PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2009) **364**, 3539–3548
doi:10.1098/rstb.2009.0186

Modelling multimodal expression of emotion in a virtual agent

Catherine Pelachaud*

CNRS, LTCI, Telecom ParisTech, Paris 75014, France

Over the past few years we have been developing an expressive embodied conversational agent system. In particular, we have developed a model of multimodal behaviours that includes dynamism and complex facial expressions. The first feature refers to the qualitative execution of behaviours. Our model is based on perceptual studies and encompasses several parameters that modulate multimodal behaviours. The second feature, the model of complex expressions, follows a componential approach where a new expression is obtained by combining facial areas of other expressions. Lately we have been working on adding temporal dynamism to expressions. So far they have been designed statically, typically at their apex. Only full-blown expressions could be modelled. To overcome this limitation, we have defined a representation scheme that describes the temporal evolution of the expression of an emotion. It is no longer represented by a static definition but by a temporally ordered sequence of multimodal signals.

Keywords: embodied conversational agent; non-verbal behaviour; emotion

1. INTRODUCTION

Embodied conversational agents (ECAs) are virtual entities endowed with conversational capabilities; i.e. they can communicate with virtual or real interlocutors using verbal and non-verbal means. By detecting and interpreting signals coming from their interlocutors, they perceive what is being said; they plan what and how to answer their interlocutors. They can do so through human-like communicative modalities, namely the voice, face, gaze, gesture and body movements. To be a full conversational partner, the ECAs ought to generate behaviours that can be decoded by their human interlocutors.

Faces can display a large variety of expressions, linked to emotional state, attitude, belief or even intentions. They can be used to display complex expressions arising from the blend of emotions such as masking one emotional state by another one. Eyebrow shape and head movement are also tightly linked to the intonational structure of the speech. Smile, gaze and other facial expressions can be social signals; their emission is related to our relationship with our interlocutors, our role or even the context of the conversation.

Several attempts have been made to create ECAs able to display communicative facial expressions. The models do not cover the full spectrum of facial expressions but still quite a large one. Most of the computational models (Pandzic & Forcheimer 2002; Ruttkay *et al.* 2003*a*; Becker & Wachsmuth 2006) lay on the work conducted by Ekman and his colleagues (Ekman & Friesen 1975; Ekman 2003). They represent the six prototypical expressions of the emotions: anger, disgust, fear, happiness, sadness and surprise. To allow ECAs to display a larger number of expressions, computational models following the dimensional representation of emotions have been proposed (Tsapatsoulis *et al.* 2002; Albrecht *et al.* 2005; Garcia-Rojas *et al.* 2006). A new expression is computed as the combination of pre-defined expressions of emotions that have been placed in the dimensional space. However, the expressions obtained through arithmetical functions may not be perceptually valid. Lately research has been undertaken to ensure that any intermediate expressions are valid (Grammer & Oberzaucher 2006; Arya *et al.* 2009; Stoiber *et al.* 2009). These studies are either based on user perception studies or on motion capture of facial activities.

Other approaches are based on appraisal theories, in particular on Scherer and colleagues' theory (Scherer 2001; Scherer & Ellgring 2007). Expressions of emotions arise from the sequence of facial actions (Malatesta *et al.* 2006; Paleari & Lisetti 2006). The temporal course of an expression of emotion no longer follows a trapezoid shape; i.e. it does not: appear (onset), remain (apex) and disappear (offset). Rather, signals are temporally displayed. An expression appears as a succession of signals. Recently a language to describe the expression of emotion as a dynamic sequence of behaviours has been proposed (Niewiadomski *et al.* in press *b*). An emotion is expressed over the whole body as a temporal arrangement of multimodal signals. The language describes not only the signals used in the expression but also the temporal constraints linking the signals.

Faces can convey complex messages that can correspond to various communicative functions (Poggi 2007) as well as blend of emotions of different types. For example a blend of emotions can correspond to the superposition of emotions, the masking of one felt emotion by another unfelt one. Expressions arising from these complex messages are obtained using a compositional approach. The face is decomposed

*catherine.pelachaud@telecom-paristech.fr

One contribution of 17 to a Discussion Meeting Issue 'Computation of emotions in man and machines'.

into facial areas. The various messages are dispatched over the areas. The resulting expression is composed of the signals over the different facial areas (Poggi & Pelachaud 2000*b*; Pelachaud & Poggi 2002; Duy Bui *et al.* 2004; Martin *et al.* 2006; Niewiadomski *et al.* 2008).

Interestingly there have also been studies on the perception of facial expressions displayed by virtual agents. They have shown that human users are able to decode expressions of emotions (Pandzic & Forcheimer 2002); they are using facial expressions to disambiguate multimodal messages (Torres *et al.* 1997); they can also distinguish when the agent is displaying expressions of felt emotions versus fake emotions (Rehm & André 2005*b*).

In this paper I present an overview of these various models. I will concentrate on the expressions of emotions, especially on the computational models of these expressions for virtual agents. I will not address the issues of determining how an emotion is triggered, neither I will discuss if and why an emotion is displayed or restrained. I will concentrate on the visual outer part of the emotion process: the multimodal expressions of emotions. Particular attention will be devoted to facial expressions. I first describe models viewing facial expressions as intonational markers (§2). I point out how facial expressions related to some communicative functions are related to emotion expressions. Then I turn my attention to the calculation of the expression of emotions for virtual agents. Computational models for generating facial expressions of emotion are presented in §3. Section 4 reports on models of behaviour expressivity. While §3 concentrates on signals shape, §4 examines qualitative value of behaviour execution. Finally, I turn my attention to complex expressions (§5). I report on studies that establish a link between the cognitive units underlying performative and facial signals (§5*a*). In this section, I also describe studies that view the face can convey several meanings simultaneously (see §5*b*). These models follow a combinatorial approach. That is the facial expression results from the combination of facial signals over different areas of the face. Section 5*c* presents computational models of facial expressions of emotions in a social context. Finally, I expose a model of expression of emotions as a sequence of behaviour signals in §6 and I then conclude the paper.

2. INTONATIONAL MARKERS

The face transmits signs of emotion but also other communicative functions, in particular related to intonation. Facial expressions can serve as intonational markers. They can help to disambiguate novel information from that previously introduced. Moreover, as we will see, some of the facial signals used as intonational cues relate to emotion expression (Poggi & Pelachaud 2000*a*).

