The Knowledge Engineering Review, Vol. 17:2, 181–196. © 2002, Cambridge University Press
DOI: 10.1017/S0269888902000218 Printed in the United Kingdom

Multimodal embodied agents

CATHERINE PELACHAUD[1] and ISABELLA POGGI[2]

[1] *University of Rome "La Sapienza", Department of Computer and System Science; e-mail: cath@dis.uniroma1.it*
[2] *University of Rome Three, Department of Education; e-mail: poggi@uniroma3.it*

1 Believable interactive embodied agents

Among the goals of research on autonomous agents one important aim is to build believable interactive embodied agents that are apt to application to friendly interfaces in e-commerce, tourist and service query systems, entertainment (e.g. synthetic actors) and education (pedagogical agents, agents for help and instruction to the hearing impaired).

A believable agent is one able to show (even, maybe, to feel?) emotions and one which has a definite personality. An interactive agent can take into account the particular user and the particular context where the interaction takes place, and is able to make up its own model of the user and the context, and interact with it by following the rules of face-to-face interaction, like turn-taking, back-channel and so forth. An embodied agent must be able to interact with the user not only through written text, but in all the modalities a human agent may use: through words, voice, gesture, gaze, facial expression, body movements and body posture (sometimes, maybe, even through touch?). But it also must be able to conceive, represent and convey all the possible meanings that natural language and multimodal interaction may convey in humans.

The list of capacities required by a believable interactive embodied agent allows us to clear up the steps to take in this field of research. Let us now first outline what these steps are in principle, and then see what has already been done in this field. In order to assess the state of the art in actual research, a workshop was organized within the *Fifth International Conference on Autonomous Agents* (Montreal, Canada, 29 May 2001): "Multimodal Communication and Context in Embodied Agents". As it will be clear from an overview of the papers in the workshop, some of the above steps have already been taken in recent work, while others have still to be taken.

2 Believable interactive embodied agents' capacities

Let us first see what should be, in our view, the overall capacities that a Believable Interactive Embodied Agent (BIEA) should be endowed with. Each of these capacities shows the agenda of studies to be carried out in order to make such an agent. The first important issues to tackle are the contents a BIEA should be able to manage in its interaction. These include:

(a) information on the specific domains the agent will be used for (a touristic agent should know everything about hotels and museums, while a mathematical pedagogical agent should be an expert on integrals);
(b) information on the agent's mental states, which includes its possible emotions (he could tell you he is happy to hear you), personality traits (she might tell you she is a jealous agent) but also beliefs (warn you when she is not sure of what she is saying), intentions (inform you she will explain something in more detail) and its social intentions (give you an order or advice, praise you or criticise you);
(c) decision capacity as to whether to provide information or withhold it: humans sometimes do not display their emotions, or they do not tell the truth, or they communicate things in an indirect or covert way (through rhetorical figures such as euphemism, metaphor or irony);

(d) decision capacity as to how to communicate information, provided that an agent has decided to communicate it. A multimodal agent may do this in multiple ways: through speech and intonation, facial expression and gaze, gesture, body movements and posture. So it has to master systems of communication in all modalities and to manage the distribution of information across modalities, deciding which is most appropriate for each context and how they can combine with each other;
(e) consideration of the conversational context. A skilled BIEA must take context into account, that is, all the requirements above need to be fulfilled for each particular User and application: Is the application directed to children, adults or computer scientists? Do we want a friendly agent or an authoritative agent? Using too simple models of communicative behaviour might produce a poor agent with which the user will rapidly feel bored. Again, some applications require caricature and over-expressive behaviours, while others need very accurate and realistic behaviours.

3 Phases of research

Research aimed at constructing BIEAs must include three phases. First, a phase of empirical research aimed at finding out the regularities in the mind and behaviour of human-like agents and at constructing models of them. Second, a phase modelling believable interactive embodied agents, where the rules found out are formalized, represented and implemented in the construction of agents. Third, a phase of evaluation of the implemented agents, aimed at testing how they fit the user's needs and preferences and, especially in a simulative approach, how similar they look to a real human agent. Each of these three phases of research must be carried out for all the topics above. Let us see what has been done already and what should yet be done.

