A Neuroscientific Approach to Emotion System for Intelligent Agents

Gunn-Yong Park[1,2], Seung-Ik Lee[1,2], and Joong-Bae Kim[1]

[1]Intelligent Robot Research Division,
Electronics and Telecommunications Research Institute,
P.O.B. 161, Gajeong-dong, Yuseong-gu, Daejeon, Korea
{pgy64257, the_silee, jjkim}@etri.re.kr
http://www.etri.re.kr
[2]Computer Software Engineer Division,
University of Science and Technology,
P.O.B. 52, Aueun-dong, Yuseong-gu, Daejeon, Korea
http://www.ust.ac.kr

Abstract. For an agent to be believable, emotion is an essential part of the human-computer interaction. In this paper, we propose a dynamic affective system, inspired from neuroscience. The system comprises four modules: an appraisal module, an emotion-generation module, an emotional expression module, and a long-term memory (LTM). The appraisal module gathers environmental stimuli and evaluates them to see if they are rewarding or punishing. Based on the results of the appraisal module and the agent's internal stimuli, the emotion-generation module generates the affective states of the agent. The emotional expression module combines and expresses emotional behaviors in accordance with the affective states. As a result, the proposed affective system can generate various emotions simultaneously and produces emotional life-like expressions.

1 Introduction

There have been several efforts to build a life-like agent for our convenience, such as the ubiquitous robotic companion [1], the visual software agent [2], and the power wheelchair for the physically handicapped persons [3]. For an agent to be more life-like and believable, an emotion is an essential part, because affective states affect rational decision-making, perception, learning, and other cognitive functions of a human [4]. According to the somatic marker hypothesis [5], the marker records emotional reaction to a situation. We learn the markers throughout our lives, and use them for our decision-making. Therefore, a believable agent needs an affective system to synthesize and express emotions. Due to the importance of the emotion in the human-computer interaction (HCI), there have been several attempts to build an emotional agent, such as Sony's 'Aibo' [6], MIT's 'Kismet' [7].

Most of the previous work to build emotional agents was inspired by ethology, and used an affect space to produce emotion. For example, Aibo has six emotions based on the affect space, called Takanishi's model [6], and generates appropriate emotional reactions to a situation. Because the emotions in the affect space have a competitive relationship with each other, the agent based on the affect space expresses only one emotion can be expressed at a time. For example, Aibo expresses only one affective state from its six emotions: happy, sadness, fear, disgust, surprise, and angry. However, contrary to the previous work, humans can have several emotions simultaneously and can express them in various ways.

According to neuroscientific studies, because of the considerable developments in temporal lobe and prefrontal cortex, a human emotional processing is more complicated than the affect space. Therefore, humans can generate and express a large range of emotion from happiness to anger. Furthermore, because people feel more comfortable with a human-like agent [4], a human-oriented agent should have an emotion generation process similar to that of a human. Hence, we propose an affective system for intelligence agents based on a neuroscience research.

This paper is organized as follows. Next Section presents the human emotion process in the aspect of both neuroscience and cognitive science. Section III describes the proposed affective system based on Section II. The experimental result and the conclusion are presented in Section IV and Section V, respectively.

2 Affective System in a Human Brain

Humans have quite well developed temporal lobe and the prefrontal cortex. Literatures [8, 9] have discovered that the development has contributed to the human's simultaneous expression of emotions and the ways they are expressed. In this section, we briefly introduce neuroscientific approaches to understanding human affective system.

Since the triune brain model [4] had been proposed, many people assumed that the limbic system is the source of emotions. However, according to the recent neuroscientific studies [8, 9, 10], emotion involves not only with the limbic system but also with the neocortex such as the orbitofrontal cortex. The amygdala and the orbitofrontal cortex are responsible for the perception, appraisal, and learning of environmental stimuli such as visual or acoustic sensory stimuli. The hypothalamus is in charge of homeostasis – maintaining the body's status quo. It controls body temperature, hunger, thirst, and other circadian cycles, and alters the affective state to obtain the desired stimuli.

