

Towards Background Emotion Modeling for Embodied Virtual Agents

Luís Morgado^{1,2}

¹Instituto Superior de Engenharia de Lisboa
R. Conselheiro Emídio Navarro, 1
1949-014 Lisboa, Portugal
(351)218317217
lm@isel.ipl.pt

Graça Gaspar²

²Faculdade de Ciências da
Universidade de Lisboa
1749-016 Lisboa, Portugal
(351)217500609
gg@di.fc.ul.pt

ABSTRACT

For the realistic simulation of embodied agents we need a model of emotion that represents both structural and dynamic aspects of emotional phenomena to serve as background support for multifaceted emotion characterization. In this paper we present an emotion model oriented towards that aim, which provides a continuous modeling of the evolution of emotional phenomena. We also illustrate how it can be used to provide different perspectives of an emotional situation, namely by identifying emotional patterns that can be characterized as discrete emotional states.

Categories and Subject Descriptors

I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence - *Intelligent agents*.

General Terms

Design, Theory.

Keywords

Virtual Agents, Emotion, Cognitive Modeling.

1. INTRODUCTION

An increasing attention has been given to emotion modeling for embodied agent implementation as growing evidence indicated the encompassing role that emotional phenomena plays in humans and other organisms. From an interaction perspective, it is well known that our gestures and speech are colored by emotional content, which gives them meaning even without us being aware of that. That emotional content plays a crucial role in verbal and non-verbal communication and in social interaction [16]. However, increasing experimental evidence from neurosciences indicates that the communicative role of emotion is rooted on the pervasive influence of emotional phenomena in multiple aspects of behavior and cognition. For instance, experimental results reported by Damásio [8] indicate that a selective reduction of emotion is at least as prejudicial for rationality as excessive emotion, and Gray *et al.* [13] reported neural evidence for a

strong highly constrained form of emotion-cognition interaction, with loss of functional specialization, indicating that emotion and higher cognition can be truly integrated. Recently, Mallet *et al.* confirmed the tight integration between emotion, cognition and behavior by precisely localizing a sub-thalamic region where motor, cognitive and emotional information is integrated [22]. In what concerns the implementation of realistic embodied agents, that kind of evidence indicates that a prescriptive approach to emotion modeling, based only on the structural classification of emotional phenomena, may be limitative, since the processes underlying emotional phenomena are not addressed, therefore strongly limiting the ability to model their dynamic aspects. From this point of view, emotion models are needed that incorporate both structural and dynamic aspects of emotional phenomena and that are able to serve as a background support for multifaceted emotion modeling and expression. It is also important that these models be able to integrate and extend agent models of different kinds and levels of complexity.

Emotion corresponds in practice to a multitude of interrelated phenomena that occur at different organizational levels, from stereotyped stimulus-response reactive levels to feed-forward anticipatory levels based on internal representations of the world (e.g. [36]). In this paper we present a framework for emotion modeling based on this view where motivational-emotional dynamics are identified and defined to provide a uniform support for modeling emotional phenomena at different levels of organization. A primary contribution of the paper is to show how it is possible to model both identifiable emotional patterns and, at the same time, the continuous dynamic nature of emotional phenomena, in a common framework able to support the development of agents of different kinds and levels of complexity. Based on this framework different facets of emotion modeling can be explored, providing a versatile support for embodied agent simulation.

The paper is organized as follows: in section 2, we present an overview of the emotion model that supports the proposed approach; in section 3, we build upon that foundational framework to present an operational concretization of the model; in section 4, we present an example to illustrate the use of the model to characterize different facets of emotional phenomena; in section 5, we establish comparisons with related work; and in section 6, we draw some conclusions and directions for future work.