Intonation is defined as the melodic feature of an utterance. It is linked to the syntax of an utterance, the attitudes of the speaker and the emotions (Scherer 1988). Facial expressions can serve as cues to signal acoustic prominence (Bolinger 1989; Beskow *et al.* 2006; Swerts & Krahmer 2008). Indeed, detailed visual–acoustic analyses have shown the relation between pitch accents and facial signals. Eyebrow shapes, head movement and gaze direction have been found to be linked to the intonational structure. There is more eyebrow activity and head movement coinciding with pitch accents. For example, raised eyebrows and frown occur with stressed word (Ekman 1989). More facial actions occur on a stressed word. Indeed, Humans use facial actions as a cue of accent signalling (Krahmer & Swerts 2004; Swerts & Krahmer 2008). Perceptual tests have been conducted to understand how facial actions participate in the perception of prominence. Through the manipulation of prominence placement, videos of two- or three-dimensional virtual agents (Krahmer & Swerts 2004) as well as videos of humans (Swerts & Krahmer 2008) have been recorded. Animations of two- and three-dimensional agents were created where facial actions of the agent (mainly eyebrow raising) were displayed at various times. In some of the videos, facial actions (e.g. raised eyebrows) did not coincide with pitch accent. The participants of the evaluation studies had to determine which word was stressed. In all the studies, be it with a two- or three-dimensional agents or with a human, visual cues help to locate pitch accent.

In many ECA systems (Cassel *et al.* 1994; Pelachaud & Prevost 1994; Pelachaud *et al.* 1996; Beskow 1997; Torres *et al.* 1997; Stone & DeCarlo 2003; Beskow *et al.* 2006), algorithms have been proposed to align facial expressions with the intonation structure of the sentence. The type of facial actions and head movements vary with the type of pitch accent (Stone & DeCarlo 2003) as well as speaker's attitude (Beskow *et al.* 2006). In Pelachaud *et al.* (1996) the algorithm used to decide when and which facial expressions to display with pitch accent is based on rules that describe and synchronize the following relationship: intonation, information and facial expressions. Each utterance to be said by the agent is decomposed in rheme (what is new in the utterance) and theme (what was already given in the discourse context) (Steedman 1991). This structure is transmitted, in synchrony, acoustically by the placement of the focal accents and their type and visually by facial expressions and head movement. The choice of signals to display depends on the type of information to be conveyed. Two types are considered: newness and goal obstruction. Raised eyebrows can be used to mark new information in the utterance (Pelachaud *et al.* 1996). It marks newness, novel information. One can notice that raised eyebrows are also part of the expression of surprise. In both cases, raised eyebrows is related to new information. On the other hand, the facial signal, frown, is linked to goal obstruction (Stone & DeCarlo 2003); it also signals an angry emotional state. Facial expressions linked to intonation is somehow related to some aspects of emotion (Poggi & Pelachaud 1999).

3. GENERATION OF FACIAL EXPRESSIONS OF EMOTIONS

Several computational models of facial expressions for virtual agents (Pandzic & Forcheimer 2002;

Ruttkey *et al.* 2003*a*; Becker & Wachsmuth 2006) are based on the prototypical expressions of emotions described in the work of Ekman and his colleagues (Ekman & Friesen 1975; Ekman 2003). There are fewer models that look at enriching the palette of expressions of an ECA. Most existing solutions compute new expressions as an algebraic combination of the facial expression parameters of predefined expressions. That is an expression is obtained by combining the expressions of other emotions. In EmotionDisc (Ruttkay *et al.* 2003*a*), the six basic emotions are uniformly distributed on a disc. The centre of the circle corresponds to the neutral expression. The spatial coordinates of any point in the disc allow the computation of the corresponding facial expression as a linear combination of the two closest known expressions. The distance from this point to the centre of the disc represents the intensity of the expression. This approach of bilinear combination of expressions has been extended to more complex models (Tsapatsoulis *et al.* 2002; Albrecht *et al.* 2005; Garcia-Rojas *et al.* 2006). These models use a dimensional representation of the emotions (Plutchik 1980; Whissel 1989). The representation can be in two-dimensional with the axes arousal and valence (Tsapatsoulis *et al.* 2002; Garcia-Rojas *et al.* 2006) or in three-dimensional with the axes arousal, valence and power (Albrecht *et al.* 2005). A new expression is obtained by combining the expressions of the closest basic emotions in the representation space. This combination is done at the level of the facial parameters and depends on the distance between the coordinates of the considered emotions.

Another computational approach uses fuzzy logic. Duy Bui *et al.* (2004) follow Ekman's approach to compute facial expressions arising from blend of emotions (Ekman & Friessen 1975). Their model applies a set of fuzzy rules to compute the mix of expressions of the six basic emotions. The face is decomposed into two parts: upper part and lower part. The fuzzy rules determine which emotion is displayed on the upper part of the face and which one on the lower part. The muscular intensity is computed by fuzzy inferences as a function of the intensity of each emotion.

Niewiadomski & Pelachaud (2007*a*,*b*) and Martin *et al.* (2006) have developed a model of complex expressions. Complex expressions correspond to different types of emotions such as masking one emotion by another, superposition of two emotions, inhibition or exaggeration of one emotion, etc. Faces are decomposed into eight areas. A set of fuzzy rules produces the final expression of a complex emotion as the combination of facial areas of several emotions (see figure 1).

Some researchers have applied a combinatorial approach to compute the facial expressions of emotions (Malatesta *et al.* 2006; Paleari & Lisetti 2006). They consider the signals to be combined in a temporal sequence. These models follow Scherer's appraisal theory (Scherer 2001). A facial expression is created as a temporal sequence of the facial signals appearing consecutively because of the cognitive evaluations predicted by Scherer's theory.

