We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

From Signals to Emotions: Applying Emotion Models to HM Affective Interactions

Rita Ciceri and Stefania Balzarotti
Università Cattolica del Sacro Cuore, Milano
Italy

1. Introduction

In 1997 Rosalind Picard defined *Affective Computing* as the «computing that relates to, arises from, or deliberately influences emotions». Since then, research has developed to provide machines and computers with emotional skills similar to those of human users: among others, emotion recognition and expression. Within this approach, emotions are thought to constitute key components to achieve more effective Human Computer Interaction (HCI): «machines may never need all of the emotional skills that people need; however, there is evidence that machines will require at least some of these skills to appear intelligent when interacting with people» (Picard, Vyzas, & Healey, 2001). The core idea is that in the same way as emotion is crucial for human intelligent and rational functioning – as most recent psychology and neuroscience research has pointed out (Bechara & Damasio, 2005) – emotional abilities should be considered also in the development of intelligent computer systems.

Within the broad research area of *Affective Computing*, it is possible to distinguish two main work areas: on one side, *emotion simulation* is aimed at implementing artificial human-like autonomous agents able to interact with the user reproducing human facial and/or vocal emotional expressions, e.g. avatars, robots and ECAs (Breazeal, 2002; de Rosis et al., 2003; Lisetti et al., 2004); on the other side, *emotional decoding* is meant to design interfaces able to recognize the user's emotional responses from real-time capturing and processing of multiple signals (Lisetti et al., 2003). Whatever the research goal is – simulation or recognition – researchers have been confronted with the concept of emotion. Modelling emotions in HCI has revealed a complex challenge and different computational models have been developed to support its complexity and the multimodal richness of human *face-to-face* communication (Bianchi-Berthouze & Lisetti, 2002; Kort, Reilly, & Picard, 2001). Although the psychological study of emotion has brought to the development of different definitions of this construct, in the last years a widespread consensus has grown in psychological literature on the complexity of the emotional process and in particular on its *componential* nature (Scherer, 2005). All major theorists have stressed the need to analyze multiple response systems, such as physiology, behaviour and experience (Levenson, 2003).

In this chapter, we'll mainly focus on *emotional decoding*, i.e. the task to construct new generation interfaces able to sense and to respond to the user's affective feedback (Picard, Vyzas, & Healey, 2001). In particular, the major purpose of the chapter is to provide hints

about the application of psychological emotion models to this research area. We'll first present a brief review of the two main theoretical approaches to the study of emotion: the *categorical* and the *dimensional* one, in the attempt to point out the main advantages and weak points when applied to the HCI. Secondly, we will focus on a dimensional semantic model (Multidimensional Emotional Appraisal Semantic Space, MEAS) which locates emotion in a four axis space, presenting its major features and application possibilities. Finally, we'll present some experimental data to support the use of our model.

2. Theoretical models of emotion: state or process?

As stated by Cowie et al. (2001) «constructing an automatic emotion recognizer depends on a sense of what emotion is. In the context of automatic emotion recognition, understanding the nature of emotion is not an end in itself. It matters mainly because ideas about the nature of emotion shape the way emotional states are described. They imply that certain features and relationships are relevant to describing an emotional state, distinguishing it from others, and determining whether it qualifies as an emotional state at all» (p.35). However, those ideas about what emotion is are not universally shared and different models of emotion are available within psychological literature. The selection and application of a certain model (the way emotion is conceived) to HCI and to the construction of artificial systems is not an irrelevant issue, since it will determine the way emotion is elicited, the way it is measured and the final performance of the system itself. Figure 1 displays three subtasks which can be differentiated within the broader scope of building an emotion detection system: signal capturing, feature extraction and semantic attribution. It should be noted that the selection of a specific model of what emotion is would influence each of these subtasks: which signals are needed, which features are the relevant ones and how they are labeled. In particular, a critical issue in the implementation of interfaces able to manage affective interactions with the user concerns the *semantic attribution*, i.e. the criteria or rules to attribute an emotional meaning to patterns of signals. In our chapter we will mainly focus on this subtask, considering patterns of signals – such as non verbal behavior and physiological responses – which are usually called *subsymbolic* and are thought to belong to the implicit channel of human communication (Cowie et al., 2001).

Figure 1. Three subtasks in the construction of emotion-sensitive interfaces.

A first theoretical approach to emotion is named *categorical* and has its roots in the evolutionary theories (Tomkins, 1962; Ekman, 1972), according to which emotions correspond to discrete and distinct units that are regulated by innate genetic-based mechanisms. Emotions are biologically determined and have evolved to address specific environmental concerns of our ancestors (for example, flight as a consequence of fear in

situations that may be dangerous for the organism). Evolutionary theorists have pointed out the existence of a reduced number of *basic* or *primary* emotions (as for example anger, fear, disgust, happiness, sadness, surprise) each characterized by a specific response pattern, i.e. basic emotions may be clearly differentiated on the basis of their profile of physiological responses and facial expressions, which are universal. What happens when such an approach is applied to the HCI field? First of all, research will focus on primary emotions so that – for example – an automatic emotion recognizer will be taught to learn to detect if the user is happy, sad, angry, disgusted and so on. Second, research will consider stable and distinct response configurations each corresponding to a certain emotional label. To date, the main approach has been the attribution of a fixed emotional label to a specific configuration of signals. For example, the well-known Facial Action Coding System (FACS; Ekman & Friesen, 1978)– or better the EMFACS - involves a stable link between configurations of Action Units and six discrete basic emotions. This approach may be synthesized by the following formula:

$$\text{if } s_{11} + s_{12} + \ldots + s_{1n}, \text{ then } e_1$$

$$\text{if } s_{21} + s_{22} + \ldots + s_{2n}, \text{ then } e_2$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$\text{if } s_{k1} + s_{k2} + \ldots + s_{kn}, \text{ then } e_k$$

where *s* are different signals and *e* the correspondent emotions: for example, if smile, then happiness; if lowered eyebrows, then anger, etc. Although they are useful to clarify the underlying core idea, these examples are obvious simplifications, since the detection and definition of pattern of signals is not so simple and immediate. Moreover, this subtask is complicated by a number of issues, such as the chance to regulate and hide non verbal signals or their intrinsic multiple communicative functions (Kaiser & Wehrle, 2001a), which however we will not consider here.

A second theoretical approach is named *dimensional.* It was started by Wundt (1905) and nowadays it is represented mainly by appraisal emotion theories and other theorists, such as Russell (1980;2003). Appraisal theories have suggested that the elicitation and differentiation of emotions are based on a process of *appraisal* of the situations that affect an individual's needs and goals (Arnold, 1960; Lazarus, 1991; Scherer, 1984). In this view emotions may be differentiated on the basis of *continuous* appraisal dimensions or criteria (and their patterns), according to which individuals evaluate situations and events. For instance, fear is generated by the evaluation of an event as new (novelty), unpleasant (valence) and exceeding the resources of the individual to cope with it (coping). This approach may be synthesized by the schema displayed in figure 2 where emotion (*e*) is the resultant of the intersection between different dimensions (d) whose values are determined by pattern of signals (s). In this view, the component-process model describes emotion as «the dynamic and constantly changing affective tuning of organisms as based on the continuous evaluative monitoring of their social and physical environment» (Scherer, 2005). Thus, emotion is considered as a *process* rather than a state: it is defined as an episode of temporary synchronization of all major subsystems of the organism functioning represented by five components – cognition; physiology; action tendencies; behavioural expression and subjective feeling. What happens when this second approach is applied to the HCI field?
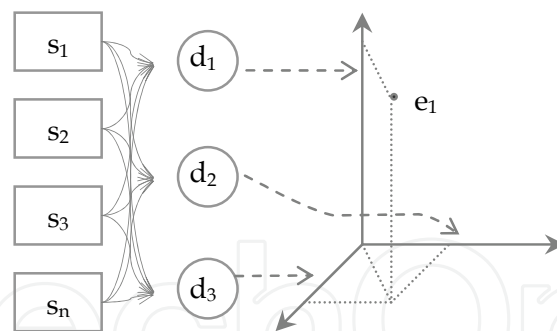
Figure 2. Dimensional models: emotions as resultant of underlying dimensions.

First of all, according to the researchers of this approach, the categorical view seems unable to account for the multiplicity of emotional signals that characterize human spontaneous interactions because it is limited to decode and interpret stereotyped sets of few basic emotions (Kaiser & Wehrle, 2001b). Since emotion is the *dynamic* product of the continuous process of appraisal, the interest is shifted from the detection and labeling of static response configurations to the modeling of the continuous modifications in multiple response systems – as for instance facial movements – with respect to specific dimensions such as novelty, pleasantness, activation and so on. These dimensions vary according to the model adopted: the definition of a reduced number of dimensions through which differentiate all emotional states and hence account for the *structure* of emotion is still a problematic issue. Ben-Ze'ev (2000) has used the expression «subtlety of emotions», to underlie how difficult the task is to account for the entire domain of human emotions through a few number of dimensions and their combinations.

3. Mutual tuning and joint action

Examining the categorical and dimensional approach, we suggested that the latter provides an alternative way to represent emotions rather than conceiving them as more or less discriminative lists of categories. Instead, the dimensional approach should enable researchers to go beyond the attribution of labels to static configurations of signals, since it is aimed at the analysis of patterns of signals in dynamically changing emotional episodes. Of course, this does not mean that prototypical (and universal) expressive configurations of basic emotions are denied; what is questioned is the usefulness of an approach which is limited to their analysis: regarding facial behaviour, Kaiser and Wehrle (2001b) have stated that «since such prototypical full-face expression patterns occur rather rarely in daily interactions, the question of whether and how single facial actions and partial combinations can be interpreted becomes crucial». In a similar way, in their contribute Cowie et al. (2001) have differentiated between full-blown and underlying emotional states, remarking the need for emotion technology to consider underlying emotion: «its priorities are pragmatic, in the sense that it has to deal with emotion as it occurs in real settings. In that context, it would be difficult to justify a policy of ignoring underlying emotion. For instance, it is a serious limitation if an alerting or tutoring system is blind to signs of emotions like boredom or anger until they become full blown» (p.35).