Cite as: Towards Background Emotion Modeling for Embodied Virtual Agents, L. Morgado, G. Gaspar, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Mueller and Parsons (eds.), May, 12-16, 2008, Portugal, pp. XXX-XXX.
Copyright © 2008, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

2. MODELING ARTIFICIAL EMOTION

The study of emotional phenomena, particularly the development of emotion models for the implementation of artificial agents has been mainly based on a perspective of emotion as a human phenomenon, possibly shared by some animals in the evolutionary continuum, according to the existence of specific brain structures. Cognitive appraisal models are a clear example of this approach, but even physiologically based models are inspired by human physiology and behavior. This anthropomorphic approach has an important drawback. Due to the tightly intertwined relation between emotion and cognition [2] emotional phenomena in this context are highly complex with multiple emotion blends. In fact phenomenological observations indicate that the complexity of emotional phenomena has its highest in humans [15]. From this point of view, we can understand why multiple emotion models and theories coexist and their difficulty to characterize emotion in a concise way, especially in what relates to the processes underlying emotional phenomena.

On the other hand, in the evolutionary continuum there is no evidence of a discontinuity in what regards the existence of emotional phenomena. On the contrary, it is well known that some simple organisms, even unicellular organisms, can present remarkable behaviors for organisms without nervous system, which from an observer point of view are easily classified as emotional [37].

Although almost unexplored, these observations give rise to the possibility that emotional phenomena are rooted not on cognitive or even on nervous structures, but on biophysical principles that are pervasive among biological organisms. In our work we explore this line of research by defining an emotion model that is inspired on basic mechanisms that have been proposed as sustaining the structure and activity of biological organisms. Underlying the proposed model is a view where “basic biological organization is brought about by a complex web of energy flows” [3]. Since the proposed model is based on the interchange between an organism and its environment, expressed mainly as energy flows, we called it *flow model of emotion*. Its overall structure will be presented next.

2.1 The Flow Model of Emotion

Two main aspects are recognized as fundamental to emotional phenomena: (i) the relation between the agent and its environment - for instance Lazarus identifies the “person-environment relationship” as the “basic arena of analysis for the study of the emotion process” [19]; (ii) the agent’s ability to cope with the current situation (e.g. [20]). Underlying these two aspects is another central concept, motivation, that is, the driving force that directs agent behavior (e.g. [4]).

In cognitive appraisal models, emotional states result from the assessment of the agent-environment relationship by deliberative mechanisms according to agent’s goals (e.g. [12, 18]). On the other hand, in behavioral models, emotional states result from a tight agent-environment coupling in the context of stimulus-response patterns and homeostatic drives (e.g. [6, 38]). These two descriptive levels have however known limitations, in particular the difficulty to model the dynamic non-linear nature that characterizes emotional phenomena [35]. Some approaches have been proposed to address this problem (e.g. [5, 18]), however they

are mainly based on parameterized intensity and decay functions defined empirically. Is there another descriptive level that can provide an adequate support to address this problem? The answer to this question could be yes, if we consider that emotional phenomena may not be rooted on cognitive or even on nervous structures, but on more fundamental biophysical principles.

To address that level, an adequate model of agent organization and internal structures and mechanisms is needed. That model must be able to represent thermodynamic aspects that occur at a biophysical level. An adequate base support is the notion of *dissipative structure* [32]. Dissipative structures are open systems governed by the interchange of energy with the environment and able to maintain themselves in a state far from equilibrium, yet keeping an internally stable overall structure. The maintenance of that internal stability in spite of environmental changes is done through feedback networks that motivate the system to act.

These feedback networks incorporate the basic notion of motivation (e.g. [4]) by maintaining in definite ranges of values internal and external variables that represent some form of energetic potential. The maintenance of a basic life support energy flow can be seen as a base motivation. From this base motivation other forms of motivation emerge according to the agent internal structure and organizational context, as is the case of *drives*, at an homeostatic level, or *desires*, at a deliberative level.

The basic flow of energy can be directly related to central aspects of emotional phenomena previously referred, namely, the agent-environment relationship and the relation between motivation and emotion. These two aspects are intrinsically associated in order to support the maintenance of the structure and activity of organisms through self-generation, a process known as *autopoiesis* [23]. This is the basic context that motivates the flow model of emotion.

From the relation between an agent’s internal motivation and its external situation, expressed through the energy flow that results from agent-environment interaction, behavioral dynamics arise that, from our point of view, constitute a particularly adequate support to characterize emotional phenomena in a way that is independent of the type or level of complexity of the agent.