Fig. 1 shows how the primate brain responds to environmental stimuli. In primates, external stimuli are decoded as a reward or a punishment. Rewards are stimuli for which an animal will work, whereas punishers are stimuli that the animal will work to avoid [8]. In the primate taste system, for example, the reward decoding occurs after several processing stages. The first stage is the primary taste cortex, that identifies the taste. Based on that result, the secondary taste cortex in orbitofrontal cortex evaluates

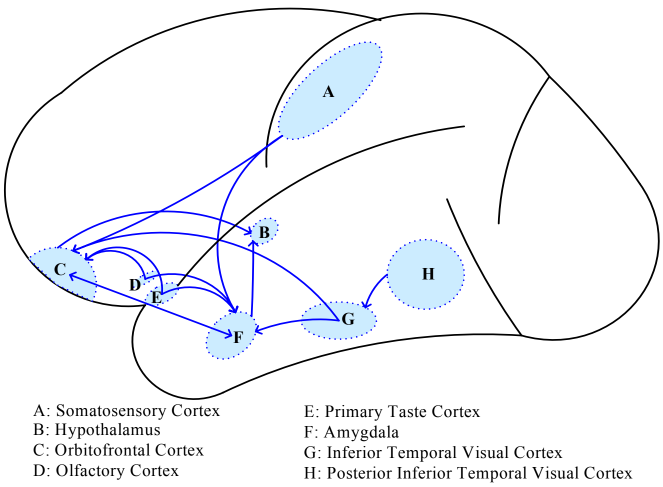


Fig. 1 Brain of the macaque monkey

its reward value. This value is decreased by feeding to satiety, since neurons in the orbitofrontal cortex decrease their responses as the reward value of the food decreases.

Rewards and punishers can be defined as reinforcers, because they change the probability of behavior [8]. There are two types of reinforcers: primary and secondary reinforcers [9]. The primary reinforcer (e.g., the taste of food or pain) is unlearned, and has innate reinforcing properties. The secondary reinforcer (e.g., the auditory or visual stimulus) develops its reinforcing properties through learning its association with primary reinforcement. For example, fear might be produced by a sound (i.e., the conditioned stimulus) that has been previously associated with a painful stimulus (i.e., the primary reinforcer). This type of learning is referred to as "stimulus-reinforcement association learning," [9] and involves with the amygdala and the orbitofrontal cortex.

According to neuroscientific studies, emotions are defined as affective states produced by rewards and punishers [8]. An animal changes its affective state and performs any actions to obtain rewards or to avoid punishers. For example, happiness could be produced by the occurrence of rewards such as an admiration of others or a pleasant touch; and frustration or anger could be produced by the termination of rewards such as the death of a loved one. As shown in these examples, the human emotional processing is involved in the occurrence, termination, or omission of the rewards (or punishers).

The affective states are classified as primary and secondary emotions [4]. The primary emotions are innate, and involve the primary reinforcers. For example, when you are surprised by a loud noise, the surprise involves with the primary reinforcer such as an acoustic sensory stimulus. On the other hand, the secondary emotions involve with the secondary reinforcers, and depend on the result of the stimulus-reinforcement association learning.

As described above, the environmental stimuli are evaluated as the primary or secondary reinforcers, and then the affective states depend on these reinforcers. For example, in the amygdala, the association learning is realized as modifiable synapses

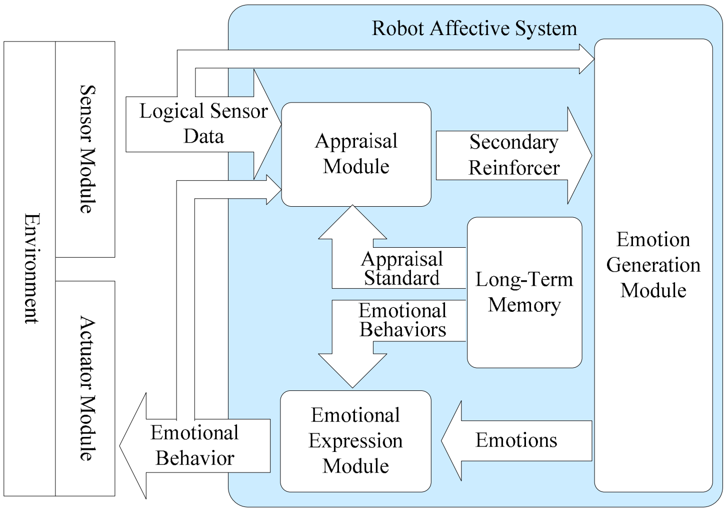


Fig. 2 Framework for the affective system.

between visual (or auditory) neurons and neurons from taste, olfactory or somatosensory primary reinforcers [9]. However, since the cognitive ability of intelligent agents is still far behind that of humans, this learning mechanism is difficult for an intelligent agent to adopt. Hence, we adopt a cognitive appraisal model as the stimulus-reinforcement association learning in the affective system.