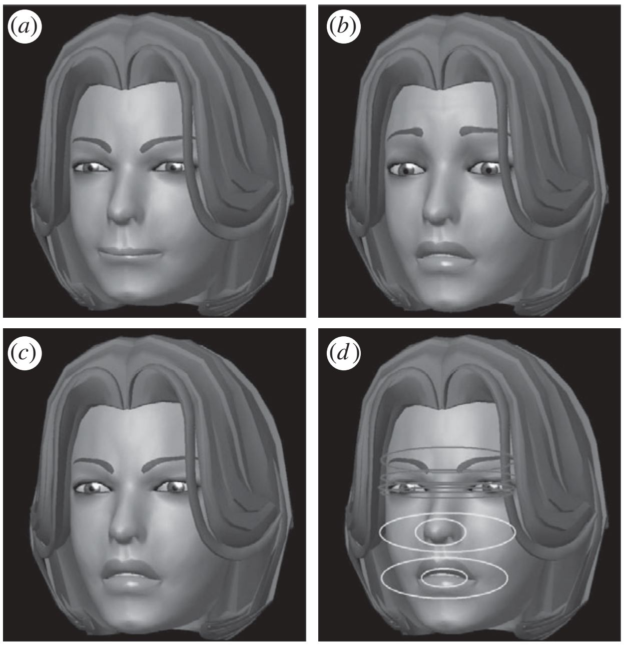

Figure 1. Superposition of two emotional states: (*a*) anger; (*b*) sadness; (*c*) superposition of anger and sadness; (*d*) facial areas of each emotion: anger is shown on the upper facial areas and sadness on the lower areas.

4. BEHAVIOUR EXPRESSIVITY

As mentioned above, many studies have reported that emotions are distinguished by facial expressions. Fewer pieces of work have reported that body posture and body expressivity are also specific to emotional states. The movement accompanying an emotional state differs when it is an angry state or a sad state (Wallbott 1998). The execution of a gesture and of an expression is also very representative of an emotional state. Behaviour expressivity, the qualitative feature of behaviour, has been studied through perceptual studies (Wallbott & Scherer 1986; Wallbott 1998) and image analysis (Castellano *et al.* 2007; Caridakis *et al.* 2008). Behaviour expressivity for dance has also been carefully looked at (Laban & Lawrence 1974).

Building an expressive ECA does require endowing it with appropriate facial expression but also with the capability to execute behaviours with different expressivities. Both behaviour shape and behaviour expressivity contribute to the representation of an emotional state.

In the domain of ECAs, Ruttkay *et al.* (2003*b*) have developed a representation scheme that encompasses behaviour styles. An agent is described along a large number of dimensions spreading from its culture and profession to its physical and emotional state. These dimensions act on the way the ECA behaves nonverbally. Depending on the setting that has been chosen for a given ECA, the system looks in a library of behaviours to select the corresponding behaviours. It allows the authors to simulate behaviour styles in the agent.

EMOTE (Chi *et al.* 2000) implements Laban's annotation scheme for dance. The model acts on the effort and shape components of the Laban movement analysis (LMA). The effort component refers to the dynamical property of movement. It encompasses four factors: weight (degree of continuity in movement), space (linear vs curvilinear trajectory of the articulators), time (degree of temporal continuity of the movement) and flow (fluidity of the movement). These parameters affect the wrist location of the humanoid. Using an inverse kinematics technique, the shape of the body is computed for each wrist position. With EMOTE, a gesture is modulated by several parameters that specify its strength, its fluidity and its tempo. It acts as a filter on the overall animation of the virtual humanoid. The model adds expressivity to the final animation. Other work is based on motion capture to acquire the expressivity of behaviours during a physical action, a walk or a run (Ménardais *et al.* 2004; Neff & Fiume 2004; Liu & Hertzmann 2005).

Based on perceptual studies conducted by Wallbott & Scherer (1986) and Wallbott (1998) a model of nonverbal behaviour expressivity has been defined that acts on the synthesis computation of a behaviour (Hartmann *et al.* 2005). Among the large set of variables that are considered in the perceptual studies, six parameters (Hartmann *et al.* 2005) were retained and implemented in the Greta ECA system (Pelachaud 2005; Bevacqua *et al.* 2007). Three of the dimensions, spatial extent, temporal extent and power, act directly on the formational parameters of behaviour (gesture, facial expression or head movement). *Spatial extent* operates on the signal amplitude; it corresponds to the physical displacement of a facial feature or of the hand position. *Temporal extent* is linked to behaviour duration; it modifies the speed of behaviour execution. The third parameter, *power*, corresponds to the dynamic property of behaviour; it is implemented as movement acceleration. Another dimension, fluidity, operates over successive behaviours of the same modality. It characterizes the degree of continuity between consecutive behaviours; i.e. how one behaviour is moulded into the next one. This model has been evaluated through two perceptual tests (Buisine *et al.* 2006). Subjects viewed various animations of the agents and had to select either which expressivity parameters have been affected or which movement quality the animation reflects. In both tests, results are mainly positive. Subjects did recognize all in all the different expressive parameters, ensuring us that each of them has a distinctive effect on the final animation of the agent. For two out of three movement qualities, they also positively attributed the animation of behaviour expressivity. The most recognized animations refer to vigorous and abrupt movement qualities. The least corresponds to a sluggish quality; indeed, there was discrepancy between the gesture shape and the movement quality that the subjects perceive.

This last model of behaviour expressivity has been used to develop distinctive agents (Mancini & Pelachaud 2008). Distinctiveness is defined here as the quantitative and qualitative features that make agents communicate differently one from another. An agent is characterized by a baseline that specifies its general behaviour tendency. For example an agent can tend to use few gestures in a slow and soft manner while another one will behave in an expansive manner. The parameters characterizing the baseline of an agent encompass this behaviour variability: how much a modality is used and how expressive are behaviours on each modality. Thus the baseline is constituted as a set of pairs, one per modality that details behaviour characteristics of usage and expressivity. The notion of local tendency has been introduced. When a given agent, described by its baseline, communicates an emotional state or an intention, its global tendency may be modulated. The local tendency is obtained by modulating locally the baseline.

Several studies have been conducted to detect manually and automatically the values of these expressivity parameters. The copy-synthesis method (Devillers *et al.* 2005) starts from a multi-level annotation of video corpora. The highest level corresponds to emotion labels whereas the lowest one corresponds to behaviour shape and movement description. The animation of a virtual agent is derived either from the high-level annotations or low-level ones. Expressivity parameters are obtained either through manual annotation (Devillers *et al.* 2005) or automatically (Castellano *et al.* 2007; Caridakis *et al.* 2008). The six parameters along with the other annotations are sent to an interpretation module that controls the agent animation. Perceptual studies were conducted to study which low-level features carry out the emotion. They allowed us to investigate the role of each multimodal behaviour in the perception of emotions (Martin *et al.* 2006).

5. COMPLEX MESSAGES

The face can convey complex messages. As said earlier it can provide information on various aspects of emotional, communicative and intentional features. We followed the approach that views a virtual agent as social and communicative interlocutor, and that, as such, ought to be endowed with human-like capabilities; we have developed algorithms aiming at generating complex messages on the agent's face.

