations, factors such as a person's physiological circadian rhythms and hormonal cycles need to be considered.
- *Activation.* Not all inputs can activate an emotion; they have to be of sufficient intensity. This intensity is not a fixed value, but depends on factors such as mood, temperament, and cognitive expectation.
- *Saturation.* No matter how frequently an emotion is activated, at some point the system will saturate and the response of the person will no longer increase. Similarly, the response cannot be reduced below a "zero" level.
- *Cognitive and physical feedback.* Inputs to the system can be initiated by internal cognitive or physical processes. For example, physiological expression of an emotion can provide feedback which acts as another input to the system, generating another emotional response.
- *Background mood.* All inputs contribute to a background mood, whether or not they are below the activation level for emotions. The most recent inputs have the greatest influence on the present mood.

Many of the properties listed here can also be accounted for by other models, which I will say more about in Chapter 7. In particular, the Cathexis model comes the closest to fulfilling the properties I articulated here. Let me caution that there can be different models or representations that satisfy a set of properties. I am deliberately not trying to establish one model or one theory of emotion in this book; I do not think there is *one* best model for all applications, nor is there sufficient understanding of human emotions to justify a comprehensive model at the level needed for computer implementation. Depending on the level of detail demanded by an application, different models will be preferable. The low-level signals I have illustrated here are good at handling emotion intensities, and may be useful both for theorists modeling emotion generation, as well as for computers that generate and regulate internal emotion signals. However, this low-level signal representation is not well-suited to address the high-level cognitive reasoning that may be involved in triggering emotions. My belief is that the latter will be better fulfilled by the use of higher-level representations which I will describe in Chapter 7, and in some cases, by some of the pattern models which I will describe in the next chapter. Throughout the remaining chapters I will illustrate different levels of representation that can be used advantageously within the framework of affective computing.

Physiological Signals

The signals I have shown so far do not represent the measurement of known physical quantities. However, there are many signals relevant to emotional responses that are physically measurable, especially by cameras, microphones, and sensors, the latter of which might be placed in physical contact with a person in a comfortable and non-invasive way. Signals gathered from four such sensors will be illustrated below, while others such as facial and vocal signals will be shown in the next chapter. Patterns of low-level signals can be combined with high-level information to recognize an affective expression, as well as to characterize an affective state.

Because people are already in physical contact with computers, augmenting their contact with sensors provides a new form of communication without much effort on the user's part. In particular, it is easy for a computer to gather signals such as the four shown in Fig. 5.7: electromyogram (EMG), blood volume pressure (BVP), galvanic skin response (GSR), and respiration, which I will say more about in a moment. The short segments shown in this figure illustrate very different responses obtained while an actress expressed two different negative emotions. Although clear differences can be seen in the signals for the two different emotions, we obtained data from the actress over 20 days, and sometimes found that the variations in the signals for the same emotion over different days were greater than the variations between the different emotions on the same day. In other words, the examples shown in Fig. 5.7 are some of the cleanest, the most illustrative of the differences; in practice, it is very hard to build a system to recognize just the differences between the emotions. I will say more about this later in the next chapter, and illustrate some recognition results on these signals.

The electromyogram (EMG) signal uses small electrodes to measure a tiny voltage from a muscle, indicating when it is contracted. The EMG shown in Fig. 5.7 measures the voltage emitted by the masseter muscle outside the jaw, which increases when the teeth are clenched in anger, as well as when there are certain other facial movements such as laughter. The EMG sensor could also have been placed elsewhere, such as on the trapezius muscle between the neck and shoulder, to sense tension in that muscle without the sensor being as visible as it is when placed on the jaw. The sharp peaks in the EMG signal in Fig. 5.7 were probably caused by the actress clenching her jaw during her expressions of anger.