3 Affective System Framework

To make an agent more user-friendly, the architecture of the system is based on neuroscience studies on primate's brains. Our affective system is shown in Fig. 2. The sensor module translates external stimuli into logical sensor information. The logical sensor data is an abstraction of the external stimuli, such as the name and position of an object. Emotional behaviors of the agent are realized by the actuator module. The affective system has a responsibility for generating and expressing affective states in according to the external or internal stimuli.

Inside the affective system, there are four modules: an appraisal module, an emotion generation module, an emotional expression module, and a long-term memory (LTM). The appraisal module describes the stimulus-reinforcement association learning of the secondary reinforcers. The function of producing affective states of the amygdala, the orbitofrontal cortex, and the hypothalamus, is implemented in the emotion generation module. The emotional expression module controls the sequential or parallel execution of emotional behaviors. These behaviors include, for example, the facial expressions of the agent or text-to-speech. Lastly, the LTM contains appraisal standard and emotional behavior for the appraisal module and the emotional expression module.

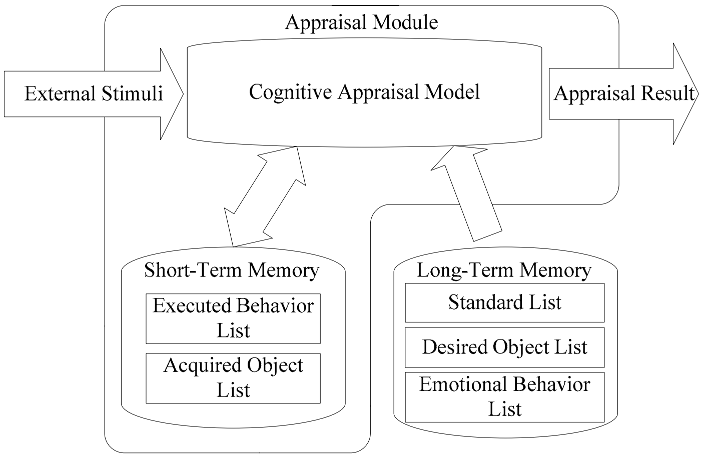


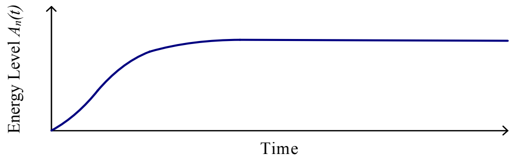
Fig. 3 Appraisal module of the robot affective system.

3.1 Appraisal Module

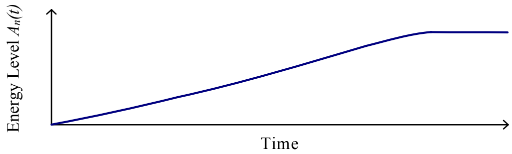
As shown in Fig. 3, the appraisal module comprises a cognitive appraisal model, a short-term memory, and an external LTM. The LTM contains a standard list and a desired object list for the cognitive appraisal model. The standard list and the desired object list are an appraisal reference for the praiseworthiness of the actions and the appeal of the objects, respectively.

The short-term memory contains an executed behavior list and an acquired object list. All the emotional behaviors executed by the agent are included in the executed behavior list. This module examines whether the result of the listed behaviors is received within the specified time interval. The acquired object list maintains the visual properties (placement, color, and name) of the recognized object. The agent obtains the physical changes – velocity, direction of moving objects – by comparing the data of objects from sensors with these of the short-term memory.

The cognitive appraisal model assesses external stimuli in accordance with the LTM and the short-term memory. For example, the facial expressions of users and the recognized objects determine the emotions related to the *fortunes of others* and *attraction* in TABLE. 1, respectively. When a recognized object is in the desired object list, the appraisal result is love. The action of the agent is related to the *attribution* in TABLE. 1. It depends on the praiseworthiness of the action of the agent. If the action is in the standard list, it is a praiseworthy one and then the appraisal result is pride. When the agent executes an emotional behavior, it is recorded in the executed behavior list, and waits for the expecting result. This result is evaluated through *prosepct relevant* and *prospect-based*. The appraisal of the action depends on whether the prospect is relevant. If it is relevant, emotions depend on whether it is confirmed within a few seconds.