From a thermodynamic point of view, to achieve its motivations an agent must apply an internal potential to be able to produce the adequate change in the environment. However, the concretization of the intended change depends also on the characteristics of the current environmental situation. That agent-environment relation can be modeled as a coupling conductance. Therefore, the process underlying motivation achievement can be modeled as a relation between an agent’s internal potential, its *achievement potential*, and the agent-environment coupling conductance, the *achievement conductance*. The achievement potential represents the potential of change that the agent is able to produce in the environment to achieve the intended state-of-affairs. The achievement conductance represents the degree of the environment’s conduciveness or resistance to that change, which can also mean the degree of environment change that is conducive, or not, to the agent intended state-of-affairs.

The achievement potential can be viewed as a force (P) and the achievement conductance as a transport property (C). The agent-environment relation underlying motivation achievement can therefore be characterized as a flow, called *achievement flow* (F),

which results from the application of a potential P over a conductance C . From the relation between achievement potential and achievement conductance, expressed as achievement flow, internal dynamics arise that underlie agent behavior.

Two independent components characterize those dynamics, the rate of change of P and the rate of change of C . However, the conductance C is not directly observable, instead, it is the resulting achievement flow F that is observable, representing an expression of situation change according to the agent-environment relation. Therefore, we consider the following two components:

$$\delta P = \frac{dP}{dt} \quad \text{and} \quad \delta F = \frac{dF}{dt} \quad (1)$$

From a qualitative point of view, we can identify basic patterns of evolution of the motivation achievement process. When $\delta P > 0$, the agent is able to handle the situation, in the sense that its achievement potential is increasing. In that case two basic patterns can be identified:

- $\delta F > 0$, a pattern of evolution where the agent is able to handle the situation ($\delta P > 0$) and the situation is evolving favorably ($\delta F > 0$), directly related to *joy/happiness*;
- $\delta F < 0$, a pattern of evolution where the agent is able to handle the situation ($\delta P > 0$) but the situation is evolving adversely ($\delta F < 0$), directly related to *frustration/anger*.

When $\delta P < 0$, the agent is not able to handle the situation, in the sense that its achievement potential is decreasing. In that case two more basic patterns can be identified:

- $\delta F < 0$, a pattern of evolution where the agent is not able to handle the situation ($\delta P < 0$) and the situation is evolving adversely ($\delta F < 0$), directly related to *apprehension/fear*;
- $\delta F > 0$, a pattern of evolution where the agent achievement potential is decreasing ($\delta P < 0$) in spite of a neutral or favorable evolution of the situation ($\delta F > 0$), which can be related to *despondency/sadness*. For instance, according to experimental studies (e.g. [19]) sadness often results from sudden losses of resources, which in our model can be represented by the achievement potential (a sudden decrease of P produces a negative variation of δP).

From the above characterization we can observe that the dynamics represented by δP and δF convey a kind of information that has an intrinsic emotional nature. These dynamics underlie the operation of agent internal processes, namely by preparing the agent to act according to the situation. That condition was referred by Frijda as *readiness for action* [10].

To explicitly represent the emotional dynamics conveyed by the change of the achievement potential and the achievement flow, we define a vectorial function ED , called *emotional disposition*:

$$ED \equiv (\delta P, \delta F) \quad (2)$$

This notion of *emotional disposition* is defined as an action regulatory disposition or tendency, but it does not constitute in itself an emotion. In the proposed model, phenomena such as emotions and moods arise from the background effect of emotional dispositions across different organizational levels, generating increasingly reach emotional phenomena, especially at self-reflexive and social levels.

2.1.1 Emotional Qualitative Characterization

A basic problem of emotion modeling is how to explain in a single model both the dynamic, continuously fluctuating nature of emotion processes and the existence of discrete labels referring to steady states [35]. To address these questions we need to formalize a qualitative characterization of emotional phenomena.