One of our first attempts regards the computation of facial expressions related to performative. We have developed a mechanism that computes, for a given agent and for each of its intentions to communicate, the performative and its corresponding facial expressions to display.

(a) *Performative faces*

Performatives are used a lot in every day language. They received and are still receiving great attention from linguists (Austin 1962; Benveniste 1974; Ducrot *et al.* 1980; Ducrot 1984; Cervoni 1987). Austin (1962) in his book '*How to do things with words*' described how a class of verbs has included in it say a value of action. These verbs are used neither to describe, nor to affirm something, but they imply the performance of an action. For example with 'I greet', the speaker is not describing a state of the world or a mental state, but the speaker is performing the act of greeting. Thus, the utterance self-contains the performance of an action (the action of greeting). As the speaker says the utterance she is performing the action. An utterance which has made its performative explicit ('I order you to do it') is explicitly an order, while the utterance ('do it') may be interpreted differently: it could be advice, a suggestion, imploring and so on. The facial expression and the voice quality of the speaker as well as the context the utterance takes place in will determine a particular interpretation. That is the listener with his knowledge of the situation, of the speaker's personality and behaviour is able to gather how the utterance should be taken. In that case signals other than the verbal one are necessary to the speaker to communicate her goal (of advising, suggesting . . .) and to the listener to interpret the speaker's goal (Poggi & Pelachaud 2000*b*).

Performatives have been analysed and represented in cognitive units using the BDI (Belief Desire Intention) formalism (Castelfranchi & Parisi 1980; Poggi & Pelachaud 2000*b*). The cognitive units we considered that distinguish performative one from the other are (Poggi & Pelachaud 2000*b*):

(i) *In whose interest* is the action requested or information provided. In an order the action is for the speaker while in advice it is for the interlocutor.
(ii) *Degree of certainty* of what is being said. For example, this cognitive unit differentiates the performatives 'suggest' and 'claim'.
(iii) *Power relationship* between the interactants. When imploring, the speaker acknowledges the listener's power over him while in a command this is the other way around and, in advice the speaker does not consider this relation.

A correspondence between cognitive unit and facial signals has been established (Poggi & Pelachaud 1999). For example the degree of certainty is marked with eyebrow shape and power relationship with head direction. Raised eyebrows mark a low degree of certainty while a frown signals a high degree. Head bent sideways is often a sign of submissiveness. For whose interest it is often signalled by gaze direction. When the action is for the other as in 'suggest' the head can be leant forward while it is kept straight ahead when the requested action is for oneself. To compute the facial expression associated with a performative, we compose it as the combination of the facial signals corresponding to the cognitive units of the performative. The combination is simply obtained by adding each non-verbal signal. In the example below, speaker is denoted A_i and listener A_j. Figures 2 and 3 display the associated facial expressions.

(b) *Conflict of facial signals*

To control the agent, we are using a representation language APML (Affective Presentation Markup Language) (DeCarolis *et al.* 2004). This language is based on Poggi's theory of multimodal communication (Poggi 2007). The elements of the language correspond to the communicative functions that the agent aims to convey. They do not provide information on the multimodal behaviours that will be used to convey them. This conversion is done with our agent system (Pelachaud 2005). The input of our system is the text to be said by the agent enhanced with information on how to convey it using APML tags. Several of these tags can be specified over a given text span. For example the utterance 'Glad to see you' can happen with a performative *greet*, a *happy* emotional state and an *emphasis*. The system translates each of these communicative functions into a set of multimodal behaviours (tables 1 and 2).

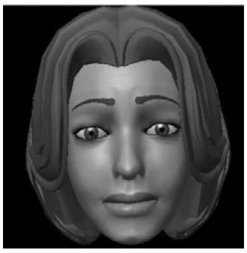

Figure 2. The imploring expression.

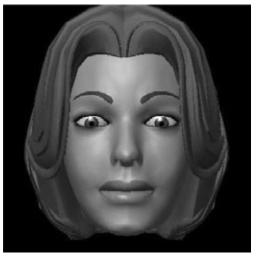

Figure 3. The ordering expression.

However, we may have two or more communicative functions that activate in the same time interval two different signals (such as frown) and (raise eyebrows) on the same facial region (eyebrow). Such a conflict may occur when a person gives an order (signalled by a frown) and at the same time talks about two elements in contrast with each other (the roses and the tulips in the sentence: 'Take the flowers, the roses NOT the tulips'). 'Contrast' may be signalled by a raised eyebrow while 'order' by a frown. Which signals will the face display: a raised eyebrow or a frown? There is a conflict at the facial signal level that must be solved before visualizing the animation. Adding both expressions to the face may create very awkward expression. When a facial signal conflict is detected, the system calls up a special module for its resolution (Pelachaud & Prevost 1994). Such a module determines which signal, of those that should be active on the same facial region, must prevail over the others. That is, the module chooses to display either the frown (of 'order') or the raised eyebrow (of 'contrast'). The conflict resolution is done only on the part of the face conflicting, here the eyebrow region. Thus, using always the same example, the performative 'order' corresponds to frown, look down while gazing at the interlocutor while contrast is marked by a raised eyebrow. The final expression as outputted by the resolution conflict module will have the head and gaze direction of 'order' and the eyebrows either of 'order' or of 'contrast' depending on the output of the conflict module. In this example the output is frown, the eyebrow shape of 'order'. Thus the final expression is obtained by combining

Table 1. Performative of imploring.

cognitive units	facial actions
A_i's request is for A_i's goal	A_i keep head right
A_i claims being in power of A_j	A_i bend head aside
A_i is potentially sad	A_i raise inner brows

Table 2. Performative of ordering.

cognitive units	facial actions
A_i's request is for A_i's goal	A_i keep head right
A_i claims having power on A_j	A_i look down at A_j
A_i is potentially angry at A_j	A_i frown

signals corresponding to different communicative functions over the various facial areas. We have already used such a combinatorial approach to create expression linked to performative. This combinatorial method allows us to create facial expressions whose meanings correspond to the combination of the meanings conveyed by the various signals that composed the expressions.