(a) Appraisal energy level in a similar mood



(b) Appraisal energy level in a different mood

Fig. 4 Characteristic graph of moods.

TABLE I
ASSESSMENT RESULT OF THE COGNITIVE APPRAISAL MODEL

Category of Appraisal result		Appraisal result	
		Positive	Negative
Attraction		Love	Hate
Attribution		Pride	Shame
Fortunes of Others		Happy for	Resentment
Prospect Relevant		Joy	Fear
Prospect-Based	Confirmed	Satisfaction	Fear-Confirmed
	Disconfirmed	Relief	Disappointment

In the appraisal module, the external and the internal stimuli are classified into 6 categories according to the cognitive appraisal model (see TABLE. 1). The energy level of an appraisal result, $A_n(t)$, depends on time and mood. In a good mood, the energy level of the positive appraisal result increases more rapidly than that of the negative one, and vice versa (Fig. 4).

The mood *Mood(t)* is defined as follows:

$$Mood(t) = \sum_{k=1}^{A} e_k \times E_k(t-1)\,, \tag{1}$$

where $E_k(t)$ is the activation level of the k_{th} emotion, A is the number of emotions, and e_k is the gain parameter of the k_{th} emotion. If $E_k(t)$ is positive, then e_k is also positive, and vice verse. When *Mood(t)* is positive, the system is in the positive mood.

3.2 Emotion Generation Module

According to Ekman's research on emotion [11], we express our elemental emotions in the same ways, and can recognize these emotions in the face of another: anger,

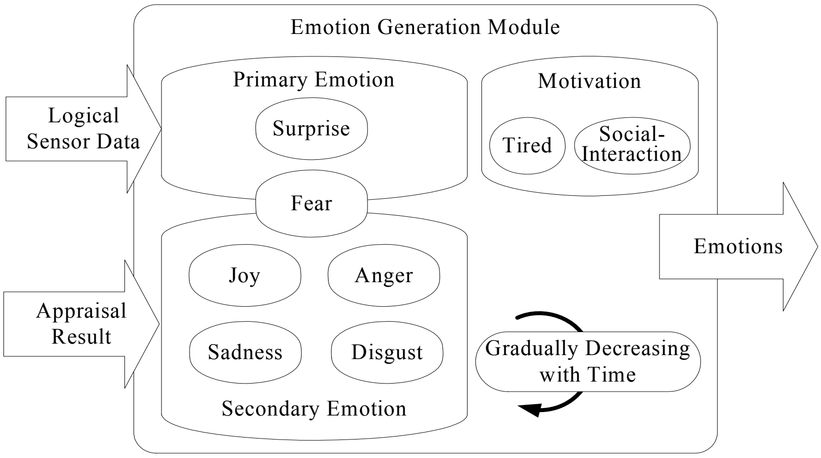


Fig. 5 Emotion–generation module of the robot affective system.

disgust, fear, joy, sorrow, and surprise. Therefore, we propose that the affective state of the agent is the blending of the six basic emotions. In the emotion-generation module, the activation level of each emotion is continuously updated. When the activation level exceeds a threshold value, the emotion is activated. All emotions above the thresholds are active in contrast to the winner-take-all scheme that allows only one emotion to be active at a time.

The emotion-generation module comprises three units: the primary emotion unit, the secondary emotion unit, and the motivation unit (Fig. 5). The primary emotions involve the primary reinforcers. They are linked directly to the logical sensor data, because the primary reinforcers are unlearned stimuli. On the other hand, the secondary emotions involve the secondary reinforcers. Therefore, they depend on the results of the appraisal module. The motivation module implements homeostatic regulations in the hypothalamus, such as ‘tired’ and ‘social interaction.’