Thus, the risk seems twofold: on one side, it could be implemented an emotion-sensitive system able to recognize prototypical emotional reactions (fear, anger, happiness, etc.) which however occur so rarely within interactions that the system is almost useless; on the other side, the system could be sensitive to macroscopic (full blown) changes rather than to

more subtle response modifications which are crucial in the management of the interaction. Communication theorists have called *attunement* the set of behavioural units through which the agents manage, maintain and coordinate their communicative interactions (Giles et al., 2001; Siegman and Feldstein, 1979). Emotional non verbal signals play a central role within this process: in this sense, Cowie et al. (2001) have talked about *convergence*. Thus, applying the attunement perspective to HCI, the interchanges between user and machine cannot be reduced to or analyzed in terms of linear sequences of expressions and recognitions (Cappella & Pelachaud, 2001), but rather as a coordinated sequence of signals through which they attempt to mutual tuning. Emotional interfaces are thought to support HM interactions within different contexts, as for instance entertainment, tutoring or information retrieval. Therefore, it is reasonable to take into consideration that emotion recognition on the side of the machine is not an end in itself, but it should subserve the intent of improving the effectiveness of the dialog between the artificial system and the human user in the accomplishment of a certain goal. There is a joint (HM) action going on between user and artificial agent and whatever its nature is – having fun in a videogame, e-learning, retrieving information – the emotional reactions of the user arise within this context. In other words, we suggest that the machine should be able to *use* emotional signals to improve its attainment of the operational target (and eventually to adapt the task accordingly): the ability to *tune* emotionally to the user (through the processing of non verbal signals) can support and enhance the joint-action (Figure 3).
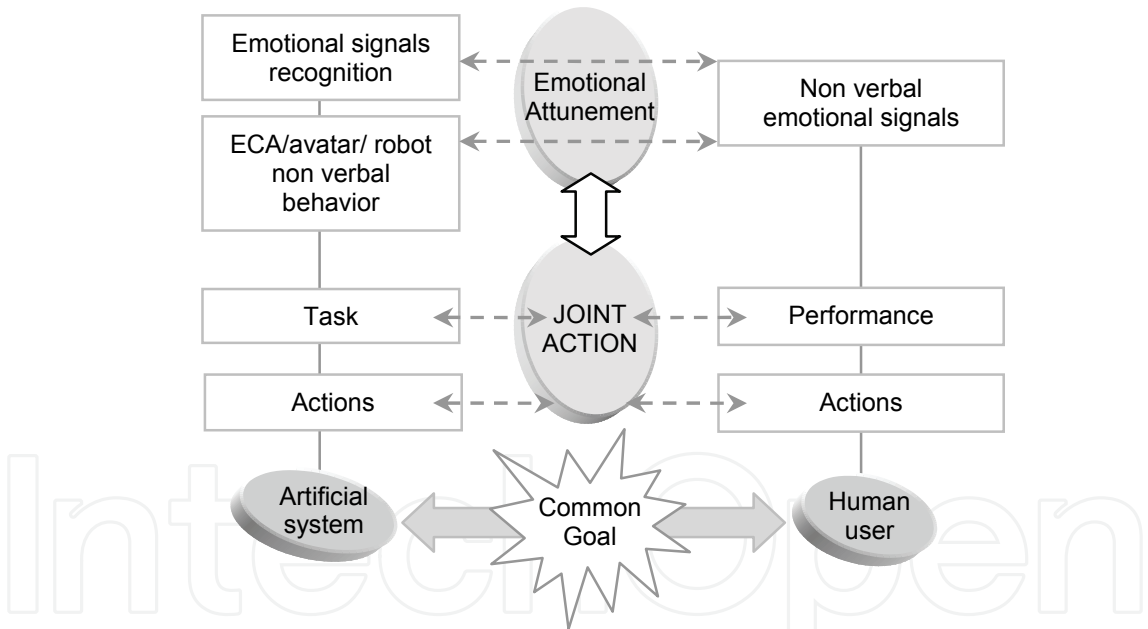

Figure 3. Emotional attunement and management of joint actions.

These two processes are conceived as simultaneous and interdependent: the artificial system and the human user are involved in a specific type of task (usually proposed by the machine) which implies certain types of action and a performance on the side of the user. The recognition of the user's emotional signals enables the machine to respond properly to his/her state through both pertinent non verbal behavior (performed for instance by an ECA) and the modification of the running task.

In this view, the correct recognition of the user's response cannot disregard contextual information such as the type of action or task and the related performance of the human:

«the concrete meaning of a facial expression can only be determined within the whole *temporal* and *situational* context» (Kaiser & Wehrle, 2001b). This also means that more *blended* emotional responses – for instance frustration, boredom, interest, satisfaction, amusement, etc. – are extremely relevant and that the attribution of an emotional meaning to a signal should always consider what's going on within the interaction (a smile could be a signal of embarrassment rather than of happiness).

4. Dimensional models: A proposal

Standing the importance of the attunement process in our modelling of HM interactions, we started working on a model of emotion which could be applied to any situation where an emotion-sensitive system would be likely to operate and could support the subtask of semantic attribution. In the following paragraphs, we'll outline its main features: since the work is still in progress, it is important to note that the model is by no means meant as exhaustive or definitive. We believe that it can provide useful hints for the research in this area, but we will also present the limits and weak points we are still trying to face. First of all, we had clear in mind that the model should entail three main properties: dimensionality, situatedness and multimodality.

Dimensionality. To model the structure of emotion, we chose a dimensional rather than a categorical approach. As reported above, the dimensional approach has accounted for the *temporal dimension* of emotion, i.e. it has underlined the importance of monitoring the constant modifications of the emotional response, which – in each moment – corresponds to the result of the continuous process of appraisal of the situation. It should be clear that if one considers (as we do) the task of recognition of emotional signals as functional to the constant process of emotional attunement in the HM interaction, the need to consider emotion as a process becomes crucial. Thus, we started to conceive the user's emotions as represented by a set of continuous axes, which are normally set to zero but that can move within a certain range of positive and negative values as a result of a change in the status of the user. Given these assumptions, a question remained: how many and which dimensions are necessary to account for the structure of emotion and differentiate between the huge number of emotional responses?

To address this question, we turned to the dimensional models of emotion proposed by psychological literature. A first dimensional model is the *circumplex model* of Russell (1980) which considered two relevant dimensions, *activation* and *valence* – the definition of these two dimensions was derived by the results of various analysis techniques (factor analysis, scaling, etc.) on emotion related words. According to this model, emotions are represented by different points which are distributed within a two-dimensional space forming approximately a circle. Emotions are then organized around two axes: the horizontal one corresponds to valence, i.e. pleasantness vs. unpleasantness of a certain emotional state; the vertical one is defined as activation, i.e. how dynamic the state is (activation vs. calm). In sum, every emotional state can be described as a specific point of intersection between these two axes. The valence and activation dimensions have been used by other psychological models of emotion (Plutchik, 1980) and have also been applied to HCI (Cowie et al., 1999). However, «representations of that kind depend on collapsing the structured, high-dimensional space of possible emotional states into a homogeneous space of two dimensions. There is inevitably loss of information» (Cowie et al., 2001). In other words,

these two dimensions alone do not allow to differentiate between all emotional states: for example, both fear and anger are characterized by high activation and negative valence.

Secondly, we directed our attention to appraisal dimensions and in particular to the Stimulus Evaluation Checks (SECs) proposed by Scherer (2001): novelty, intrinsic pleasantness, goal conduciveness, coping potential, norm/self compatibility. According to this author, all different emotions may be derived from a particular pattern of these appraisal dimensions. More recently, Scherer (2005) has proposed an instrument for the assessment of subjective experience (Geneva Emotional Wheel, GEW) «going beyond a simple valence-arousal space in order to be better able to differentiate qualitatively different states that share the same region in this space» (p.721). The structure given to the emotional categories included in the instrument is based on appraisal dimensions (or SECs) and in particular on goal conduciveness (vs. obstructiveness) and coping potential (low vs. high power), since – according to Scherer – research has demonstrated that these are the dimensions which have the strongest impact on emotion differentiation.

Starting from the theoretical issues (based on empirical findings) presented above, we chose to locate emotion in a four axis space defined by: novelty, valence, coping and arousal. Because of the reference to the appraisal framework, we named our model *Multidimensional Emotional Appraisal Semantic Space* (MEAS, Ciceri & Balzarotti, 2007). Thus, in our model, every emotion can be described by the formula: e_j (x,y,z,k), where x, y, z, and k define the coordinates in a four-dimensional space. Although they are theoretically conceived as continuous, due mainly to practical reasons, the MEAS axes can assume five different discrete intensity values, ranging from -2 to +2. A graphical representation of the model is displayed in Figure 4.
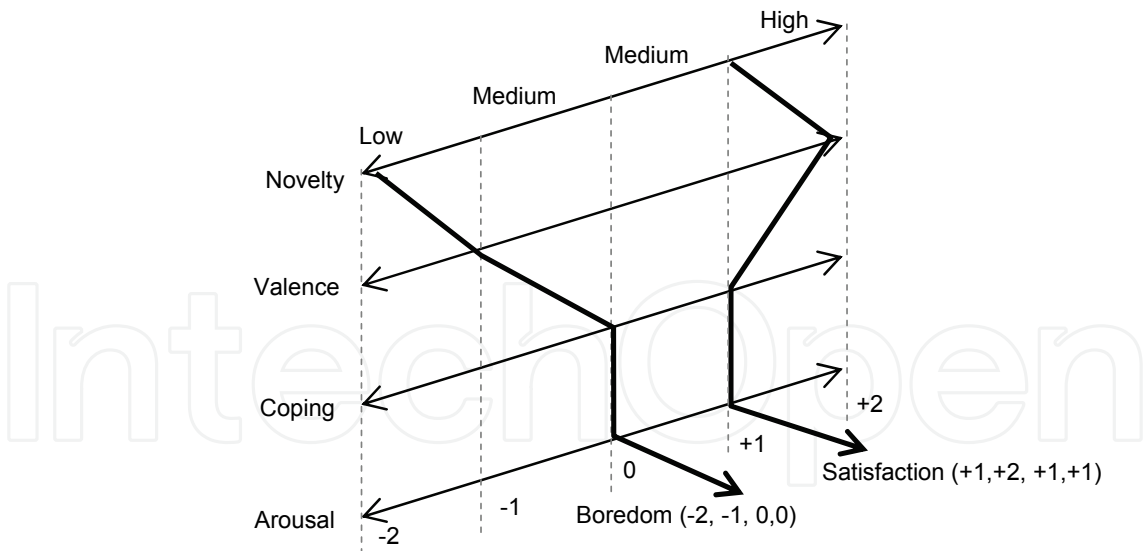

Figure 4. The MEAS model: emotions are conceived as the intersection of four emotional axes (novelty, valence, coping and arousal) which can assume different levels of intensity.