(c) *Expressions of emotions in a social context*

When interacting we take into account (being more or less conscious about it) several factors such as the social and affective relationships we maintain with our interlocutors. We are not impulsive and have learned to control our behaviour. That is we know when an expression can be shown in which circumstances and to whom. Ekman has introduced the term 'display rules' to refer to rules dictated by our social, cultural and professional background. These rules are linked to each culture and society. They govern how behaviours can be shown in various contexts. ECAs are virtual entities capable of communicating with human users. They are often viewed as social entities. As such it is necessary to model adequately their behaviours taking into account their social and cultural context.

Social context has been incorporated in agent system (André *et al.* 2004; Ballin *et al.* 2004; Johnson *et al.* 2004). Prendinger & Ishizuka (2001) model social role awareness in ECAs. They specify a management of facial expression with a set of rules named social filter programs. These filters are based on social conventions (e.g. politeness) and personalities of the interlocutors. They define the intensity of an expression as a function of social thread (power and distance), user personality (pleasant, extrovert) and emotion intensity. The results of this filter allow one to increase, to decrease the intensity of a facial expression or even to inhibit it completely.

De Carolis *et al.* (2001) have built a reflexive agent able to adapt its expressions of emotions depending on the conversational settings. The emotional expressions of the agent depend on emotional factors (e.g. valence of the emotion, social acceptation, emotion of the interlocutor) and scenario factors (personality, intentions, relationships, types of interaction). The reflexive agent applies management rules that determine in which context can felt emotions be shown or not (De Carolis *et al.* 2001). When an emotion cannot be displayed, its expression is inhibited.

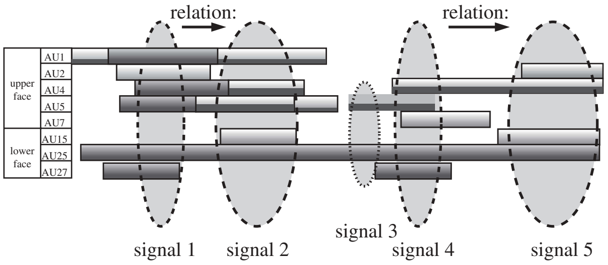

Figure 4. Excerpt of annotation from a video corpus. Facial action, hand and gaze behaviours are annotated.

Niewiadomski & Pelachaud (2007) extended this notion of management rules. Based on the analysis of a video corpus, they annotate the type of complex emotion expressions displayed (masking, inhibition, superposition). In a previous study, Rehm & André (2005*a*) annotated the same corpus with politeness theory tags. Statistical measures were made to obtain the link between facial expression types and politeness tags. The rules that drive the facial expressions of the agent encompasses which expression types should be used in which condition. Depending on the context, the agent can express itself through great variety: inhibition of the emotion felt, masking the emotion felt by another one, or display a fake expression (Niewiadomski & Pelachaud 2007).

6. SEQUENCE OF EXPRESSIONS

Apart from preliminary works by Paleari & Lisetti (2006) and Malatesta *et al.* (2006), most of the other works in virtual agents consider that facial expressions appear as a full-blown expression. Computationally, these expressions are characterized by a trapezoid temporal course where they appear, remain and disappear. All the facial actions composing the expressions follow the same temporal pattern.

Some careful analyses of video corpus have highlighted that the expressions of emotions are dynamic and result as sequences of signals (Keltner 1995; Bänziger & Scherer 2007). An expression is not viewed as a static expression at its apex, rather it is a succession of signals arranged in time interval.

We have ourselves conducted the annotation of video corpus (Niewiadomski *et al.* in press *a*). Facial expressions are annotated using FACS, Facial Action Coding System (Ekman *et al.* 2002). Hand and arm gestures as well as head movements are annotated using free language (e.g. head nod, raise arm up). Figure 4 shows an example of such an annotation. From a careful analysis of the annotations, some patterns emerge in the set of signals. For example a signal is always followed by another one or a signal always starts the sequence, and so on. These patterns are gathered as temporal constraints linking the signals. Apart from the temporal constraints, there can be constraints on the appearance of signals such that

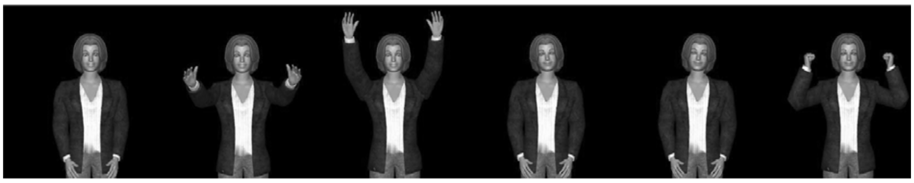

Figure 5. Sequence of signals for the cheerful emotion.

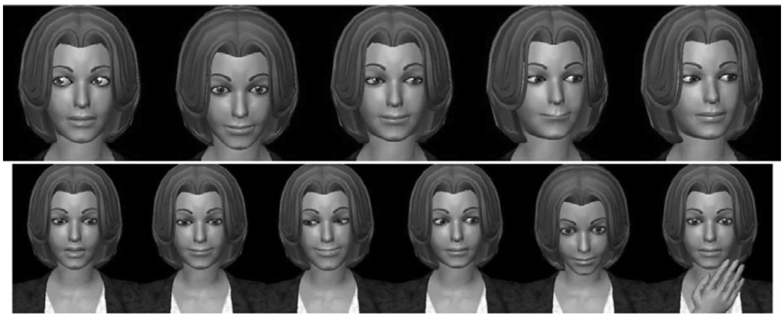

Figure 6. Two different sequences of signals for embarrassment.

two signals cannot appear together, or that the appearance of one signal always coincides with the appearance of another one. These constraints are clustered as the appearance constraints.

We have developed a symbolic language to encompass this temporal quality (Niewiadomski *et al.* in press *b*). The language allows us to describe the set of signals that appears in the expression of emotions. The signals can be a gaze direction, a hand gesture, a facial signal, a head movement or a torso movement. A set of operators has been defined to specify the temporal and the appearance relationships between these various signals. These relationships describe the constraints that need to be satisfied between the signals in a sequence. For example the constraints make explicit when a signal should start in relation with the starting or ending time of other signals. The signals could overlap completely, partially, or not at all. A signal may always be followed by another one or vice versa. The language captures the temporal and appearance constraints.

For a given emotion, our algorithm generates a sequence of signals from this symbolic description. Each signal is specified by its facial parameters as well as its temporal information. Figure 5 illustrates the sequence of signals related to the cheerful emotion. This algorithm can generate different animations, each satisfying the temporal and appearance constraints, for a given animation. It is a first step towards the creation of non-repetitive behaviours for virtual agents. Figure 6 shows two sequences for a same emotion, embarrassment.