(a) Emotion. A lot of precious work has been done in the realm of emotions, not only from a theoretical and empirical point of view, in the endless literature in the psychology of emotions (Scherer, 1988; Tomkins, 1982; Ekman, 1982; Lazarus, 1991), but also in the field of emotion simulation (Ortony *et al*., 1988; Elliott, 1992; Frijda & Swagerman, 1987; Picard, 1997; Paiva, 2000). Models of emotions have been proposed that can be implemented in emotional agents (Ball & Breese, 2000; Martinho *et al*., 2000).
(b) Personality. Models of personality in computer simulation have mainly followed routes quite different from the classical models in psychology (Freud, Jung). After the brilliant work by Carbonell (Carbonell, 1980), more recent research has interestingly modelled and implemented different possible personalities in artificial agents (Castelfranchi *et al*., 1997; Trappl & Petta, 1997) and taught them to guess the user's personality in user models research (Isbister & Nass, forthcoming, Castelfranchi *et al*., 1998).

    The topic of the decision as to whether to display information or not has been so far explored especially in the emotion literature (Pennebaker, 1982; Rimé, 1987; Ekman, 1982) and somewhat in the literature on deception and secrets (Bok, 1982; Castelfranchi & Poggi, 1998) and on indirect speech acts. The issues addressed so far are mainly what are the effects of sharing emotions (Rimé, 1987; Pennebaker, 1982); the existence, in different cultures, of different "display rules" that determine which emotions one should display or not in different situations; and the cues to deception that unmask people's face when they try to intensify, de-intensify or mask the expression of emotions (Ekman, 1982). More recent work has tried to find out what are, besides Ekman's "display rules", the factors that affect people's decision to display their emotions or not. The effect of these factors on the construction of multimodal agents, and the differences in how people express their different attitudes to different interlocutors are the object of more recent research (Carolis *et al*., 2001; Prendinger & ishizuka, 2001). The use of euphemism and of indirect expression of affect-inducing contents is also shown in discourse generation systems (deRosis & Grasso, 2000).
(c) Multimodal Communication. After the (aware or non-aware) decision to communicate some content, the agent must plan how to communicate it and choose, first, the elements, in whatever modality, that best convey this information, and then their possible combination: the agent in fact

may communicate each single item of meaning either in one or another modality, or in more modalities at the same time. So both the elements of the communication systems in all modalities have to be studied, and the way these elements may be produced simultaneously or in sequence. Research on what is presently called multimodality heavily leans on much of the research on so-called non-verbal communication. In this field, work has been done both on single modalities and on the relationship between the different modalities. Milestones that are quoted recurrently in all papers, both in the simulation era and before, include for example the names of Kendon and McNeill for gestures (Kendon, 1980; McNeill, 1992); Ekman and Friesen (Ekman, 1982); Chovil (1991) and Fridlund (1994) for facial expression, Argyle and Cook (1976) and Beattie (1981) for gaze, Scheflen (1973) and Mehrabian (1969) for body posture, Hall (1966) for proxemic behaviour. About the synchronization of different modalities, especially word and gesture, important works are those by Kendon (1993); Duncan (1974); Rimé and Schiaratura (1991); Knapp (1972), Condon and Ogston (1967); Poggi and Magno Caldognetto (1997). More recent work is being done to analyse specific multimodal behaviour in view of creating autonomous multimodal agents. Some good examples of this kind are the works by Cassell (2000; Cassell *et al.*, 1999); Lester (2000); André and Rist (2000). But more work has to be done yet; in the next section we outline what in our view should be the goals of research to be reached in order to construct an interactive multimodal embodied agent.