The blood volume pressure (BVP) signal is an indicator of blood flow, gathered using a technique known as photoplethysmography, which shines infrared light onto the skin and measures how much of it is reflected. The BVP shown in Fig. 5.7 was taken from a small sensor worn on the fingertip. The

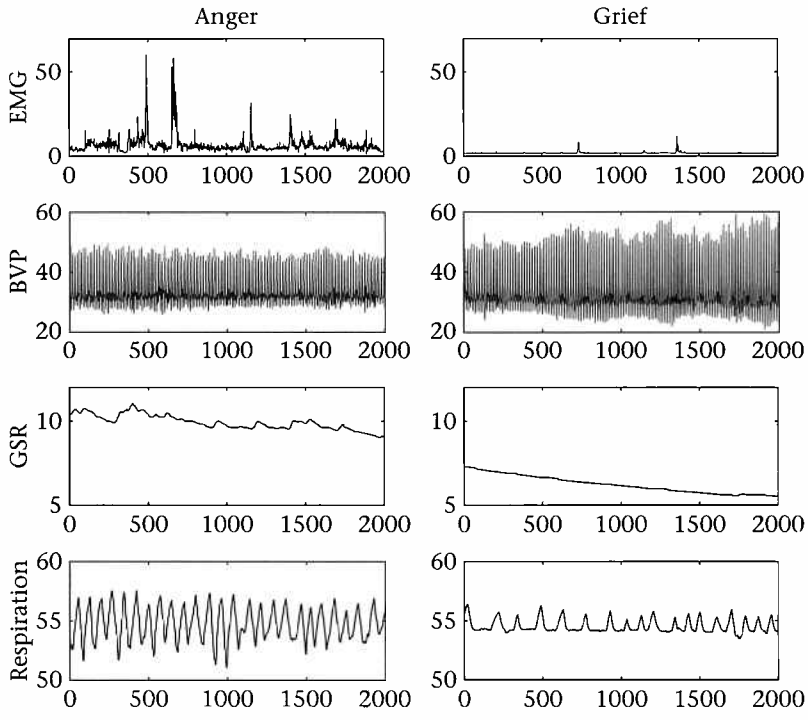

Figure 5.7
Examples of physiological signals measured from an actress while she consciously expressed anger (left) and grief (right). From top to bottom: electromyogram (microvolts), blood volume pulse (percent reflectance), galvanic skin conductivity (microSiemens), and respiration (percent maximum expansion). All of these signals can be gathered from sensors on the surface of the skin, without any pain or discomfort to the person. These signals were sampled at 20 samples a second, using a ProComp system from Thought Technology Ltd. Each box shows 100 seconds of response.

BVP waveform exhibits the characteristic periodicity of the heart beating, since each beat forces blood through the vessels. The overall envelope of the signal tends to pinch when a person is startled, fearful, or anxious. An increase in the BVP amplitude is caused when there is greater blood flow to the extremities, such as when a person relaxes.

The galvanic skin response (GSR) signal is an indicator of skin conductivity, and is measured via two small silver-chloride electrodes. An imperceptibly small voltage is applied and then conductance is measured between the two electrodes. The signal in Fig. 5.7 was gathered by placing these electrodes on two fingers of the actress's hand. Reliable signals can also be obtained from electrodes placed on the feet, if it is desired to keep the hands free from sensors. GSR tends to increase when a person is startled or experiences anxiety, and is generally considered a good measure of a person's overall level of arousal.

The respiration signal is sensed using a long thin velcro belt worn around the chest cavity, which contains a small elastic that stretches as the subject's chest cavity expands. The amount of stretch in the elastic is measured as a voltage change and recorded as a percent of its maximum change. The respiration sensor can either be placed over the sternum for thoracic monitoring or over the diaphragm for diaphragmatic monitoring. In Fig. 5.7 the signal was gathered using diaphragmatic monitoring. From the waveform, the depth of the wearer's breathing and the rate of respiration can be obtained.

To be handled by the computer, all human signals need to first be converted from their continuous form to a digital form. If these signals are facial or gestural motions, then they are usually gathered by a video camera and digitized at 30 frames a second. A speech waveform is gathered by a microphone and typically sampled at 16 KHz, with 16 bits per sample. Physiological signals such as those described above contain much lower frequencies than voice, and can be sampled reliably at only 20 Hz, usually with 32 bits per sample. Muscle potential changes can be sampled at 20 Hz to get large changes due to stress, but should be sampled at 1 KHz if it is desired to capture fine changes associated with fatigue, such as lactic acid buildup. After the sampling process, the computer has a representation of the signal as a sequence of binary numbers, which can then be analyzed to try to determine characteristics of the signal that correlate with expression of a particular emotion.