An evaluation study has been conducted to measure the recognition rate of the emotional state from these sequences of expressions Niewiadomski *et al.* in press *b*). Fifty-three subjects (means age 28 years, from various European countries) viewed animations corresponding to eight emotions (anger, anxiety, cheerfulness, embarrassment, panic fear, pride, relief and tension) in a random order. The animations were shown twice. Subjects had to select the label of the emotion the animation represents from a closed list of eight labels. The results show the recognition level of each expression above the chance level. No confusion was made between positive and negative emotions and between emotions with duchenne or non-duchenne smiles. Subjects could recognize expressions of emotions that were not all part of the prototypical emotions. These results are encouraging. We are currently working on annotating more videos and elaborating a smoother and more realistic animation system.

7. CONCLUSION

In this paper I have drawn a fast panorama of the various existing technologies to create expressions of emotions. Computational models of emotional expressions implement a large variety of facial expressions. Models follow the different emotion theories. Some models are based on discrete representation of emotion; others on the dimensional models; and finally others on appraisal theories. These computational works tackle different issues in ECA behaviour models.

I have also presented work whose aims are to encompass a virtual agent with subtle and complex expressions. Combinatorial approach is used. An expression is the result of the arithmetic combination of facial signals spread over the face. Each signal shape is linked to a given meaning. The message carried by the expression corresponds to the combination of meanings attached to each signal.

So far, expression of emotions has had a temporal course of trapezoid form. The expression appears, remains and disappears on the face. To endow virtual agents with more dynamical behaviours and to give them the possibility to transmit a large variety of emotions, a sequential model of expressions has been introduced. Expressions are no more defined by a static representation; rather they are constituted as a succession of signals that appear dynamically.

We are pursuing our work on creating virtual agents able to display complex messages. We are extending our work to the whole body. It is our belief that agents should be able to convey very subtle messages to be a long-term conversational companion. Using few expressions limits the interaction. Endowing agents with large variety of expressions ensures more naturalness in the agent's behaviour. Such a result cannot be achieved without a model that decides which emotions the agent should display. This paper has not addressed this very important issue but it focused on the creation of facial expressions of emotions.

I am very grateful to Isabella Poggi whose collaboration has always been and is still very rich and enlightening. All the work done in the Greta agent would not have been possible without the contributions of numerous colloborators among which are Elisabetta Bevacqua, Sylwia Hyniewska, Maurizio Mancini and Radoslaw Niewiadomski. This work has been partially funded by the IP-CALLAS project IST-034800 (http://www.callas-newmedia.eu) and by the STREP SEMAINE project IST-211486 (http://www.semaine-project.eu).

REFERENCES

Albrecht, I., Schroeder, M., Haber, J. & Seidel, H.-P. 2005 Mixed feelings: expression of non-basic emotions in a muscle-based talking head. *Virtual Real.* (special issue) **8**, 201–212. (doi:10.1007/s10055-005-0153-5).

André, E., Rehm, M., Minker, W. & Buhler, D. 2004 Endowing spoken language dialogue systems with emotional intelligence. In *Affective dialogue systems* (eds E. André, L. Dybkjaer, W. Minker & P. Heisterkamp), pp. 178–187. Berlin/Heidelberg, Germany, Springer Verlag.

Arya, A., DiPaola, S. & Parush, A. 2009 Perceptually valid facial expressions for character-based applications. *Int. J. Comput. Game Technol.* Article ID 462315. (doi:10.1155/2009/462315)

Austin, J. L. 1962 *How to do thinks with words*. London, UK: Oxford University Press.

Ballin, D., Gillies, M. F. & Crabtree, I. B. 2004 A framework for interpersonal attitude and non-verbal communication in improvisational visual media production. In *First European Conf. on Visual Media Production* (*CVMP*), pp. 203–210, London, UK.

Bänziger, T. & Scherer, K. 2007 Using actor portrayals to systematically study multimodal emotion expression: the GEMEP corpus. In *Affective computing and intelligent interaction*. Lecture Notes in Computer Science, no. 4738/ 2007. Berlin/Heidelberg: Springer.

Becker, C. & Wachsmuth, I. 2006 Modeling primary and secondary emotions for a believable communication agent. In *Int. Workshop on Emotion and Computing, in conj. with the 29th annual German Conf. on Artificial Intelligenz* (*KI2006*), pp. 31–34, Bremen, Germany.

Beneveniste, E. 1974 *Problèmes de linguistique générale*. Paris, France: Gallimard.

Beskow, J. Animation of talking agents. 1997 In *Proc. ESCA Workshop on Audio-Visual Speech Processing* (eds C. Benoit & R. Campbell), pp. 149–152. Rhodes, Greece.

Beskow, J., Granstrom, B. & House, D. 2006 Visual correlates to prominence in several expressive modes. In *Proceedings of Interspeech*, pp. 1272–1275, Pittsburg, PA.

Bevacqua, E., Mancini, M., Niewiadomski, R. & Pelachaud, C. 2007 An expressive ECA showing complex emotions. In *AISB'07 Annual convention, workshop "Language, Speech and Gesture for Expressive Characters"*, pp. 208–216, Newcastle UK.

Bolinger, D. 1989 *Intonation and its uses*. Stanford, CA: Stanford University Press.

Buisine, S., Abrilian, S., Niewiadomski, R., Martin, J.-C., Devillers, L. & Pelachaud, C. 2006 Perception of blended emotions: from video corpus to expressive agent. In *The 6th International Conference on Intelligent Virtual Agents, Marina del Rey, USA*.

Caridakis, G., Raouzaiou, A., Bevacqua, E., Mancini, M., Karpouzis, K., Malatesta, L. & Pelachaud, C. 2008 Virtual agent multimodal mimicry of humans. In *Language resources and evaluation, special issue on Multimodal Corpora for Modelling Human Multimodal Behavior*, vol. 41 (eds J.-C. Martin, P. Paggio, P. K'uhnlein, R. Stiefelhagen & F. Pianesi), pp. 367–388. Berlin, Heidelberg: Springer.

Cassell, J., Pelachaud, C., Badler, N. I., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S. & Stone, M. 1994 Animated conversation: rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Computer Graphics Proc. Annual Conf. Series*, p. 413–420. ACM SIGGRAPH.