4 Multimodality in humans

To state the research steps needed to reach a model of multimodal communication, we can see the human body as composed of various parts, each of which bears its specific communicative repertoire. In fact, a human-like agent uses several communication systems, residing in different body organs. The head produces head movements; the eye region (eyebrows, eyelids, eyes) produces gaze; the mouth, words and prosody (acoustically) and visemes, smiles, grimaces (visually); shoulders, arms and hands produce gestures, trunk and legs postures, movements, orientations and proxemic signals. All these signals may be produced at once in multimodal communication. Therefore, two things are worth studying in order to produce multimodal agents:

1. the structure of each communication system by itself; that is, it is necessary to provide descriptions of the different mode-specific languages of the body;
2. the rules stating how the agent chooses to convey meanings via one or the other modality, and how signals in different modalities synchronize or anticipate each other.

4.1 Mode-specific languages

Let us start with the first issue. In our hypothesis, each system of signals makes up a "mode-specific communication system", that is, a set of rules to link signals to meanings. The links between signals and meanings may be either "codified" or "creative" (Magno-Caldognetto & Poggi, 1995; Poggi & Magno-Caldognetto, 1997). In the former the signal–meaning link is coded in memory (for example in words or symbolic gestures) on a biological or cultural basis (say, both the biologically determined gesture of shaking fists up and the "Churchill" gesture for victory mean happiness for an achievement), and a whole set of these links makes a "lexicon". In the latter, what is coded in memory is only a small set of inference rules about how to create a new signal starting from a given meaning, or about how to retrieve a meaning from a given signal (like in mime or in the creation of new words in natural languages).

A first task to accomplish in order to make artificial agents that really communicate multimodally is, then, to find out the rules that link signals to meanings. For "creative" communication systems one

has to find out the inference rules that state how new signals may be created by a speaker and understood by an addressee. Some studies accomplishing this task are, for example, McNeill (1992); Magno Caldognetto and Poggi (1995); Sowa and Wachsmuth (2001); Yan (2000); Cassell (to appear).

According to our hypothesis, though, many of the systems of communication above are of a lexical kind: not only words or symbolic gestures, as it is generally accepted, but also other kinds of gestures (say, beats or affect displays) and even gaze, facial expression, posture shifts, have each a precise meaning that is coded in the speakers' memory: in our opinion, it is only thanks to this that we can understand each other. In fact, each signal in each modality can be attributed some meaning, and this meaning can be restated in a verbal language. For instance, the gesture of shaking fists up means "I am exulting"; a beat, "this is the comment of my sentence"; a frown may mean either "I am worried" or "I am angry at you"; a posture shift means something like "I am shifting to another topic", and so forth. To the objection that non-verbal signals are the domain of homonymy, polysemy and vagueness, one could object that also for apparently very different meanings of the same signal a common core meaning can be found in all its occurrences, as has been shown by (Kendon, 1995; Poggi, 2001; Poggi and Pelachaud, 2000).

For "codified" systems, that is, "lexicons", the task is then to compile lexicons of the systems in all modalities. Some examples in this field are the dictionaries of sign languages and, more recently, the flourishing of dictionaries of symbolic gestures for many different cultures (Morris, 1977) all over the world: Morris *et al.* (1979) for the Mediterranean area; Tumarkin (2001) for Japanese gestures; Kreidlin (2001) for the Russian; Payratò (1993) for the Catalan; Posner and Serenari (2001) for Berlin gestures; Poggi (2001) for the Italian.

But we can write down lexicons for other systems also; Ekman and Friesen's FACS (1978), for instance, can be considered a lexicon of the face; a sketch of a lexicon of gaze is Poggi, Pezzato and Pelachaud (1999); and also lexicons of very specific systems may be written, like the fragment of lexicon of "performative faces" by Poggi and Pelachaud (1998), the lexicon of the orchestra conductor's gestures (Boyes-Braem & Braem, 1999) and face (Poggi & Mastropasqua, 1999), the lexicon of deictic gestures and gaze (Johnson *et al*., 2000).

A lexicon usually also includes an "alphabet", that is, a set of sub-lexical components, the rules of production of signals that, variously combined simultaneously or in sequence, form all the possible signals of the lexicon.