It should be remarked that unlike the other models outlined (mainly concerned with subjective experience), our model was meant as an instrument to attribute emotional meaning to patterns of signals (semantic attribution). In other words, what we tried (and are still trying) to develop is a system of rules able to link signals to underlying emotional

dimensions so that 1) emotional episodes can be differentiated from non emotional behaviour; 2) an automatic emotion recognizer using these rules may be able to monitor the responses of the user continuously, detecting the relevant episodes of changes in the user's state. Thus, as it will be shown concerning *multimodality*, the MEAS system and its scoring are tightly linked to the multimodal analysis of both physiological and non verbal signals, whose modifications provide information for the scoring of the axes.

Besides theoretical assumptions on the structure of emotion, the selection of the four axes (novelty, valence, coping and arousal) was dependent on the possibility to score them on the basis of observable responses. In this sense, the most problematic dimension seems the coping one, since «coping» represents an internal state of the individual, i.e. the evaluation of his/her personal resources to cope with the situation. The issues regarding the scoring of the axes will be further considered in the following paragraphs.

The MEAS model presents a number of limitations: a) first of all, a four axis space is complicated to handle and it seems likely to be a long time before automatic emotion recognizers will incorporate all the subtleties considered in it: nonetheless, we believe that it could provide useful directions to work on, which may be absent in simpler models; b) although it is based on the analysis of previous studies on emotion structure, the selection of the dimensions is still arbitrary and others axes could be more effective at differentiating emotions; c) the definitions of rules linking signals to dimensions is still in progress, as we will show in the next paragraphs.

5. Embodiment: situatedness and action

The application of the concept of attunement to HCI represents an attempt to shift empirical attention from a perspective of uncontextual emotion recognition to its grounding in joint actions between human and artificial agents in which the two participants act in coordination with each other to accomplish goals that are part of their joint activities (Clark, 1996; Brock & Trafton, 1999). In this view, the emotion detection system to be implemented is not an automatic labeler, but rather an artificial agent able to exploit emotional signals to manage the interaction with the user properly tuning to his/her current state. To do this, the system has to be provided with contextual information in some way, since HM Emotional interactions are always *situated* within a certain context, as for instance, videogame-playing, information retrieval services, e-learning, etc. Within HCI, the context may be defined by the particular nature of the interaction itself, i.e. the type of task to accomplish which in turn determines two key elements: the common goal that user and artificial agent share (have fun, give/receive information, tutoring) and the actions performed. The core idea is that it is almost impossible to attribute the accurate emotional meaning to the user's behaviour unless this contextual information is available.

Besides the attunement perspective, the central role attributed to *action* is also justified by the reference to the recently formulated *enactive* view (Varela et al., 1991). Within this psychological approach to cognition, it has been theorized that the mind is embodied, i.e. the body – and the sensorimotor processes – constitutes the medium of the interaction with the world and for this reason it plays a central role in cognition and in adaptation. Since the organism inevitably interacts with the world through his own body, action directly influences perception, categorization and other cognitive processes: «cognition depends upon the kinds of experience that come from having a body with various sensorimotor

capacities […]. By using the term *action* we mean to emphasize once again that sensory and motor processes, perception and action, are fundamentally inseparable in live cognition» (Varela et al., 1991). The concept of embodied (body-mediated) interaction has become increasingly relevant within the domain of HCI. Concerning simulation, for instance, the Embodied Conversational Agents (ECAs) are defined as embodied because they have bodies (similar to the human ones) through which they can communicate multimodally with the user using voice, face, gaze, posture, etc. (Pelachaud & Poggi, 2002). Concerning recognition, interfaces and artificial agents are provided with sets of sensors able to detect a wide range of body signals and human motion. Moreover, besides multimodal interfaces, a new research area has started to work on *enactive interfaces* (ENACTIVE NoE), i.e. interfaces which are thought to interact through perception, action and motion rather than through icons and symbols (e.g. Reactive Robots).

In this perspective, to accomplish the definitions of possible rules of semantic attribution which are part of the MEAS model, two different factors were considered in addition to patterns of (physiological and non verbal) emotional signals. The first factor corresponded to the event, i.e. the type of stimulus/task (what the user is doing): for instance, signals of frustration may be expected within a difficult learning task that exceeds the user's abilities; signals of boredom such as yawning may be expected within repetitive tasks, or within unsuccessful entertainment contexts and so on. According to appraisal theories, emotion is generated by the cognitive evaluation of certain situations: through the systematic manipulation of types of events in a computer game (losing a ship, passing to the next game level), Van Reekum, et al. (2004) have demonstrated the influence of appraisal dimensions (intrinsic pleasantness and goal conduciveness) on physiological reactions and vocal behaviours.

The second factor corresponded to action and included 1) the performance to the task (e.g. doing right/wrong, win/lose, etc.); 2) body movements signaling approach or withdrawal. As to this last component, it has been showed that emotions are linked to different action tendencies, e.g. avoid, approach, interrupt, change strategy, attempt, reject, etc. and involve states of action readiness (Frijda, 2007). Both these elements are used in the scoring of the MEAS dimensions: 1) the performance in the task provides information about two axes, valence and coping; for instance, doing wrong may change from positive to negative the valence of a smile and usually involves low coping; 2) body movements signalling approach and withdrawal provide information about the three axes of novelty, valence and coping; for instance, approaching may signal interest and curiosity towards something new, whereas withdrawing may indicate unpleasantness and low coping (turning one's head), or even low novelty (boredom).

6. Multimodality

Multimodal interfaces are a new-generation class of interfaces designed to recognize two or more combined user input modalities – as for example speech and facial movements – relying on recognition-based technologies increasingly improved by the development of new sensors and input/output devices now becoming available (Oviatt, 2002) together with new algorithms for real-time pattern detection, as in facial tracking (McKenna & Gong, 1996; Cohn, Zlochower, Lier, & Kanade, 1999; Anisetti, Bellandi, Beverina, Damiani, 2005). In this way, they are thought to support more efficient and powerful HM interaction. Multimodality seems an essential property especially for interfaces aimed at automatic

emotion detection, since (as stated at the beginning of the chapter) psychological literature has highlighted the *componential* nature of the emotional process (Scherer, 2005) stressing the need to analyze multiple response systems (Levenson, 2003).

Computation research has based empirical investigation including the combination of multiple modalities collecting various body measures that may capture aspects of an affective state in the attempt to reproduce the multimodal richness of the emotional process. Combining all the features extracted might also provide consistent information in order to increase the accuracy and reduce the error related to the use of single parameters (Picard, Vyzas, & Healey, 2001). In this sense, multimodality is tightly linked to the two subtasks of signal capturing and feature extraction (Picard, Vyzas, & Healey, 2001), i.e. how emotion is measured. In the last decades, these two subtasks have been faced by several studies and research teams (e.g. HUMAINE NoE), which however will not be considered here since they are beyond the purpose of this chapter.

What we are interested in is the translation of multimodal signals into an emotional semantic, i.e. a model to link patterns of signals belonging to different response system to emotional meanings. The MEAS model is designed to receive information from different types of signals (as the ones which multimodal interfaces should handle): a first distinction concerns two macro categories, *physiological* and *communicative* responses.

Many researchers have introduced the interesting and important use of *biosignals* as possible indicators of emotion and arousal (Prendinger et al., 2003; Picard, Vyzas, & Healey, 2001). The Autonomic Nervous System (ANS) is one of the most elicited apparatus from which understand and detect arousal, which is either consciously or mainly unconsciously modified. Since 1953, researchers have proposed some correlations between physiological signals and ANS responses and nowadays it is possible to infer from these signals useful information to understand that something is changing in the emotional condition (Kim et al, 2004; Picard et al, 2001): among the most exploited input signals, skin temperature, electrodermal activity, heart rate, the respiratory rate, etc. However, the chance of differentiating configurations of physiological signals specific for each emotion is still controversial (Cacioppo & Tassinary, 1990) and it may be questioned whether a system relying exclusively on biosignals could be able to discriminate the emotional reactions of the user - or at least reactions which do not fall under the basic emotion categories (Picard, Vyzas, & Healey, 2001). Besides the investigation of specific physiological correlates of emotional categories, other studies have showed that physiological changes are linked to emotional dimensions such as valence and arousal (Cacioppo et al., 2000; Picard, Vyzas, & Healey, 2001).

For these reasons, in our model, physiological signals act as a generic and transverse layer besides the specific channels used to realize the communicative interaction between Human and Machine. In other words, the principal use of the physiological signals we do in our framework is as detectors of the beginning of emotional arousal in the human agent: for this reason, biosignals provide information for the scoring of the *arousal* dimension. Since we are still working on the translation of two types of signals (HR and EDA) into one single arousal axis, we will mainly focus on communicative responses.

Communicative non verbal signals ranging from facial expression, vocal features, gestures, gaze direction, posture have been widely exploited in HCI (for a review about voice and face see Cowie et al., 2001). Psychological research on multimodal communication has suggested that human expressive systems are characterized not only by semantic