One of the applications I described earlier involves the use of physiological signals to gather responses such as frustration or distress from consumers trying out products. This application was inspired while a student was playing the video game DOOM, and we noticed that he exhibited more pronounced responses when there was a problem working the game controls than during any other event in the game. A short segment of three of his physiological signals can be seen in Fig. 5.8, which shows his GSR, BVP, and EMG over 5 minutes of time. His stress is initially signaled by the jaw-clenching peak in the EMG, which remains high during the minute and a half where the software controlled navigation keys failed to work as he expected. The point labeled "give up" is where he stopped the game and started over. After the game gets going, we see a constriction in his BVP, indicative of lowered blood flow to his extremities, and an increase in the GSR, indicating a state of higher arousal.

Summary

This chapter has overviewed issues of representation in affective systems, specifically the need for a mixture of representations, spanning low to high levels of processing. In particular, I emphasized that systems which can recognize, express, and "have" emotions will employ processing that involves

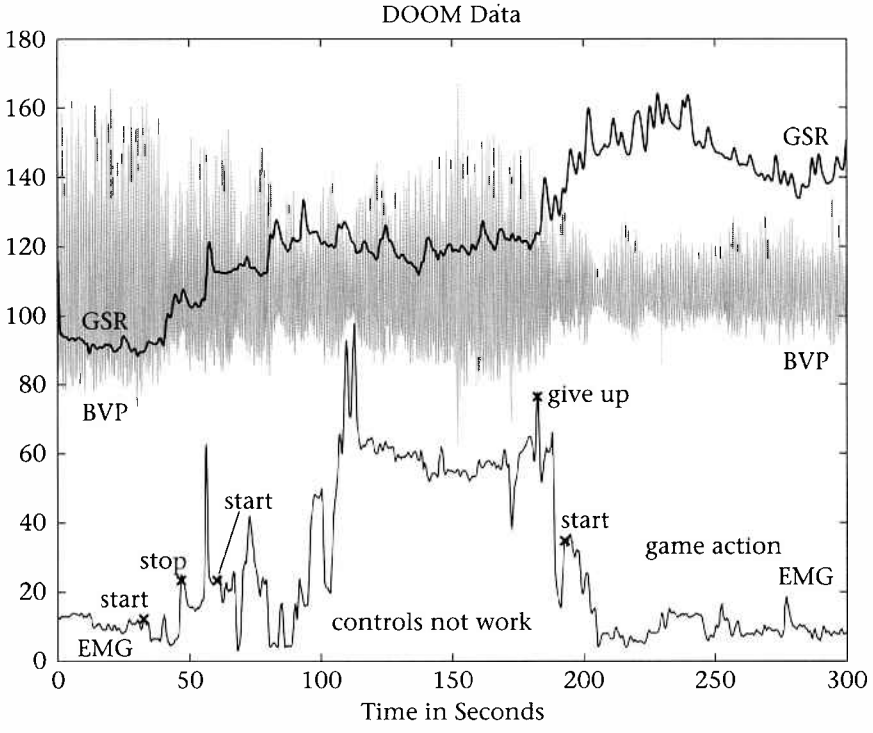

Figure 5.8
Three physiological signals measured from a student playing the game DOOM on the computer. The most significant response is shown in the center, where EMG peaked during loss of use of the controls.

both high-to-low and low-to-high transformations, from symbol to signal and from signal to symbol.

I have proposed the use of a low-level signal representation, in an abstract sense for representing intensities of emotions and moods, and in a physical sense, for representing waveforms measured of physiological changes characteristic of emotional states. The representation presented here is flexible, and accounts for many of the properties of emotions outlined in earlier chapters. I described certain questions it raises for theorists, as well as how it is suited for some applications and not for others. The next chapters will highlight higher-level means of representing affective information, especially for use in recognition and expression of affect, and for giving computers the ability to reason about emotion generation.