A task somehow included in the construction of lexicons is, then, the discovery of "alphabets" of non-verbal systems. In this view, important examples are Laban's notation (Laban & Lawrence, 1974) and Birdwhistell's (1952) system. One seminal work is Stokoe (1978), who proposed the notion of "formational parameters" of signs in sign languages of the deaf. He found out that each sign is produced by a particular handshape, a movement, a location, to which orientation was then added. Formational parameters have since then been discovered in different gestural systems of the Hearings Impaired: Calbris (1990) found them for French gestures, Kendon (1988b) for the Australian Aboriginal sign language, Sparhawk (1978) for Persian gestures, Romagna (1999) for Italian gestures, Ekman *et al.* (1969; Johnson *et al*., 1975) for American gestures. Another system for coding and annotating non-verbal items is MPEG-4 notation (Doenges *et al*., 1997).

More recently, the notion of formational parameters has been applied to gaze (Poggi *et al*., 1999), by singling out the parameters and values that pertinently describe each item of gaze: eye direction, eye opening, humidity, eyebrow movements and so on. Finally, formational parameters were found also to describe the communicative gestures of touch (Poggi *et al*., in preparation).

4.2 Signal synchronization

The other important issue to study before constructing multimodal agents is to assess how the signals coming from the communication systems of the different modalities mix up in real interaction. Just thanks to the fact that we are endowed with communication systems in different modalities, we can use multiple signals at the same time: we may in the same instant utter a word, move our trunk towards our interlocutor, look at him and while raising our eyebrows, and open and drop hands. What

determines which signals we will perform at each moment of our discourse? How do we choose to communicate some content through words or other signals, or through both?

Various scholars have dealt with the relationship between speech and other modalities; see Kendon (1988a, 1995, 1997a, b); Freedman (Freedman & Grand, 1977); Rimé & Schiaratura (1991); McNeill (1992); Ekman (1979).

Some factors affecting the choice and synchronization of signals are: the presence or absence in the different modalities of a signal apt to convey the intended meaning, the likeliness of occurrence of that signal as opposed to others or the appropriateness of a specific signal in the particular context at hand.

4.3 How to study mode-specific languages and multimodal synchronization

Both the task of constructing lexicons of mode-specific languages and the task of studying choice and synchronization rules require a wide use of annotation systems and tools to analyse real videotaped data. Some useful tools to transcribe real data are Media Tagger (Max Planck Institute for Psychologistics); Com-Trans (Ingenhoff & Schmitz, 2001); Poggi and Magno Caldognetto (1996); some new ones are presented in this Workshop (Martin *et al.*, Kipp, see below). These tools are useful because they allow studying the precise timing of signals against each other, thus providing relevant information about the planning and distribution of multimodal signals.

5 A workshop on multimodal embodied agents

Let us see how the papers in the Workshop on "Multimodal communication and context in embodied agents" fill in the slots of the research outlined above.

Some of them represent research at all the steps outlined above, thus covering the whole route from the theoretical and empirical study of personality and emotions, and the discovery of specific mode languages, to the implementation of existing systems.

Others focus more on empirical research directly applied to the construction of agents, on the annotation of empirical data or on the evaluation of existing agents. All contributions, though, even when focusing on a particular aspect, are pertinent to the implementation of systems.

The contents of the present Workshop follow the topics outlined above. Many of the papers tackle more than one of these topics, because they present a complete system architecture. Yet, we have decided to cluster the papers based on the workshop topics.

5.1 Basic research and annotation on communicative behaviour

As for the basic research needed to create a multimodal communicative repertoire to implement in an agent, the paper by Cassell *et al.* is a good representative of basic empirical research specifically devoted to the simulation of particular multimodal behaviours. Kipp and Martin *et al.* present coding schemes for annotating gestures, designing tools for coding, viewing and analysing multimodal interactions, while Cassell *et al.* and Pelachaud and Poggi present two ways of finding out multimodal repertoires.