redundancy – multiple signals belonging to different systems converge to transmit the same message – but also by semantic interdependence, in which each signal in relation to the others participates in the meaning construction (Ciceri, 2001). For example, one can shake his head to deny and smile at the same time to lower the negativity of the situation or to signal its only partial unpleasantness. We can observe action units like lip corner depressor and lowering eyebrows to signal disgust and at the same time an approaching movement towards the screen to signal simultaneously attention and willingness to better explore. The coordinated use of signals belonging to different response systems enables humans to soften, to stress, to modify the expression of their emotional states to accomplish different communicative intentions. For these reasons, it becomes relevant to focus on the role of different signal systems in emotional expression, as showed in figure 5.
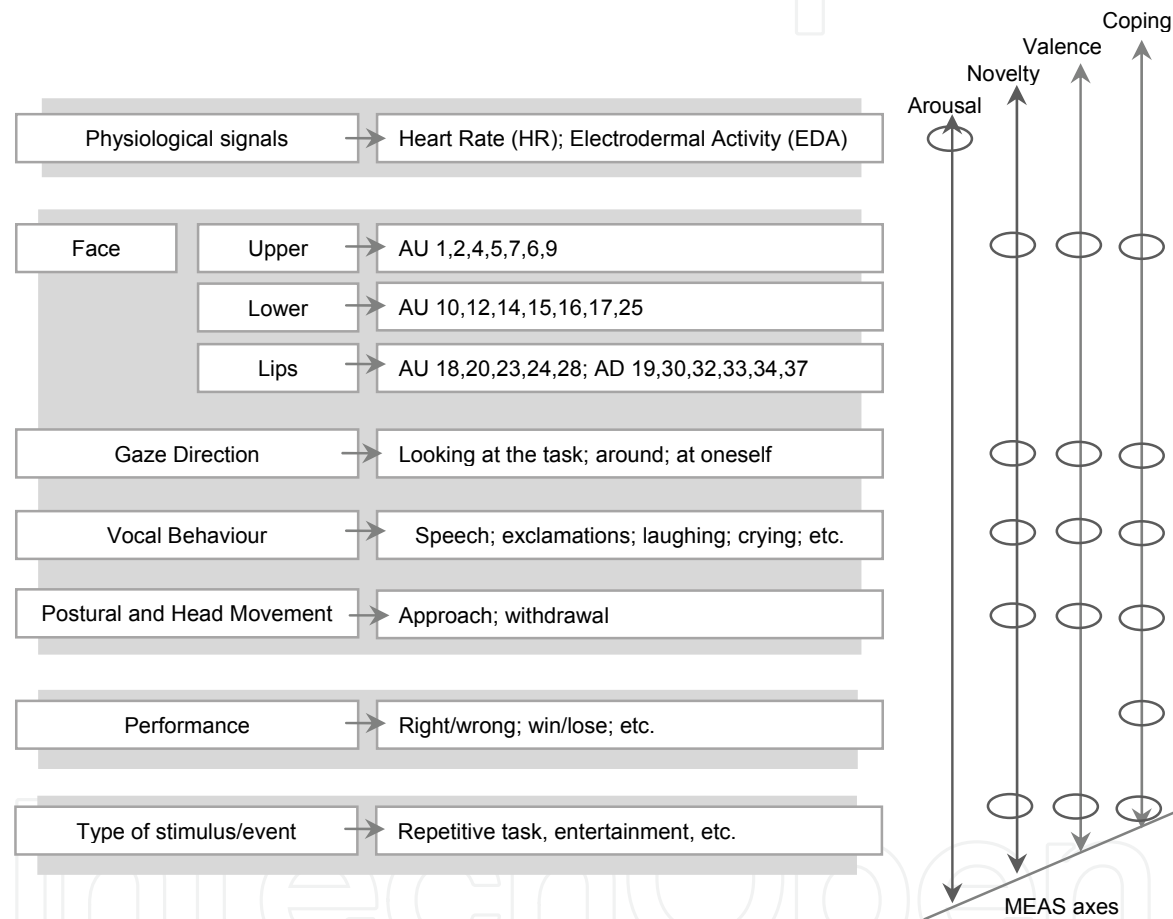

Figure 5. Response systems and emotional axes.

The figure displays the different types of signals considered in our model and the emotional axes which is possible to score on the base of each of them. Physiology provides information to the arousal dimension, whereas communicative signals to novelty, pleasantness and coping. The coping dimension is scored also on the base of the performance to the task. Finally, the type of eliciting event or stimulus may influence novelty, pleasantness and coping.

The procedure of scoring of the four MEAS dimensions is always based on a preceding analysis of the signals belonging to the different response systems taken into account (non verbal and physiological). Since all signals are characterized by large inter-individual

variability, this level of analysis should always take into account a *baseline* level for each subject. Although the procedure to compare a certain response to the baseline is widespread in the analysis of physiological signals, it is less common in behavioural analyses. In this case, the risk is to overestimate (or underestimate) a certain behaviour – since it may be part of the individual's «personal style» of expression.

Regarding non verbal behaviour, the first level of application of the MEAS method consists in a behavioural *frame by frame* micro-analysis (25 fps) of the clip obtained from the video recording of the human subject. This analysis is based on the previous elaboration of a coding grid, i.e. the selection of behavioural units to be extracted. The left side of figure 4 displays the major categories of the Behaviour Coding system (BCS), through which we conduct our analyses (Ciceri, Balzarotti, & Colombo, 2005). The BCS considers four macro-categories:

1) *Facial movements*: the fundamental muscle movements that comprise the Facial Action Coding System (Ekman & Friesen, 1978) were selected. We considered action units relating to upper and lower face and to lip movements (20 AUs and 10 ADs). For each unit, intensity is rated on a 3 point scale (Low, Medium, High);

2) *Gaze direction:* this category considers where the gaze – and thus the attention – of the subject is directed: for instance, the subject may look at the task (mainly at the computer screen), at the keyboard, or he may be distracted and look around, or at oneself (this usually happens when the subject is wired);

3) *Posture and head movement*: behavioural units of approaching/withdrawing are considered (see 5);

4) *vocal behaviour:* since the coding system is applied to a video, we restrict here the analysis to the recording of the use of speech or other kinds of vocalizations. However, more subtle analyses on supra-segmental features may be conducted extracting the vocal stream.

Reliabilities between different coders need to be calculated. Moreover, statistical analysis performed through THEME Software (Magnusson, 2001) is used for the detection of multimodal recurrent pattern analysis (T-pattern). The number of different T-patterns detected in a behavioural stream, their average and/or maximum length (number of event types involved) and their average/maximum level (number of hierarchical levels in a pattern) may be used as measures of complexity or overall synchrony.

Non verbal signals provide information for the scoring of three axes: novelty, pleasantness and coping. At this second level of analysis, the three axes derived from SECs are scored on a 5 point rating scale by a group of three judges to describe the user emotional state. The judges do the scoring on the base of the MEAS rules, which link pattern of signals to positive or negative scores on a certain emotional dimension: examples are showed in Table 1. Novelty is scored in correspondence of behavioural units signalling the individual's evaluation that a change in the pattern of external or internal stimulation occurred (e.g. raising eyebrows, brow lowering, widening eyes, etc.) or that a stimulus is already known, well explored and too expected (moving gaze away, yawning, etc). Valence is scored in correspondence of behavioural units signalling the evaluation that a stimulus event as pleasant (e.g. smile, approach tendencies) or unpleasant (nose wrinkle, avoidance tendencies, etc.). Coping is scored in correspondence of behavioural units signalling the evaluation that the coping potential, i.e. the degree of control over the event is high

(approach tendencies, nodding, right performance, etc.) or low (withdrawal, closing eyes or moving gaze away, wrong performance, etc.). The level of intensity attributed to the emotional axes depends on both the level of intensity (High, Medium, Low) and the number of the behavioural units activated. Inter-judge agreement is calculated.

At present, we have available a group of rules which we derived from indications found in literature (Kaiser & Wehrle, 2001a; Wehrle et al., 2000; Wallbott, 1998) and from the results of the previously mentioned analysis of patterns and configuration of multimodal signals applied to the MEED database (Ciceri, Balzarotti, Beverina, Manzoni, & Piccini, 2006) as we'll describe in the next paragraph. However, we are working at the development of this initial group to a more extended set of rules. Moreover, although at present the MEAS method requires manual annotations (with the aid of software for the analysis of observational data) and human expert judges who receive a training to perform the coding, in the future the rules may be tested and implemented through connectionist architectures to build up the semantic of an emotion detection interface[1]. Facing the problem of «the definition of appropriate models of the relation between realistic emotions and the coordination of behaviours in several modalities» – though applied to the construction of believable ECAs – Martin et al. (2006) have stated that «regarding the analysis of videos of real-life behaviours, before achieving the long-term goal of fully automatic processing of emotion from low levels (e.g. image processing, motion capture) to related behaviours in different modalities, a manual annotation phase might help to identify the representation levels that are relevant for the perception of complex emotions».

7. MEAS multimodal database

We conclude the chapter presenting the grounding of the MEAS system on the analysis of a multimodal database (Multidimensional Emotion Ecological Database, MEED; Ciceri, Balzarotti, Beverina, Manzoni, & Piccini, 2006): audiovisual corpora have been proposed as an important source of knowledge on multimodal behaviour occurring during real-life complex emotions (Douglas-Cowie et al., 2003).

The MEED database includes the collection of naturally occurring samples of emotions (audio, video and physiological data) elicited by two different kinds of stimuli in experimental situations: a) in a first condition, 36 subjects were video recorded while watching segments extracted from three famous movies; b) in a second experimental setup, 24 subjects were video recorded while playing with five different kinds of computer games especially developed to elicit emotional reactions. In this chapter we will focus on this second condition since it better represents a typical HM interaction and it offers an experimental situation where the subject is involved not just as a simple observer since he has to act and react. The videogames were developed through the manipulation of appraisal dimensions and types of game events in a similar way as in Van Reekum et al (2004).

---

[1] In partnership with the Laboratory of Neural Networks (Laren) directed by prof. Bruno Apolloni, State University, Milano.

Dimensions	Rules (+1, +2)	Rules (-1, -2)
Novelty: behavioural units signalling that a change in the pattern of external or internal stimulation occurred	AU 1, AU2, AU 1+2 AU 1+2+17 AU 1+2+25+26 AU 5, AU 1+5 AU 4, AU 4+7 AU 5+ head backward AU 7 + head forward AU 7+20+23 AU 25, 26,AU 16+25 Head/posture forward	AU 41 AU 43 Look away Head backward Head tilted Posture backward
Valence: behavioural units signalling pleasantness or unpleasantness	AU 12 AU 6+12 AU 6+12+25 AU 6+12+20 AU 6+14+20 Approaching	AU 15 AU 6+4 AU 6+17 AU 9 AU 9+16+25 AU 10 AU 6+7+9 AU 16+20+25 Withdrawal Head turned
Coping: behavioural units signalling that the degree of control over the stimulus is high or low	AU 6+20+24 AU 23+24 AU 6+12+right performance Nodding	AU 2 AU 1+4+10 AU 15 + 17 AU 4+17 AU 5+17 AU 17+22 AU 17+23 +head backward/turned Shaking head Wrong performance

Table 1. Examples of signals used to score the MEAS dimensions.

We started from the analysis of emotional responses to address three questions which lied at the base of the task to define rules linking signals to underlying emotional axes:
1) which behavioural systems are used most frequently during HM interactions and thus which signals are the most relevant to consider (multimodality);
2) which behavioural units are usually associated in the same behavioural patterns (multimodality);
3) how these patterns vary in correspondence to certain types of stimulus/event (embodiment);
4) how behavioural units dynamically change in time (attunement).