Castelfranchi, C. & Parisi, D. 1980 *Linguaggio, conoscenze e scopi*. Bologna, Italy: Il Mulino.

Castellano, G., Villalba, S. D. & Camurri, A. 2007 Recognising human emotions from body movement and gesture dynamics. In *ACII*, pp. 71–82.

Cervoni, J. 1987 *Lénonciation*. Linguistique nouvelle. Paris, France: Presses Universitaires de France.

Chi, D. M., Costa, M., Zhao, L. & Badler, N. I. 2000 The EMOTE model for effort and shape. In *Siggraph 2000, Computer Graphics Proceedings* (ed. K. Akeley), pp. 173–182. Boston, MA: ACM Press/ACM SIGGRAPH/ Addison Wesley Longman.

DeCarolis, B., Pelachaud, C., Poggi, I. & de Rosis, F. 2001 Behavior planning for a reflexive agent. In *IJCAI'01*, Seattle, USA.

DeCarolis, B., Pelachaud, C., Poggi, I. & Steedman, M. 2004 APML, a mark-up language for believable behavior generation. In *Life-like characters. Tools, affective functions and applications* (eds H. Prendinger & M. Ishizuka), pp. 65–85. Berlin/Heidelberg: Springer.

Devillers, L., Abrilian, S. & Martin, J.-C. 2005 Representing real life emotions in audiovisual data with non basic emotional patterns and context features. In *1st Int. Conf. Affective Computing & Intelligent Interaction* (*ACII'2005*), *Beijing, China, 22–24 October 2005*.

Ducrot, O. 1984 *Le dire et le dit*. Paris, France: Les Editions de Minuit.

Ducrot, O. *et al.* 1980 *Les mots du discours*. Paris, France: Les Editions de Minuit.

Duy Bui, Heylen, D., Poel, M. & Nijholt, A. 2004 Combination of facial movements on a 3D talking head. In *Computer Graphics International*, pp. 284–291.

Ekman, P. 1989 The argument and evidence about universals in facial expressions of emotion. In *Handbook of social psychophysiology* (eds H. Wagner & A. Manstead), pp. 143–164. Chichester, UK: Wiley.

Ekman, P. 2003 *Emotions revealed*. New York/London: Times Books (US)/Weidenfeld & Nicolson.

Ekman, P. & Friesen, W. 1975 *Unmasking the face: a guide to recognizing emotions from facial clues*. Eaglewood Cliffs, NJ: Prentice-Hall, Inc.

Ekman, P., Friesen, W. & Hager, J. 2002 *Facial action coding system: the manual & user' guide*. Salt Lake City, UT: A Human Face.

Garcia-Rojas, A., Vexo, F., Thalmann, D., Raouzaiou, A., Karpouzis, K., Kollias, S., Moccozet, L. & Magnenat-Thalmann, N. 2006 Emotional face expression profiles supported by virtual human ontology. *Comp. Anim. Virtual Worlds* **17**, 259–269. (doi:10.1002/cav.130).

Grammer, K. & Oberzaucher, E. 2006 The reconstruction of facial expressions in embodied systems. ZIF Mitteiluagen 2/2006, pp. 14–31.

Hartmann, B., Mancini, M. & Pelachaud, C. 2005 Implementing expressive gesture synthesis for embodied conversational agents. In *Gesture in Human–Computer Interaction and Simulation, 6th International Gesture Workshop, GW 2005, Berder Island*, pp. 188–199.

Johnson, W. L., Rizzo, P., Bosma, W., Kole, S., Ghijsen, M. & van Welbergen, H. 2004 Generating socially appropriate tutorial dialog. In *ISCA Workshop on Affective Dialogue Systems*, pp. 254–264.

Keltner, D. 1995 Signs of appeasement: evidence for the distinct displays of embarrassment, amusement, and shame. *J. Pers. Soc. Psychol.* **68**, 441–454. (doi:10.1037/0022-3514.68.3.441).

Krahmer, E. & Swerts, M. 2004 More about brows. In *From brows till trust: evaluating embodied conversational agents* (eds Z. Ruttkay & C. Pelachaud) pp. 191–216. Norwell, MA: Kluwer Academic Publishers.

Laban, R. & Lawrence, F. C. 1974 *Effort: economy in body movement*. Boston, MA: Plays, Inc.

Liu, K. & Hertzmann, A. 2005 Learning physics-based motion style with inverse optimization. In *Computer Graphics Proc. Annual Conf. Series*. ACM SIGGRAPH.

Malatesta, L., Raouzaiou, A., Karpouzis, K. & Kollias, S. 2006 MPEG-4 facial expression synthesis based on appraisal theory. In *3rd IFIP Conf. in Artificial Intelligence Applications and Innovations*.

Mancini, M. & Pelachaud, C. 2008 Distinctiveness in multimodal behaviors. In *Seventh International Joint Conf. on Autonomous Agents and Multi-Agent Systems, AAMAS'08, Estoril, Portugal*.

Martin, J.-C., Niewiadomski, R., Devillers, L., Buisine, S. & Pelachaud, C. 2006 Multimodal complex emotions: gesture expressivity and blended facial expressions. *Int. J. HR.* (special issue) **20**, 477–498.

Ménardais, S., Kulpa, R., Multon, F. & Arnaldi, B. 2004 Synchronization for dynamic blending of motions. In *Proc. of ACM, SIGGRAPH/EUROGRAPHICS. Symposium of Computer Animation*, pp. 325–335, France.

Neff, M. & Fiume, E. 2004 Methods for exploring expressive stance. In *Proc. ACM, SIGGRAPH/EUROGRAPHICS Symposium of Computer Animation*, pp. 49–58, Grenoble, France.

Niewiadomski, R. & Pelachaud, C. 2007*a* Fuzzy similarity of facial expressions of embodied agents. In *Proc. of the 7th Int. Conf. on Intelligent Virtual Agents* (*IVA*) (eds C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis & D. Pélé), pp. 86–98. Berlin, Heidelberg: Springer Verlag.

Niewiadomski, R. & Pelachaud, C. 2007*b* Model of facial expressions management for an embodied conversational agent. In *2nd Int. Conf. on Affective Computing and Intelligent Interaction* (*ACII*), pp. 12–23, Lisbon, Portugal.

Niewiadomski, R., Ochs, M. & Pelachaud, C. 2008 Expressions of empathy in ECAs. In *Intelligent Virtual Agents, IVA'08*, Tokyo, Japan, pp. 37–44.