The activation level of the i_{th} primary emotion $E_i(t)$ at time t depends on the level of logical sensor data, and defined as follows:

$$E_i(t) = E_i(t-1) + \sum_{m=1}^{M} w_m \times S_m(t) - \delta(t) \,, \tag{2}$$

where S_m is the level of the m_{th} logical sensor data, w_m is the weight of the m_{th} sensor data, $\delta(t)$ is the decay factor of the emotion, and M is the number of logical sensors related to the emotion. For example, if the emotion ‘surprise’ depends on the microphone level, the agent is surprised by a loud noise. In this system, the elements of the surprise are a loud noise and an abrupt change in the visual stimulus.

On the other hand, the secondary emotions depend on the appraisal result of the appraisal module, since it is related to the stimulus-reinforcement association learning. The activation level of the j_{th} secondary emotion at time t, $E_j(t)$, is affected by the energy level of appraisal result $A_n(t)$ and that of the motivation $M_l(t)$, as follows:

$$E_j(t) = E_j(t-1) + \sum_{n=1}^{N} a_n \times A_n(t) + \sum_{l=1}^{L} c_l \times M_l(t) - \delta(t) \,, \tag{3}$$

where a_n is the weight of the n_{th} appraisal result, c_l is the weight of the l_{th} motivation, N is the number of related appraisal results, and L is the number of related motivations.

The motivation unit implements homeostatic factors, such as 'tired' for taking a rest and 'social interaction' for interacting with people. In the emotion-generation process, as shown in (3), the activation level of the emotion $E_j(t)$ is affected by the activation level of the motivation $M_l(t)$. The motivation level $M_l(t)$ increases gradually when internal or external stimuli do not satisfy the desired stimuli of each motivation. As shown in (3), the motivation controls the emotion to execute appropriate actions, and consequently tries to obtain the required stimuli.

3.3 Emotional Expression Module

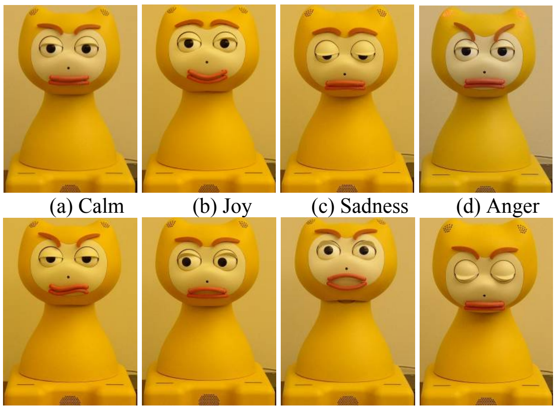
The emotional expression module controls the emotional actions of the agent. It determines appropriate affective behavior using the LTM. The LTM contains two types of emotional behaviors: concurrently executable behaviors, and sequentially executable behaviors. An execution of a sequential behavior depends on which activation level of the emotion is the highest. That is, the emotional behavior related to the highest level has priority over the others. In the concurrently executable behaviors, all behaviors that exceed the thresholds are performed.

The emotional expression module should deal with competing rewards, goals, and priorities. Furthermore, the selection process among the emotional behaviors can respond to many different types of reward. Therefore, the emotional behavior has properties, such as the expected result, if any; the processing type (e.g., sequential execution or parallel execution); and the waiting time. The expected result is the available stimulus from the emotional behavior, and the waiting time is the maximum time allowed to receive the expected stimulus from the environment.

4 Experimental Results

To evaluate the proposed system, it was applied to the Philips iCat, which is an experimentation platform for HCI. The iCat has four touch sensors, two microphones, a CCD camera, and 13 servos for facial postures. The affective system runs on a 3 GHz PC running Windows, and communicates with the iCat through a USB port. The affective system produces the agent's emotional states in accordance with the internal and external stimuli, such as interactions between the user and the iCat. Fig 6 illustrates the basis facial postures based on emotional states.

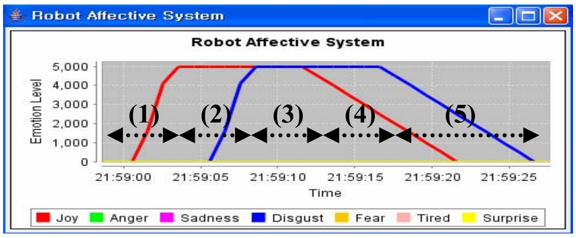
As shown in Fig. 6, the LTM contains eight basis facial postures, each of which represents a designated emotion. The agent's facial expression is defined as a blending of basis facial postures (Fig. 7). For example, at the beginning of (2) in Fig. 7 (a), the intelligent agent detects a desired object, and the activation level of joy increases. The facial expression of joy is in (2) of Fig. 7 (b). As depicted in (3) of Fig. 7 (a), disgust is activated because of the recognition of a hateful object or the execution of a blameworthy action, and the facial expression in (3) of Fig. 7 (b) is a blending of two



(a) Calm (b) Joy (c) Sadness (d) Anger

(h) Disgust (i) Fear (j) Surprise (f) Tired

Fig. 6 The basis facial postures of the iCat.