A preliminary *frame by frame* micro-analysis (25 fps) was performed as previously explained through the Behaviour Coding System (BCS) by two independent judges who had received a training in the analysis of multimodal behaviour. The analysis was conducted with the aid of The Observer 5.0® software for the analysis of observational data (NOLDUS, The Netherlands). The computerized procedure of analysis allowed the automatic extraction of behavioural indexes such as rate (number of occurrences/total time) and duration which were used as dependent variables to submit to statistical analysis. Moreover, inter-judge reliability was calculated (averaged Cohen's Kappa=.89). Figure 6 displays the mean rates calculated for each behavioural category in correspondence of the five different types of videogames.
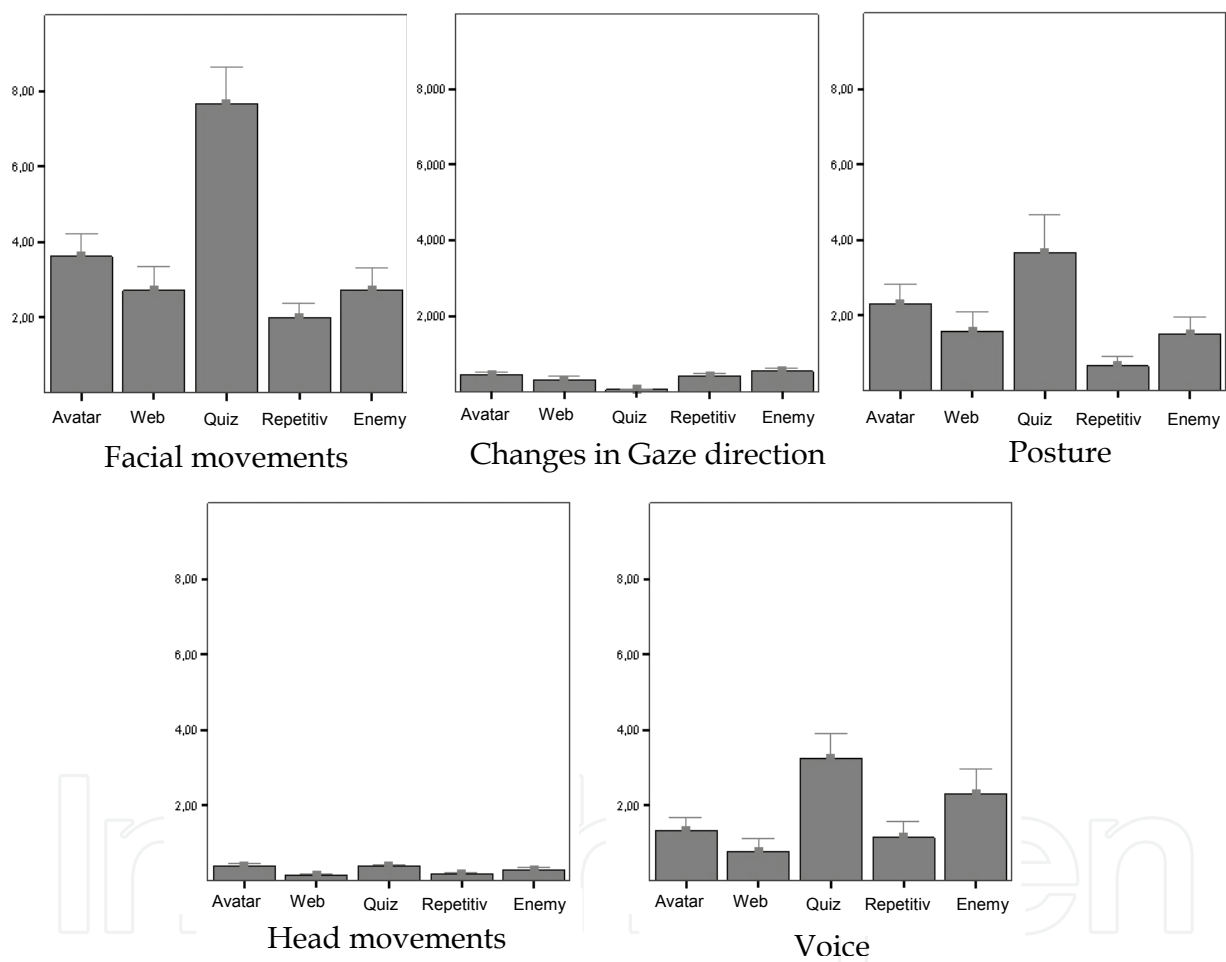
Figure 6: Mean rates calculated for Facial Movements, Gaze Direction, Posture, Head Movements and Vocal Behaviour in correspondence of five different types of video games.

Within the different multimodal signals considered in our analysis, mean rates clearly indicated that some behavioural categories – face, voice and posture – are more frequently exhibited than others. In particular facial movements seemed the most used expression system to signal emotion and thus the most informative. These results were confirmed by a repeated measure ANOVA which showed a significant main effect of the behavioural category ($F=24.807$; $df=4$ $p < .001$). Moreover, Bonferroni adjusted pairwise comparisons confirmed that face differed from all the other expressive systems.

A second significant effect concerned the type of stimulus (F = 26.402; df=4; p < .001) and its interaction with the type of behavioural category (F = 16.873; df=16; p < .001), i.e. the rate of exhibition of a certain behavioural category is influenced by the kind of running task. To better explain the role of the stimulus, we provide hereafter a brief description of the games used in our study. Five different activities were selected in order to simulate different interactive levels on one side and to elicit specific emotional reactions on the other. The first purpose required the manipulation of both the kind of procedural actions on the side of the player and the interactive structure of the game. The second goal involved the manipulation of game events through underlying appraisal dimensions (Van Reekum et al., 2004).

1) In the first activity (avatar), participants interacted with and listened to an avatar constituted by a computerized Italian speaking voice which guided participants across the different computer activities.

2) The second activity (web) consisted in the exploration of a university web site: in particular, some pages regarding university courses were selected as assumed to be emotionally neutral. As to the expected interaction structure, we supposed a low interaction level as participants were confronted with a non finalized task simply requiring to explore a controlled number of virtual pages.

3) The third activity (quiz) was constituted by a quiz game: fifteen questions of general culture were presented to subjects who had to select the right answer among four alternatives.

4) The fourth activity (repetitive) was a boring game: in this game, subjects moved a rabbit character on the screen and had to collect a large number of carrots (50). Carrots appeared one by one and always in the same positions, hence creating a repetitive task.

5) The last activity corresponded to a classical video game: subjects controlled a rabbit character that had to collect carrots while avoiding an enemy. The game presented four different levels of difficulty. The players won points for every carrot collected and every level successfully completed. From the possible events in the game that may elicit particular patterns of appraisal, three different types of events were selected: losing a life (being captured by the enemy or hitting an obstacle), bonus appearing and passing successfully to the next game level. The first type of event is obstructive to the goal of winning the game, while the latter conducive in the pursuit of gaining points.

The games were structured to provide participants with different kind of interactions: for instance, the quiz game had a very quick question/answer structure, hence very similar to a conversational exchange; in the second and third game participants moved and controlled a rabbit and had to act «as if» they were this character to reach specific goals (e.g. gain carrots, avoid the enemy). In the third game, five different interactive episodes (positive vs. negative bonus, invisibility bonus, poisoned carrot, new level) were added where the computer used verbal messages to interact with the participant and this active intervention was aimed at eliciting the subject's response. The whole session lasted about 30 minutes.

Our experimental results showed that the user's communicative responses changed with respect to the interactivity level of the tasks and the related emotional events. More in detail, it is possible to observe that high interactive tasks (e.g. the quiz game, the avatar interaction and the enemy game) elicited more communicative signals, not only when compared to the web control condition, but also with the repetitive-boring task. In particular, Bonferroni adjusted pairwise comparisons showed that the quiz game differed from all the other activities, eliciting the highest number of responses. Therefore, the evaluation of the emotional experience requires the elaboration of situated rules closely linked to the ongoing task and the subject's performance, which can influence the relevance of an event: think, for

example, how different can be a wrong answer during a job interview and a wrong answer during a game.

Secondly, THEME Software 5.0® was used for the detection of multimodal recurrent pattern analysis (minimum number of occurrences=15; α=.0005). The principal aim of the Theme software is to provide aid in discovering and understanding the structure of behavioural streams (Magnusson, 2001).

Thus, whereas the previous statistical analysis considered the mean rate of behavioural categories, this analysis was aimed at examining the *temporal structure* and the associations between specific classes of behavioural units. For instance, figure 7 shows the temporal occurring of all the behavioural units scored for one subject: time is represented by the x-axis, whereas the number of scored events is displayed by the y-axis. Behavioural units are concentrated in the initial part of the HM interaction (avatar and quiz game), whereas few responses characterize the central section (web and repetitive task). At present, we are working on this kind of temporal analysis of behavioural responses, on the detection of the patterns of association between different units and on the possibility to differentiate a reduced number of emotional episodes though the application of the MEAS system (not all behaviour is necessary emotional).
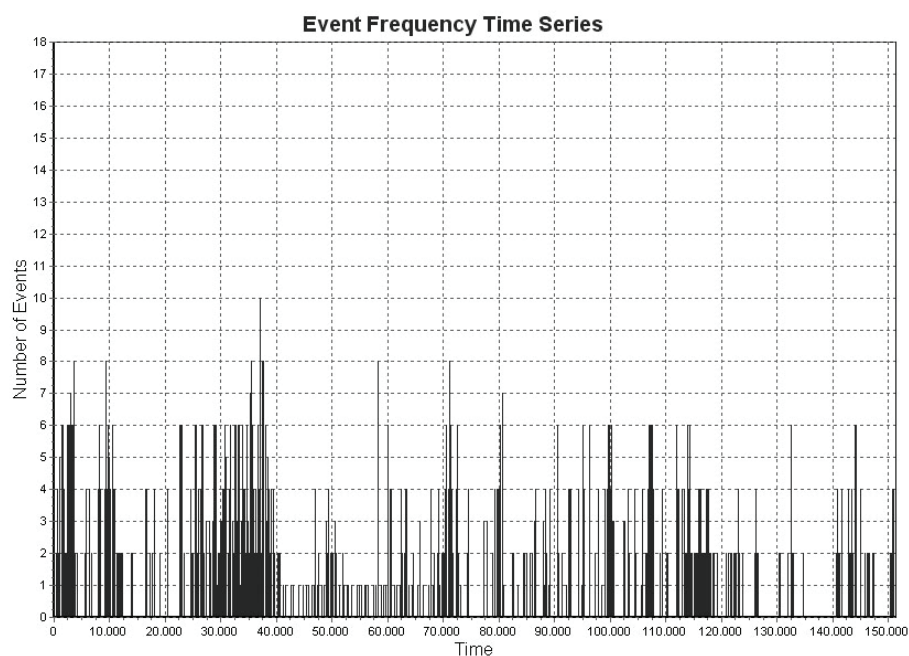