5.1.1 Jean-Claude Martin, Sarah Grimard, and Katerina Alexandri, "On the annotation of multimodal behavior and computation of cooperation between modalities"

In particular, Martin *et al.* propose a standardised coding scheme to encode multimodal corpora producing XML descriptions. The work of Martin *et al*. is embedded in the TYCOON theoretical framework, which allows one to observe, modify and define the cooperation type between modalities during multimodal human-machine interactions: for each multimodal session, the coding scheme annotates speech, gesture, synchrony and the state of the graphics modality. Starting from an example of a multimodal session, the authors explain their coding scheme and the several steps involved in a coding session. Moreover, once a session has been annotated it can be analysed using several

measurement metrics, which compute statistical information on the number of referenceable objects in each modality, on the number of multimodal references and so on. An important feature of TYCOON is the notion of salience of a referenceable object in a given modality, a parameter that defines how a given object is explicitly referred to in a given modality. Statistics of the salience of references to an object across different modalities and statistics of the salience of references in a modality across all objects can also be computed. This approach allows the exchange of annotated documents in an easy way, which responds to today's requirements.

5.1.2 Michael Kipp, "From human gesture to synthetic action"

Among the papers focusing on the annotation of communicative behaviour, Kipp illustrates AnViL (Annotation of Video and Language), a tool for annotation of gesture and posture. In this tool, three hierarchical levels of organization are distinguished for gesture annotation: gesture-phase, gesture-phrase and gesture-group. Based on the previous literature, gestures are classified into 5 categories (emblems, adaptors, deictics, illustrators and beats), each described as for different categories (name, form, function, concomitant speech, verbalisation, other gestures one can be confused with). Within posture two positions are distinguished (upright and relaxed) and four kinds of movement (legs-cross, legs-uncross, upper body and whole body). The AnViL tool, which is also provided with bookmarks, an online coding manual, view options and file import for speech transcription and rhetorical structure classification, is applied to the analysis of the German TV show *Literary Quartet*, where four persons present newly published books. Gestures and postures are annotated and synchronisation is assessed to the syntactic structure of concomitant speech. The analysis shows a high inter-judge reliability of the annotation and provides interesting information on the occurrence of gestures in different subjects, their typological distribution in different categories and their idiosyncratic combinations. A repertoire of 20 gestures resulting from this research is then implemented in "Book Jam", a book-sales scenario with a book-selling agent presenting books to two buyers, where gestures and posture are generated in the different nodes or leaves of the communicative plan.

5.1.3 Justine Cassell, Yukiko I. Nakano, Thimothy W. Bickmore, Candace L. Sidner and Charles Rich, "Annotating and generating posture from discourse structure in embodied conversational agents"

A detailed case of empirical research aimed at agent implementation is Cassell *et al*. Here, the communicative import of posture shifts is studied thoroughly by substantive empirical research, and then implemented in the Real Estate Agent (REA). In order to assess occurrence and function of posture shifts, monologues and dialogues were videotaped where subjects respectively describe rooms or give directions to their house, or discuss a possible common task. In the transcripts, segment boundaries and turn boundaries were coded, as were posture shift start and end points, and estimated energy. Results show that posture shifts occur more frequently at turn boundaries than within turns, much more frequently at discourse boundaries than within discourse, and are both more frequent and more energetic in monologues than in dialogues. Curiously enough, listeners also perform many more posture shifts when the speaker changes the topic than when he does not. The results of the research are then applied to implement posture shifts in REA, an embodied conversational agent that interacts with users: it takes input from a microphone and two cameras, and its user model interprets this multimodal input and sends the resulting semantic representation to a Dialogue Manager, Collagen, which updates a model of the discourse state (focus attention, interaction history, agenda of next possible actions). On this basis, the REA agent plans when to perform posture shifts – for example, during a topic change REA will shift posture in 26% of cases; only 8% of cases when the topic is the same. Cassell *et al.* provide a good example of how collecting, analysing and exploiting empirical data from human-human interaction with regard to verbal/non-verbal inter-dependencies (timing, co-occurrence, context) can best lead to the implementation of realistic embodied agents.