As can be seen in figure 1.a, at a given instant $t = \tau$ an emotional disposition vector has a quality, defined by its orientation (or argument) and an intensity defined by its module. That is:

$$Quality(ED) \equiv \arg(ED) \quad (3)$$

$$Intensity(ED) \equiv |ED| \quad (4)$$

Each quadrant of the two dimensional space $\delta P \times \delta F$ can be directly related to a specific kind of *emotional disposition quality* (figure 1.b) according to the four main patterns of evolution, previously described.

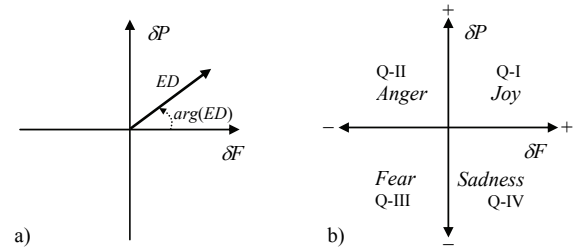


Figure 1. Vector ED as a function of δP and δF (a); relation between ED quadrants and emotional quality tendency (b).

It is important to note that the emotional tendency associated to each quadrant (joy, anger, fear, sadness) is only indicative of its main nature, since the quality of the emotional disposition is continuous. This is consistent with phenomenological well-known emotion blends. In this way it is possible to use the same model to explore the existence of specific emotional patterns and, at the same time, to explain the continuous nonlinear nature of emotion processes, as will be illustrated in section 4.

2.1.2 From Emotional Dispositions to Emotion

As previously referred, emotional dispositions do not represent emotions, instead they act as a background over which more complex emotional phenomena are formed. The kind of emotional phenomena depends on agent's internal structures and processes and on the possible kinds of interaction with the environment. From this point of view emotional phenomena may have multiple forms of expression, as phenomenological observations indicate. To categorize these multiple forms of emotional phenomena Damásio proposes a hierarchy with four main levels of differentiation [9]: (i) innate, automatic survival supporting mechanisms; (ii) pain and pleasure behaviors; (iii) drives and motivations; (iv) emotions proper. Emotions proper progressively differentiate at different levels of complexity into three main emotional categories, *background emotions*, *primary emotions* and *social emotions*. Emotional dispositions can be related to what Damásio considers as *background emotions*.

From an architectural perspective, emotional dispositions play a key role by acting as a common currency (e.g. [34]) conveying emotional dynamics across internal mechanisms and processes at different levels of organization. Among other aspects, emotional

dispositions can be used to regulate agent internal activity [28], feeding mechanisms such as attention focusing and abstraction level control [25], and to modulate autobiographical memories used for adaptation and learning [26]. In the proposed model, it is from the interplay of these different aspects that complex emotions arise, in the sense of *primary emotions* as proposed by Damásio.

3. MODELING EMOTIONAL AGENTS

Although inspired by biophysical analogies, the main aim of the proposed model is to support the development and implementation of emotional agents, independently of their kind or level of complexity. Therefore it is necessary to concretize the base notions of the model in a computationally tractable way.

The first aspect that we need to address is the notion of energy. In thermodynamics, energy is usually defined as the capacity to produce work. In the context of the proposed model, energy can be defined as the capacity of an agent to act or, in a wide sense, to produce change. Considering an agent as a dissipative structure, that change is oriented towards the achievement of motivations, driven by internal potentials and expressed through energy flows.

For the implementation of software agents this thermodynamic level of description is however difficult to use directly due to its level of detail. For that reason we present a more abstract level of representation where these notions are characterized in a way that can act as a bridge between different levels of organization.

3.1 Internal Representational Structures

Energetic potentials can aggregate to form composite potentials. These composite potentials can represent different elements of the internal representational structures of an agent, such as a perception, a memory or an intention. Therefore they are broadly called *cognitive elements*. In this context we are using the term *cognitive* in the sense proposed by Maturana and Varela, as a global property of an agent expressed through the ability of effective action in a given environment [23]. This broader sense of cognition, distinct from the specific sense of symbolic information processing commonly associated to deliberative agent models (e.g. [31]), allows to relate functionally similar concepts at different levels of organization, which is an important aspect of the proposed model.