Figure 7. Number of behavioural units scored and temporal dimension

THEME results showed a high number of non random temporal patterns in the event time series of behavioural units. Thus, the detected patterns indicated that non verbal behaviour during computer interaction is highly temporally structured. A qualitative analysis distinguished two different kinds of patterns: patterns connecting behavioural units belonging to the same category (e.g. facial movements) and patterns connecting units belonging to different categories (e.g. facial and vocal behaviour), indicating a multimodal organization and structure. Figure 8 shows one of the most frequent modal patterns (occurrences=17; length=8; duration =5116 frames), combining AU 15 and AU 17 and a multimodal one (occurrences=15; length=6; duration =15287) where vocal behaviour (speech) and facial movements AU 15+17 (lip corner depressor and chin raise) are connected.
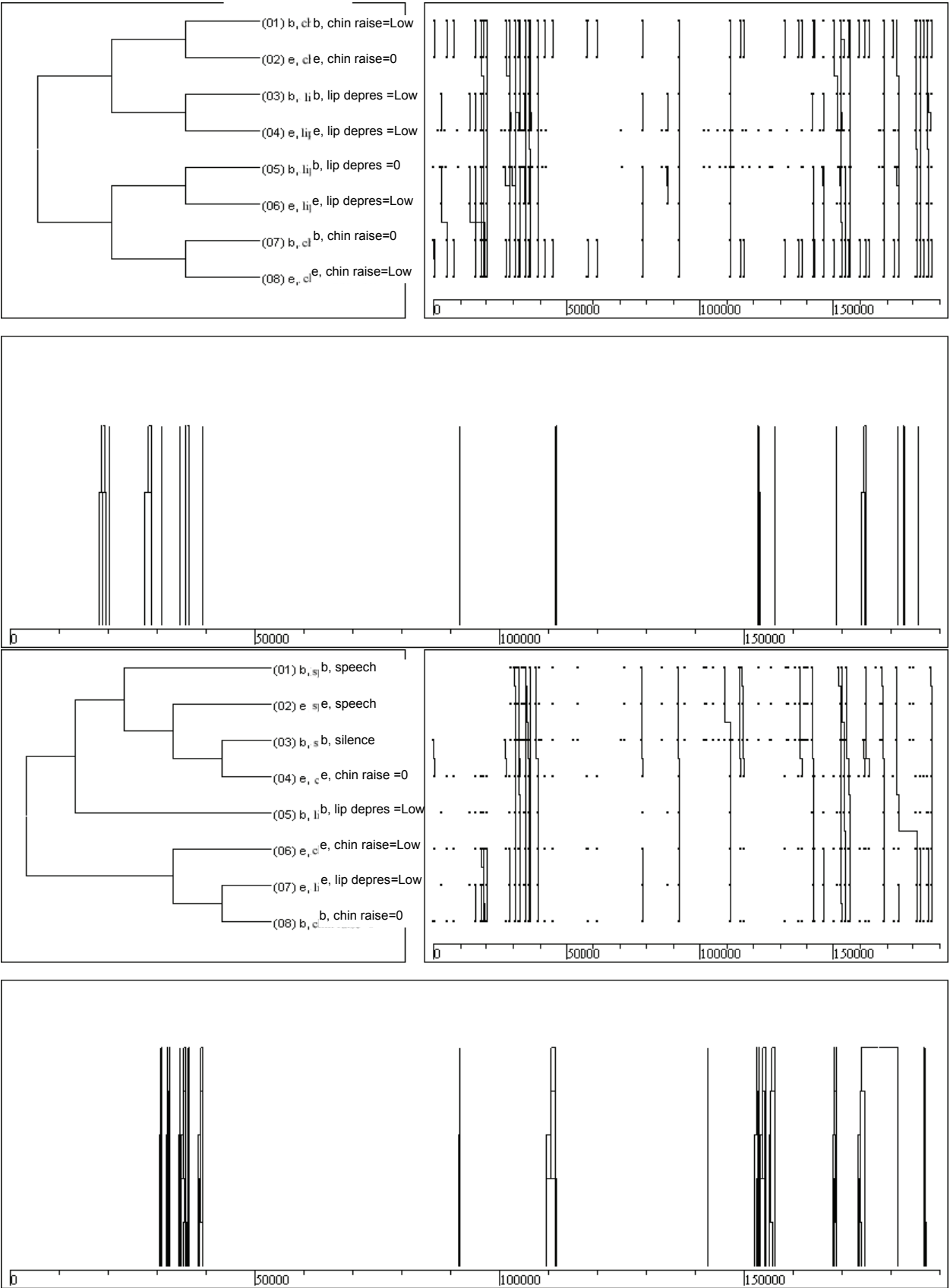
Figure 8. Statistical analysis of recurrent behavioural patterns.

Three diagrams are presented: 1) the pattern tree graph shows the event types composing the pattern (signalling begin or end) listed in the order in which they occur, while the pattern connections are on the left of the event list showing how many hierarchical levels are involved; 2) the connection graph deals with frequency and real-time distribution of events in the pattern: dots represent event occurrences and the lines connecting the dots represent pattern occurrences; 3) the instance graph provides information about the real-time structure of the pattern.

Finally, starting from the multimodal analysis explained above, the three axes of the MEAS systems (the fourth dimension, i.e. arousal, concerns physiological data which are not taken into account here) were scored by two judges who had received a previous training. Inter-judge reliability was calculated (Cohen's Kappa=.79). Figure 9 displays the durations (% interval) for each emotional dimension in correspondence of five different video activities (i.e. the percentage of time during which the dimensions were «active»).
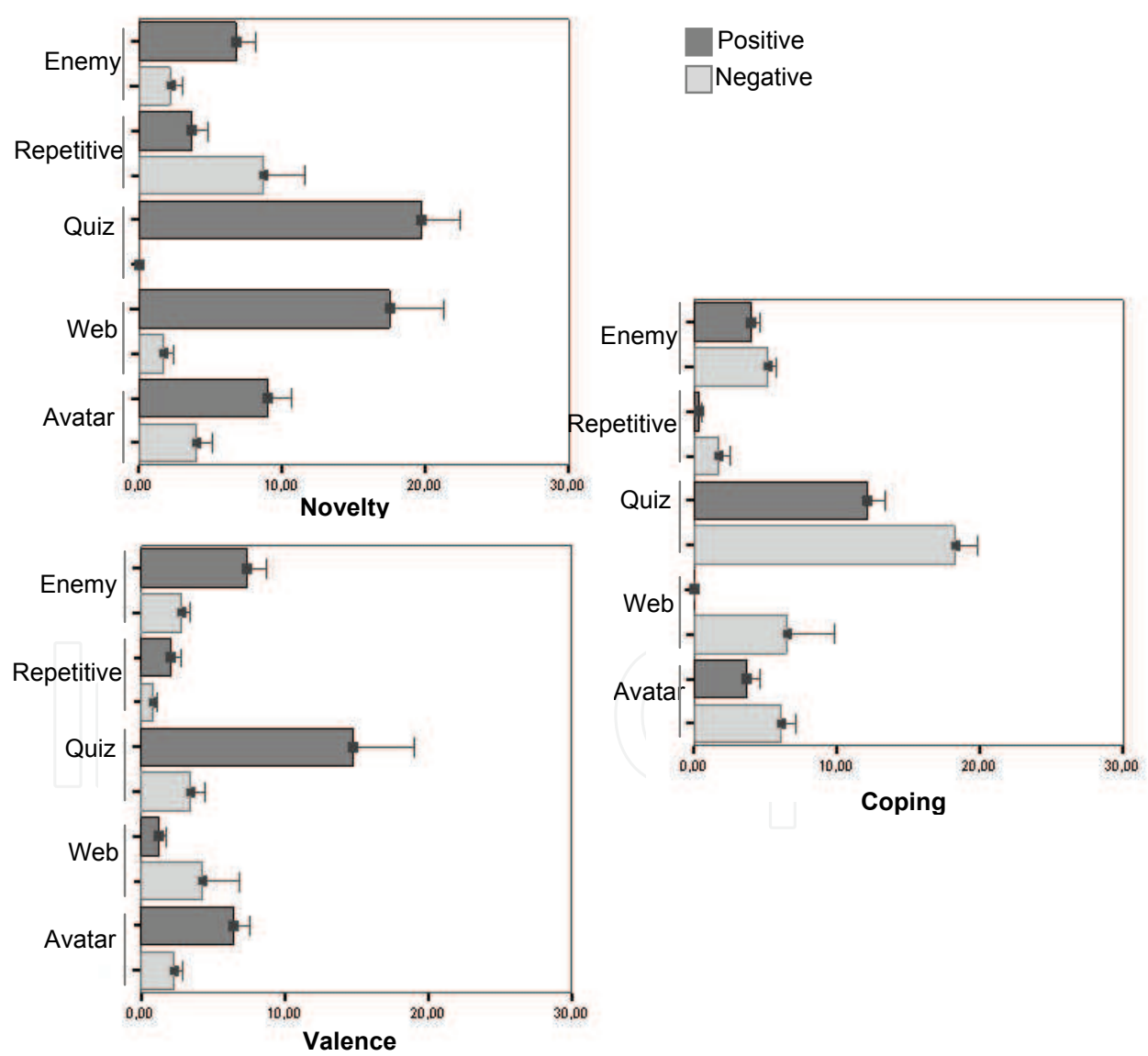

Figure 9. MEAS emotional axes: durations (% interval) for positive (+1, +2) and negative (-1, -2) scores.

First, our results showed that all three emotional axes were active for less than the 30% of the total duration of the activity, whereas they were set to zero for the remaining time: this means that through the application of the MEAS coding system a reduced number of emotional episodes was selected. Thus, emotion is a dynamic and continuous process which characterizes the whole HM interaction, yet not continually. In other words, emotional episodes arise all along the computer interaction without covering its entire duration but being restricted to limited intervals of time (Scherer, 2005). In particular, our results showed that emotional reactions occurred mostly within high interactive activities (for instance, the quiz game) and in response to manipulated eliciting events (Van Reekum et al., 2004).