Niewiadomski, R., Hyniewska, S. & Pelachaud, C. In press *a* Evaluation of multimodal sequential expressions of emotions in ECA. In *Proc. Int. Conf. on Affective Computing and Intelligent Interaction* (*ACII*), *Amsterdam, The Netherlands*. Berlin/Heidelberg: Springer Verlag.

Niewiadomski, N., Mancini, M., Hyniewska, S. & Pelachaud, C. In press *b*. Emotional behaviours in embodied conversational agents. In *A blueprint for an affectively competent agent* (eds E. Roesch, K. R. Scherer & T. Bänziger).

Paleari, M. & Lisetti, C. L. 2006 Psychologically grounded avatars expressions. In *First Workshop on Emotion and Computing at KI 2006, 29th Annual Conf. on Artificial Intelligence, Bremen, Germany*.

Pandzic, I. S. & Forcheimer, R. (eds) 2002 *MPEG4 facial animation – The standard, implementations and applications*. Chichester, UK: John Wiley, Sons.

Pelachaud, C. 2005 Multimodal expressive embodied conversational agent. In *ACM Multimedia, Brave New Topics session*, Singapore.

Pelachaud, C. & Prevost, S. 1994 Sight and sound: generating facial expressions and spoken intonation from context. In *Proc. of the 2nd ESCA/AAAI/IEEE workshop on Speech Synthesis*, pp. 216–219, New Paltz, NY.

Pelachaud, C. & Poggi, I. 2002 Subtleties of facial expressions in embodied agents. *J. Vis. Comput. Animat.* **13**, 301–312. (doi:10.1002/vis.299).

Pelachaud, C., Badler, N. I. & Steedman, M. 1996 Generating facial expressions for speech. *Cognit. Sci.* **20**, 1–46.

Plutchik, R. 1980 *Emotions: a psychoevolutionary synthesis*. Harper & Row, New York.

Poggi, I. 2007 *Mind, hands, face and body. A goal and belief view of multimodal communication*, volume Körper, Zeichen, Kultur. Berlin: Weidler Verlag.

Poggi, I. & Pelachaud, C. 1999 Emotional meaning and expression in performative faces. In *Int. Workshop on Affect in Interactions: towards a New Generation of Interfaces, Annual Conf. AC'99 of the EC I3 Programme*, Siena, Italy, October 1999.

Poggi, I. & Pelachaud, C. 2000*a* Emotional meaning and expression in animated faces. In *Affect in interactions* (ed. A. Paiva). Berlin, Germany: Springer-Verlag.

Poggi, I. & Pelachaud, C. 2000*b* Facial performative in a conversational system. In *Embodied conversational characters* (eds S. Prevost, J. Cassell, J. Sullivan & E. Churchill), pp. 155–188. Cambridge, MA: MIT press.

Prendinger, H. & Ishizuka, M. 2001 Social role awareness in animated agents. In *Proc. of the 5th Int. Conf. on Autonomous Agents*, Montreal, Canada, May–June 2001.

Rehm, M. & André, E. 2005*a* Informing the design of embodied conversational agents by analysing multimodal politeness behaviors in human–human communication. In *Workshop on Conversational Informatics for Supporting Social Intelligence and Interaction*, 2005.

Rehm, M. & André, E. 2005*b* Catch me if you can: exploring lying agents in social settings. In *AAMAS*, pp. 937–944, 2005.

Ruttkay, Z., Noot, H. & ten Hagen, P. 2003*a* Emotion disc and emotion squares: tools to explore the facial expression face. *Comput. Graph. Forum* **22**, 49–53.

Ruttkay, Z., van Moppes, V. & Noot, H. 2003*b* The jovial, the reserved and the robot. In *Proceedings of the AAMAS03 Ws on Embodied Conversational Characters as Individuals*, Melbourne, Australia, 2003.

Scherer, K. 1988 *Facets of emotion: recent research*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Scherer, K. R. 2001 Appraisal considered as a process of multilevel sequential checking. In *Appraisal processes in emotion: theory, methods, research* (eds K. Scherer, A. Schorr & T. Johnstone), pp. 92–119. New York, NY: Oxford University Press.

Scherer, K. R. & Ellgring, H. 2007 Multimodal expression of emotion: affect programs or componential appraisal patterns? *Emotion* **7**, 158–171. (doi:10.1037/1528-3542.7.1.158).

Stoiber, N., Séguier, R. & Breton, G. 2009 Automatic design of a control interface for a synthetic face. In *IUI*, pp. 207–216, 2009.

Steedman, M. 1991 Structure and intonation. *Language* **67**, 260–296. (doi:10.2307/415107).

Stone, M. & DeCarlo, D. 2003 Crafting the illusion of meaning: template-based generation of embodied conversational behavior. In *Computer Animation and Social Agents*, pp. 11–16. Brunswick, NJ.

Swerts, M. & Krahmer, E. 2008. Facial expressions and prosodic prominence: comparing modalities and facial areas. *J. Phon.* **36**, 219–238. (doi:10.1016/j.wocn.2007.05.001).

Torres, O., Cassell, J. & Prevost, S. 1997 Modeling gaze behavior as a function of discourse structure. In *Machine conversation* (ed Y. Wilks), *First International Workshop on Human–Computer Conversations*, *Bellagio, Italy.* Kluwer.

Tsapatsoulis, N., Raouzaiou, A., Kollias, S., Cowie, R. & Douglas-Cowie, E. 2002 Emotion recognition and synthesis based on MPEG-4 FAPs in MPEG-4 facial animation. In *MPEG4 facial animation – the standard, implementations and applications* (eds I. S. Pandzic & R. Forcheimer). Chichester, UK: John Wiley & Sons.

Wallbott, H. 1998 Bodily expression of emotion. *Eur. J. Soc. Psychol.* **28**, 879–896. (doi:10.1002/(SICI)1099-0992(1998110)28:6<879::AID-EJSP901>3.0.CO;2-W)

Wallbott, H. G. & Scherer, K. 1986 Cues and channels in emotion recognition. *J. Pers. Soc. Psychol.* **24**.

Whissell, C. 1989 The dictionary of affect in language. In *Emotion: Theory, Research, and Experience, volume 4: The Measurement of Emotions* (eds R. Plutchik & H. Kellerman), pp. 113–131. San Diego, CA: Academic Press, Inc.