Cognitive elements play different roles in agent internal activity. Three main roles can be identified: *observations*, *motivators*, and *mediators*. *Observations* result from perception processes, representing the current environmental situation. They can also result from simulated experience [26]. *Motivators* represent intended situations, acting as motivating forces driving agents' behavior. *Mediators* describe the media that support action, forming an interface between internal processing and action.

According to the organizational level, cognitive elements may have different kinds of instantiation. For example, in simple agents motivators may correspond to regulatory potentials, while in deliberative agents motivators may correspond to desires or high-level goals. In the same way, in simple agents mediators may correspond to direct mappings between perception and action, while in deliberative agents planning processes produce sequences of mediators that are translated by action processes into concrete action. In what concerns overall internal structure and processes, simple agents are composed of a fixed number of

cognitive elements and very simple internal processes. Their behavior is directly guided by the dynamics resulting from the achievement potentials and flows [27], leading to basic adaptive behavior such as the *kineses* of some organisms (e.g. bacterial *chemotaxis*) [37]. In more complex agents, internal representations are dynamically formed and changed, constituting an internal model (e.g. [31]) based on which high-level cognitive processing, such as reasoning and decision-making, can occur.

3.2 Cognitive Space

To concretize the flow model of emotion, an approach to agent modeling was developed that follows a signal based metaphor in order to abstract the underlying thermodynamic characterization [29]. A main feature of that approach is the definition of a signal space [33] that enables the description of agent's internal representational structures by means of geometrical concepts, therefore designated *cognitive space*. Conceptually, the cognitive space notion is related to a similar notion used in psychology to describe and categorize mental constructs based on a geometrical n-dimensional space (e.g. [30]) and to the notion of conceptual space proposed by Gärdenfors [11] based on a geometric treatment of concepts and knowledge representation. In particular, Gärdenfors proposes the notion of conceptual space as a way to overcome the opposition between the traditional, symbolic representations and the connectionist, sub-symbolic representations. This same motivation underlies the notion of cognitive space in our model. Relevant distinctions from these approaches exist however, since the proposed notion corresponds to a signal space where not only qualitative but also quantitative information can be described and interaction between the elements represented in the space is possible.

3.3 Modeling Emotional Dynamics

The motivation achievement process, from which emotional dispositions arise, can be described independently of the organizational level at which it occurs by a relation between a current situation, represented by an *observation*, and an intended situation, represented by a *motivator*. The internal activity of an agent is consequently guided by the maximization of the change (flow) that leads to the reduction of the distance between *observations* and *motivators* through the use of *mediators*.

In the cognitive space this process can be described by the movement of cognitive elements, where motivators and observations correspond to specific positions and mediators define directions of movement, as illustrated in figure 2 (φ_1 and φ_2 represent the base dimensional signals that define the space and $\sigma(t)$ represents a cognitive element at some instant $t = \tau$ [29]).

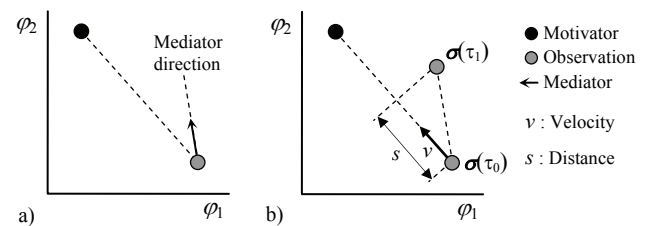


Figure 2. Elements participating in the achievement of a motivator in a two-dimensional cognitive space.

Cognitive elements may have different forms of concretization according to the organizational level, however the cognitive space notion provides a support for uniform representation. In particular, the distance and velocity in cognitive space can be directly related to the base notions of potential and flow associated to the emotional disposition notion, as illustrated in figure 2.b.