Second, performing a repeated measure ANOVA, a main effect concerning the type of dimension was found (F=8.685; df=2; p<.01): in particular, the novelty dimension – which is linked to emotions such as interest, surprise, curiosity, boredom, etc. – was the most scored and differed from both coping and valence. Thus, although all dimensions were informative, novelty seemed to have the largest weight (this also because it was often simultaneously active when coping and valence were scored). These results seem consistent with Scherer's Stimulus Evaluation Checks (SECs, 1984), according to which novelty is the first appraisal dimension activated within the emotional process.

Third, positive scores recorded significantly longer durations than negative ones (F=12.527; df=1; p<.01). Both positive and negative scores provided information about the subject's emotional response: for instance, high levels of novelty characterized all the activities, but the quiz game, the web exploration, the interaction with the avatar and the final videogame were characterized by high durations of positive scores (i.e. surprise, interest, etc.), whereas the repetitive task totalized a high duration of the negative ones (i.e. boredom).

Finally, results showed a main effect of game (F=38.461; df=4; p<.001): in particular, the quiz game was the activity where the all the axes were most active and differed significantly from all the other activities. This result highlighted once again the role of the ongoing task in eliciting specific emotional reactions, as already explained concerning the analysis of multimodal behaviour. In sum, our results showed a global coherence between the manipulated events on one side and the emotional reactions (behavioural units and axes) on the other, thus supporting the use of the MEAS scoring procedure.

We conclude this section by presenting an example of synchronic analysis: figure 10 displays a video-segment (44 sec) corresponding to the first set of five questions of the quiz game. We decided to extract the example from the quiz game since previous analyses had demonstrated that this activity obtained the highest number of responses. The figure presents: a) the image sequence extracted from the video of one of our subjects and the related analysis of the main multimodal behavioural units exhibited; b) the scoring of the MEAS axes; c) the task events and the performance of the subject. The synchronic analysis shows that each task event, i.e. the appearance of a new question was followed by the activation of behavioural units concerning novelty, as for instance eyebrow raising and head moving towards the screen and successively by units signalling a low coping potential, as for instance lip corner depressor and chin raise. Despite the facial movements exhibited before answering, the subject gave the right answer to all five questions and showed positive signals such as a smile: on the base of both the performance (right) and the behavioural response, the coping and valence dimensions were then scored as positive (pleasantness and high coping). Thus, the emotional reactions of the subjects were *tuned* to the contextual elements, such as task events and action performance.
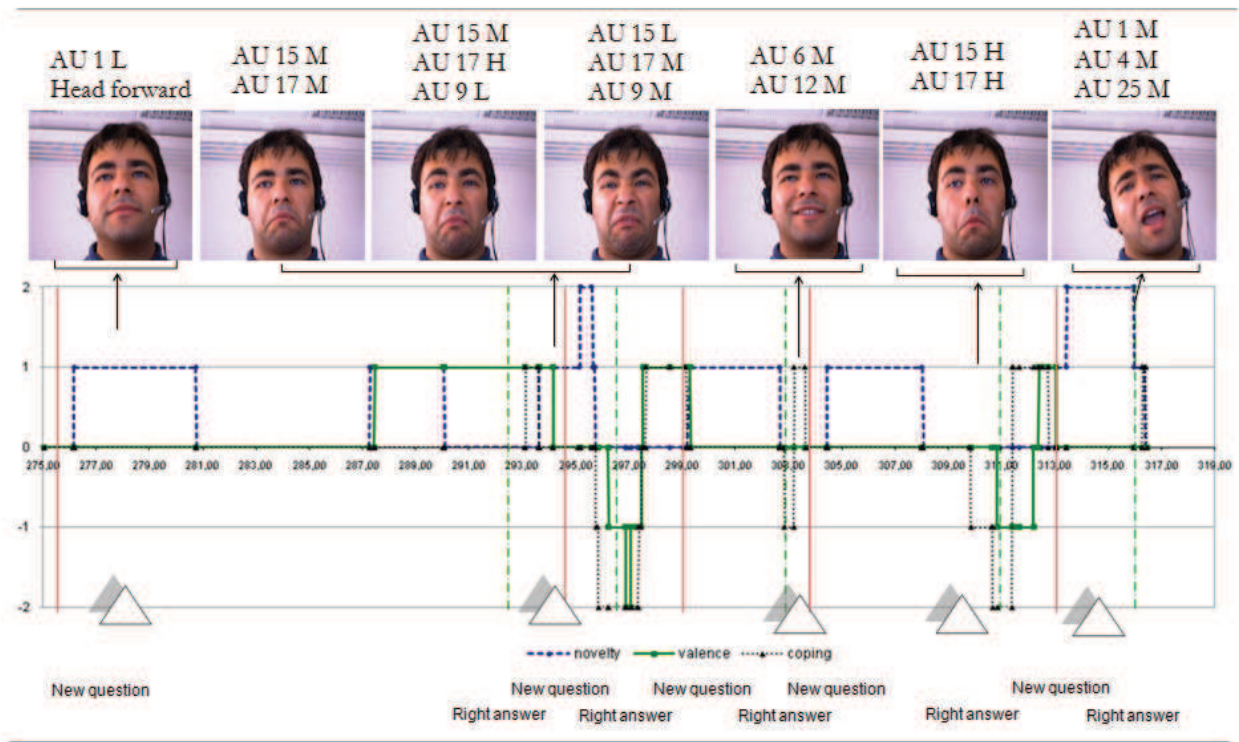
Figure 10. Synchronic analysis: an example of behavioural analysis and MEAS scoring.

8. Conclusions

In this chapter we presented a dimensional semantic model (Multidimensional Emotional Appraisal Semantic Space, MEAS) which locates emotion in a four axes space, in the attempt to detect rules linking pattern of signals to underlying emotional axes, such as novelty, valence, coping and arousal. Although the development of this set of rules is still in progress, this model is aimed at providing hints in the work area of *Affective Computing* concerning emotion decoding, i.e. the implementing and design of automatic emotion recognizers. In particular, within this field of studies, we addressed the subtask of semantic attribution: once the machine is able to capture and to process the multimodal signals (and pattern of signals) exhibited by the human user during the interaction, how is it possible to attribute to them an emotional meaning (or in other words, to label them)?

First of all, it is important to note that despite the use of the term «rule», the MEAS scoring system is not meant as a set of fixed and stable laws to be rigidly applied to every type of HM interaction and context. In fact, this would disclaim one of the principles on which the system is based (embodiment) and the more general conception of the HM interaction which is here proposed. Concerning the former – as previously explained – the MEAS system is thought as strictly linked to the context, that is to the type of running task and to the actions performed by the human user. Concerning the second, in our view, the machine should use the user's emotional signals to be able to *tune* to his/her emotional state (process of attunement). Moreover, as clearly showed by our data, the users themselves show emotional responses which are highly influenced and congruent with the type of eliciting

stimuli and the way they are appraised. Therefore, the MEAS rules are flexible and may change according to these contextual elements adjusting to them.

Second, the MEAS system is designed to record the continuous modifications of emotional dimensions rather than the number of appearances of certain types of emotion categories, since it is based on a theoretical conception of emotion as a process rather than a state. As stated by Russell (2003): «The ecology of emotional life is not one of long periods of non-emotional "normal" life punctuated by the occasional prototypical emotional episode. [..] Emotional life consists of the continuous fluctuations in core affect, in pervasive perception of affective qualities, and in the frequent attribution of core affect to a single Object, all interacting with perceptual, cognitive, and behavior processes. Occasionally, these components form one of the prototypical patterns, just as stars form constellations». Thus, prototypical or basic emotions such as happiness, anger or fear occur occasionally, whereas emotion may be better conceived as a continuous modification of emotional dimensions (as, for instance, valence and arousal which form Russell's core affect).

The adoption of a dimensional model leaves open the issue concerning emotional labels. According to Cowie et al. (2001) «category labels are not a sufficient representation of emotional state, but they are probably necessary». Of course, emotional labels may be attributed to each intersection between the different axes considered (as for instance in figure 4). However, not all intersections correspond to different types of emotions, since they may simply represent a shift in the intensity of a certain emotional response. We won't consider here the translation of our dimensions into labels since we are still working on the preceding operation, i.e. the translation of signals into dimensions. Nonetheless, in our view, besides the usual contraposition between dimensions and categories, it should be pointed out the need to consider emotion as a continuous process, thus working on models actually able to account for its dynamical changing.

Another issue we leave open is the relation between non verbal communicative behavior and physiological signals. We are still working on the possibility to translate different physiological measures, such as heart rate and electrodermal activity[2], into a single dimension which should signal the beginning of the physiological arousal of the individual. Although in our model we do not consider biosignals as informative about the specific nature of the emotional response (which is signaled by the expressive-motor system), they are nonetheless an important source of information about the individual's state of body activation. Moreover, since emotion is thought to produce coordinated modifications in multiple response systems (Scherer, 2001), we are working on the hypothesis that a total absence of behavioural expressions together with a high level of activation may indicate that a process of regulation is going on – see for instance, the regulatory strategy of suppression studied by Gross and Levenson (1993).

Finally, the MEAS system includes the dimension of coping to differentiate emotion. As already stated about the main limits of the model, this dimension was the most problematic one, since it concerned an internal state of the individual, i.e. the evaluation of his/her own resources to cope with a certain event. For this reason, it seemed more difficult to link this dimensions to observable behaviour and the scoring of the coping axis was mainly linked to the action (approach vs. withdrawal) and the performance of the user. Nonetheless, we

---

[2] In partnership with Sensibilab, Politecnico di Lecco.

believe that this dimension may provide useful information to an emotional interface, since it signals the capability of the human user to accomplish a task or not. For instance, if we imagine an e-learning tutoring context, it would be crucial for the virtual tutor to detect when the task proposed to the user is too difficult and to change it accordingly.

9. Acknowledgments

10. References

Anisetti, M.; Bellandi, V.; Damiani, E. & Beverina, F. (2005). Facial identification problem: a tracking based approach, *Proceedings of the First International Conference on Signal-Image Technology and Internet-Based Systems*, IEEE SITIS, Yaoundé, Cameroon.

Arnold, M.B. (1960). *Emotion and Personality, Psychological Aspects*. Columbia University Press, New York.

Bechara, A. & Damasio, A. R. (2005). The somatic maker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, 52, 2, 336-372.