(a) Activation level of joy (red) and disgust (blue)

(1) (2) (3) (4) (5)

(b) Sequence of facial expression in accordance with (a)

Fig. 7 The activation level of emotions and facial postures.

basis facial postures (joy and disgust). As shown in (4) of Fig. 7 (a), when the desired object has disappeared or has been perceived for a long time, the activation level of joy decreases, and the emotional expression is in (4) of Fig. 7 (b).

As described above, the facial expression of the iCat is a blending of basis facial postures. In this system, external stimuli are evaluated according to the emotional reactions and each emotional expression is generated in accordance with the emotion activation level. There are two types of emotional expression: a sequential or parallel behavior. According to the affective states of the agent, the agent performs sequentially executable behaviors, such as emotional exclamations or simple text-to-speech, or a linearly combined parallel behaviors, such as facial expressions, and sequential behaviors. Hence, in contrast to previous work, the proposed system can generate and express various emotions continuously.

5 Conclusion

We have presented an affective system based on neuroscience and cognitive theory. The proposed affective system evaluates which external stimuli are rewarding or punishing. According to the appraisal result, this system generates and expresses various emotions simultaneously. Therefore, the emotional behavior, such as facial expressions, is not a predefined emotional behavior, but the blending of concurrently executable behaviors or the sequential execution of sequentially executable behaviors. Hence, the agent can produce human-like emotional expressions continuously.

In future, we intend to implement personality and association learning of the secondary reinforcer. Personality has an important role in emotion processing and can change the emotional reaction to the environment. Therefore, the believable agent should have characteristic personality. Furthermore, we will also focus on the association learning for the emotional response of newly discovered objects.

References

1. Y. G. Ha, J. C. Sohn, Y. J. Cho, and H. S. Yoon, "Towards a Ubiquitous Robotic Companion: Design and Implementation of Ubiquitous Robotic Service Framework," ETRI Journal, vol.27, pp.666-676 (2005)
2. Y. Nozawa, H. Dohi, H. Iba and M. Ishizuka, "Humanoid Robot Presentation Controlled by Multimodal Presentation Markup Language MPML," IEEE Int'l Workshop on Robot and Human Interactive Communication, Japan (2004)
3. K.H. Kim, H. K. Kim, J. S. Kim, W. H. Son, and S. Y. Lee, "A Biosignal-Based Human Interface Controlling a Power-Wheelchair for People with Motor Disabilities," ETRI Journal, vol.28, pp.111-114 (2006)
4. R. W. Picard, Affective Computing, the MIT Press, USA (1997)
5. A. Damasio, Descarte's error: Emotion, reason and the human brain, NewYork, NY: Harper Collins (1994)
6. R. C. Arkin, M. Fujita, T. Takagi, R. Hasegawa, "An Ethological and Emotional Basis for Human-Robot Interaction," In Proceedings of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems (2002)
7. C. Breazeal, Sociable machines: Expressive social exchange between humans and robots, Ph.D. dissertation, Dept. Elect. Eng. And Computer. Sci., MIT, Cambridge (2000)
8. E. T. Rolls, "The neural basis of emotion," International Encyclopedia of the Social and Behavioral Sciences, Eds. N. J. Smelsner and P. B. Baltes, Pergamon: Amsterdam, pp.4444-4449 (2002)
9. E. T. Rolls, "Vision, emotion and memory: from neurophysiology to computation," Cognition and Emotion in the Brain, Eds. T. Ono, G. Matsumoto, R. R. Llinas, A. Berthoz, R. Norgren, H. Nishijo, and R. Tamura, Elsevier Science: Amsterdam, pp.547-573 (2003)
10. S. Greenfield, Brain Story, BBC World Wide Publishing (2000)
11. P. Ekman, "Are there basic emotions?" Psychological Review, 99, pp.550-553 (1992)