In simple agents the achievement potential may directly represent the available energy, or some other kind of resources needed for action. As the complexity of an agent internal organization increases the achievement potential becomes increasingly subjective, in the sense that it depends on internal deliberative processes and internal models of the world. At that level, the achievement potential can be related to the resources that an agent is able to mobilize to cope with the observed situation, that is, the agent internal situation. The difference between observed and internal situation corresponds in cognitive space to a distance. However, that distance may not be directly measurable. What the agent can measure based on past and present observations is the rate of change of those situations, that is, velocity. In this way, the emotional disposition notion is concretized in cognitive space as follows:

$$ED \equiv (\delta P, \delta F) \text{ where } \delta P = v_p \text{ and } \delta F = \frac{dv_F}{dt} \quad (5)$$

where v_p is a velocity that represents the rate of change of the achievement potential and v_F is a velocity that represents the achievement flow.

The dynamics δP and δF can be positive or negative. However, the sign of these dynamics only indicates the direction of the movement (flow). It can be a movement towards a motivator (convergent) or a movement away from a motivator (divergent). Therefore, the overall dynamics δP and δF can be characterized based on the convergent and divergent components as follows:

$$\delta P = \delta P^+ - \delta P^- \text{ and } \delta F = \delta F^+ - \delta F^- \quad (6)$$

3.3.1 Valence and Affective Quality

The convergent and divergent dynamics (flows) convey a hedonic quality, associated to the increase or decrease of agent's well being in relation to the achievement of its motivators. The dynamics δP^+ and δF^+ express improvement in the achievement conditions, corresponding to a positive valence, while δP^- and δF^- express deterioration of the achievement conditions, corresponding to a negative valence.

This valence aspect, pleasant vs. unpleasant or positive vs. negative, constitutes an *affective* quality [24]. The most favorable affective situation occurs when both $\delta P^+ > \delta P^-$ and $\delta F^+ > \delta F^-$. In the emotional disposition plane this situation corresponds to *ED* vectors located in quadrant Q-I (joy/happiness). In the same way, the most unfavorable affective situation occurs when both $\delta P^+ < \delta P^-$ and $\delta F^+ < \delta F^-$, corresponding to *ED* vectors located in quadrant Q-III (apprehension/fear).

3.3.2 Cumulative Emotional Dispositions

The instantaneous tendencies that emotional dispositions represent can be accumulated along the time to characterize prevalent dynamics and long lasting emotional patterns. To describe these patterns we define a cumulative emotional disposition ED_c as:

$$ED_c \equiv (P, F) \quad (7)$$

$$P = \int_t (\gamma_P^+ \delta P^+ - \gamma_P^- \delta P^-) dt \quad (8)$$

$$F = \int_t (\gamma_F^+ \delta F^+ - \gamma_F^- \delta F^-) dt \quad (9)$$

where the sensitivity coefficients γ_P^+ , γ_P^- , γ_F^+ and γ_F^- determine the influence of δP^+ , δP^- , δF^+ and δF^- signals, respectively. In this way it is possible to have a parameterized control of positive and negative influences, namely to simulate aspects such as mood and personality.

The emotional dispositions, instantaneous and cumulative, form a background support for emotional characterization, as will be exemplified next.

4. Example of Emotional Characterization

To illustrate the application of the proposed model we will consider a situation where an agent had to face different achievement conditions to reach an intended situation.

The experimental framework was composed by an environment with obstacles and a moving target. The goal of the agent was to reach the target. The disposition of the obstacles and the movement of the target were controlled during the experiment to produce different achievement conditions. At the beginning (until instant A), the agent succeeded in reducing the distance to the target. However, after a while, the agent encountered obstacles, forcing it to move away from the target (B). After that, the agent succeeded again in reducing the distance to the target (C). Again, after a while, new obstacles appeared, but in this case the target also started to move away (D). Finally, the target changed direction and started to move towards the agent (E).

The agent was able to observe the current external situation and its available resources and to form an internal memory of past situations. From these observations the agent was able to determine the rate of situation change, that is v_F . The agent was also able to estimate the rate of achievement potential change v_p , based on the evolution of its resources and on the estimated distance to the target. The position of the target was represented internally as a motivator. The available resources were used to estimate the maximum distance reachable. The achievement potential was therefore calculated as the difference between the maximum distance reachable (the available potential) and the distance to the motivator (the required potential). Agent's resources to reach the target were limited. From the velocities v_p and v_F emotional dispositions were generated at each time step. The results are shown in figure 3.