Ben-Ze'ev, A. (2000). *The Subtlety of Emotions,* Massachusetts Institute of Technology Press, Cambridge, MA.

Breazeal, C. (2002). *Designing Social Robots*, MIT Press, Cambridge, MA.

Bianchi-Berthouze, N. & Lisetti, C. (2002). Modelling multimodal expression of user's affective subjective experience. *User Modelling and User-Adapted Interaction,* 12, 1, 49-84.

Brock, D. & Trafton, J.G. (1999). Cognitive representation of common ground in user interfaces. In: *User modeling: Proceedings of the seventh international conference*, Kay, J. (Ed.), Springer-Wien, New York, NY.

Cacioppo, J.T. & Tassinary, L.G. (1990). Interring Psychological Significance from Physiological Signals. *American Psychologist*, 45, 16-28.

Cacioppo, J.T.; Tassinary, L.G., Berntson, G.G. (2000). *Handbook of Psychophysiology*, 2nd Edition, Cambridge University Press, New York.

Cappella, N.J. & Pelachaud, C. (2001). Rules for responsive robots: Using human interactions to build virtual interactions. In: *Stability and Change in Relationships*, Vangelisti, A.L.; Reis, H.T. & Fitzpatric M.A., (Eds.), Cambridge University Press, New York.

Ciceri, M.R. (Ed), (2001). *Comunicare il Pensiero*, Omega Edizioni, Torino.

Ciceri, R.; Balzarotti, S. & Colombo, P. (2005). Analysis of the human physiological responses and multimodal emotional signals to an interactive computer, *Proceedings of AISB 2005 Annual Conference, Agents that want and like: emotional and motivational roots of cognition and action*, 12-15 April 2005, Hatfield.

Ciceri R.; Balzarotti, S.; Beverina, F.; Manzoni, F. & Piccini, L. (2006). MEED: The challenge towards a Multimodal Emotional Ecological Database, *Proceedings of LREC 2006,*

Multimodal Corpora: From Multimodal Behavior Theories To Usable Models, May 2006, Genova, Italy.

Ciceri, R. & Balzarotti, S. (2007). MEAS: Multidimensional Emotional Appraisal Semantic space. In: *Knowledge-based Intelligent Information and Engineering Systems. Lecture Notes in Computer Science*, (Vol. 4693/2007, pp. 395-402), Springer Verlag, Berlin/Heidelberg.

Clark, H. H. (1996). Arranging to do things with others. In: *Conference Companion of the Conference on Human Factors in Computing Systems – CHI'96*, Association for Computing Machinery.

Cohn, J.F.; Zlochower, A.J.; Lien J. & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS Coding. *Psychophysiology*, 36, 35-43.

Cowie, R.; Douglas-Cowie, E.; Apolloni, B.; Romano, A. & Fellenz, W. (1999). What a neural net needs to know about emotion words. In: *Computational Intelligence and Applications*, Mastorakis, N. (Ed.), (pp. 109-114), World Scientific & Engineering Society Press.

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

De Rosis, F.; Pelachaud, C.; Poggi, I.; Carofiglio, V. & De Carolis, B. (2003). From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, 59, 1, 81-118

Douglas-Cowie, E.; Campbell, N.; Cowie, R. & Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech Communication*, 40, 33-60.

Ekman, P. (1972). Universals and Cultural Differences in Facial Expression of Emotion. In: *Nebraska Symposium on Motivation*, Cole, J.R. (Ed.), (pp. 207–83), University of Nebraska Press, Lincoln, Vol. IXX.

Ekman, P. & Friesen, W. (1978). *Facial Action Coding System (FACS): A technique for the measurement of facial movement*, Consulting Psychology Press.

Frijda, N.H. (2007). *The laws of emotion*. Erlbaum, Mahwah.

Giles, H.; Shepard, C.A. & Le Poire, B.A. (2001). Communication Accommodation Theory. In: *The new handbook of language and social psychology*, Robinson, W.P. & Giles, H. (Eds.), (pp. 33-56), Wiley, Chichester, UK.

Gross, J. J. & Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behaviour. *Journal of Personality and Social Psychology, 64*, 970-986.

Kaiser, S. & Wehrle, T. (2001a) Facial expressions as indicators of appraisal processes. In: *Appraisal processes in emotions: Theory, methods, research*, Scherer, K. R.; Schorr, A. & Johnstone, T. (Eds.), (pp. 285-300), Oxford University Press, New York.

Kaiser, S. & Wehrle, T. (2001b). The role of facial expression in intra-individual and inter-individual emotion regulation. In: *Emotional and Intelligent II: The Tangled Knot of Cognition*, Cañamero, D. (Ed.), (pp. 61-66), Papers from the 2001 AAAI Fall Symposium, Technical Report FS-01-02, AAAI Press, Menlo Park, CA.

Kim K. H.; Bang S. W. & Kim S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, 42, 419-427.

Kort, B.; Reilly R. & Picard, R.W. (2001). An affective model of interplay between emotions and learning: reengineering educational pedagogy—building a learning companion, *International Conference on Advanced Learning Technologies,* ICALT 2001, Madison, WI.

Lazarus, R.S. (1991). *Emotion and Adaptation,* Oxford University Press, New York.

Levenson, R. W. (2003). Blood, sweat, and fears: The autonomic architecture of emotion. In: *Emotions inside out*, Ekman, P.; Campos, J. J.; Davidson, R. J. & de Waal, F. B. M. (Eds.), (pp. 348–366), The New York Academy of Sciences, New York.

Lisetti, C. L.; Nasoz, F.; LeRouge, C.; Ozyer, O. & Alverez, K. (2003). Intelligent affective interfaces: A patient-modelling assessment for tele-home health care. *International Journal of Human-Computer Studies*, 59, 245-255.

Lisetti, C. L.; Brown, S.M.; Alverez, K. & Marpaung, A. H. (2004). A social informatics approach to human robot interaction with a service social robot, *IEEE Transactions on Systems, Man and Cybernetics*, Special Issue on Human-Robot Interaction, 195-209.

Magnusson, M. S. (2001). Discovering hidden time patterns in behavior: T-Patterns and their detection, behavior research method. *Instruments and Computers*, 32, 1, 93-110.

Martin, J.C.; Buisine, S.; Pitel, G. & Bernsen, N.O. (2006). Fusion of Children's Speech and 2D Gestures when Conversing with 3D Characters. *Signal Processing Journal*, Special issue on multimodal interfaces, 86, 12, 3596-3624.

McKenna, S. J. & Gong, S. (1996). Tracking faces, *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition*, pp. 271-276.

Oviatt, S.L. (2002). Multimodal interfaces. In: *Handbook of Human-Computer Interaction*, Jacko, J. & Sears, A. (Eds), Lawrence Erlbaum, New Jersey.

Pelachaud, C. & Poggi, I. (2002). Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation,* 13, 301-312.

Picard, R.W. (1997). *Affective Computing*, The MIT Press, Cambridge, MA.

Picard, R.W. & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transaction on Pattern Analysis and Machine Intelligence*, 23, 10, 1175-1193.

Picard, R. W.; Vyzas, E. & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions Pattern Analysis and Machine Intelligence,* 23, 10, 1185-1191.

Plutchik, R. (1980). A psychoevolutionary theory of emotions. *Social Science Information*, 21, 529–553.

Prendinger, H.; Mayer, S.; Mori, J. & Ishizuka, M. (2003). Using bio-signals to measure and reflect the impact of character-based interfaces, Workshop on Assessing and Adapting to User Attitudes and Affect: Why, When and How?, in conj. with User Modeling (UM-03), pp. 39-44, Johnstown, USA.

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

Russell, J.A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, 110, 1, 145-172.

Scherer, K.R. (1984) On the nature and function of emotion: A component process approach. In: *Approaches to Emotion*, Scherer, K.R. & P. Ekman (Eds.), (pp. 293–317), Erlbaum, Hillsdale, NJ.

Scherer, K.R. (2001). Appraisal Considered as a Process of Multi-Level Sequential Checking. In: *Appraisal Processes in Emotion: Theory, Methods, Research*, Scherer, K.R.; Schorr, A. & Johnstone, T. (Eds.), (pp. 92–120), Oxford University Press, New York.

Scherer, K.R. (2005). What are emotions? And how can they be measured? *Trends and developments: Research on emotions,*, 44, 4, 695-729.

Siegman, A.W. & Feldstein, S. (1979). *Of Speech and Time,* Erlbaum, Hillsdale.

Tomkins, S.S. (1962). *Affect, Imagery, Consciousness: Vol. I, The Positive Affects*, Springer, New York.

Van Reekum, C.M.; Johnstone, T.; Banse, R.; Etter, A.; Wehrle, T. & Scherer, K.R. (2004). Psycho physiological responses to appraisal dimensions in a computer game. *Cognition and Emotion*, 18, 5, 663-688.

Varela, F.; Thompson, E. & Rosch, E. (1991). *The Embodied Mind*. MIT Press, Cambridge, MA.

Wallbot, H.G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, 28, 879-896.

Wehrle, T.; Kaiser, S.; Schmidt, S. & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, 78, 1, 105-119.

Wundt, W. (1905). Grundzüge der physiologischen Psychologie, Engelmann, Leipzig.

Affective Computing

Edited by Jimmy Or

ISBN 978-3-902613-23-3

Hard cover, 284 pages

Publisher I-Tech Education and Publishing

Published online 01, May, 2008

Published in print edition May, 2008

This book provides an overview of state of the art research in Affective Computing. It presents new ideas, original results and practical experiences in this increasingly important research field. The book consists of 23 chapters categorized into four sections. Since one of the most important means of human communication is facial expression, the first section of this book (Chapters 1 to 7) presents a research on synthesis and recognition of facial expressions. Given that we not only use the face but also body movements to express ourselves, in the second section (Chapters 8 to 11) we present a research on perception and generation of emotional expressions by using full-body motions. The third section of the book (Chapters 12 to 16) presents computational models on emotion, as well as findings from neuroscience research. In the last section of the book (Chapters 17 to 22) we present applications related to affective computing.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Rita Ciceri and Stefania Balzarotti (2008). From Signals to Emotions: Applying Emotion Models to HM Affective Interactions, Affective Computing, Jimmy Or (Ed.), ISBN: 978-3-902613-23-3, InTech, Available from: http://www.intechopen.com/books/affective_computing/from_signals_to_emotions__applying_emotion_models_to_hm_affective_interactions

INTECH

open science | open minds