5.1.4 Catherine Pelachaud and Isabella Poggi, "Multimodal believable agent"

On the other side, Pelachaud and Poggi, arguing for the necessity of writing down the modes of the specific lexicons, propose to start from a deductive taxonomy of meanings: guess what are in principle

all the possible kinds of information an agent may need to provide to other agents for its adaptive goals, and then see how they are conveyed in the different modalities. They distinguish three classes of information: on the world, on the speaker's identity and on the speaker's mind. Information on the world concerns the concrete or abstract events a speaker communicates about, their actors and objects and time and space relations among them, and it is provided mainly through words and syntactic structure, but also by deictic, iconic and symbolic gestures, and even by gaze, voice or head or body movements, for instance in pointing at things or persons by eye or chin direction, in squeezing eyes to mean "small" or "difficult", lengthening a vowel to mean "long", in referring to somebody by imitating his movements. Information on the speaker's identity (sex, age, socio-cultural roots, personality) is provided through physiognomic traits, acoustic parameters of voice, and posture. Information on the speaker's mind concerns the speakers beliefs, goals and emotions. Information on the speaker's beliefs include the degree of certainty of the beliefs being mentioned (frowning to mean "I am serious in stating this", opening hands to mean "this is self-evident") and the source of the beliefs mentioned (memory, inference, communication), communicated by looking up when making inferences or snapping fingers while trying to remember. The goals the speaker mentions are the performative of his sentence, conveyed by performative verbs, intonation or performative facial expression, topic/comment distinction, conveyed by batons, eyebrow raising, intensity or pitch of tonic vowel; discourse rhetorical relations (counting on fingers, or repeating the same intonational contour in listing; posture shift to convey topic shift); turn-taking and back-channelling (raising hand or gaze for asking turn; nodding for providing back-channel). Information about the speaker's emotions is conveyed by affective words and gestures, intonation, facial expression, gaze and posture. Pelachaud and Poggi argue that starting from this deductive semantic taxonomy is a useful way to start building "mode-specific" lexicons that embodied agents may be endowed with.

5.2 From believability to expression

The papers by Allbeck and Badler, by Marsella, Gratch and Rickel, by Traum and Rickel, and by Heylen and Nijholt are interesting examples of how one can set up a dialogue model and cognitive model of the agent to make it more believable.

5.2.1 Jan M. Allbeck and Norman I. Badler, "Consistent communication with control"

The topic of what are the requirements to make an agent believable is tackled by Allbeck and Badler, who illustrate the multiple variables entering in the communication process. They set up the background to derive consistent behaviour from a cognitive model of the agent. After defining "believability" as meaning "accepting as real", the authors introduce the notion of consistency in an agent's behaviour. They describe its different aspects: on the one hand the agent's behaviour should respect what we would expect from its human counterpart; on the other hand the agent should behave as an individual in particular. Moreover, the authors are precise that their goal is to derive this consistent behaviour from cognitive models of the agent's mind. To this effect they propose a parametrized agent model that would act with consistent behaviours. The first step of such an elaboration is to look at the effect cognitive processes have over the display of verbal and non-verbal signals, that is, how our age and status or even our gender affect our behaviour. They report on how gaze pattern or interpersonal distance, for example, are very culturally dependent; the social role one desires to pursue often orients us towards a given acting role that is learned and well defined: we expect a doctor to look neat and to talk in a precise manner; but we will not expect the same thing from a mechanic. Not respecting these rules violates our expectations and may put us in a state of alarm. Our emotion, mood and personality are also very important factors that affect our behaviour. Moreover, the context, defined in this paper as everything that we can perceive (object, people, events of a situation), is extremely important in shaping our expectations of how one should behave in a particular situation. But context is very difficult to represent as it requires us to consider all events, actions and so on but also to establish what are their effects over our feeling and our memory. Once we have established the various cognitive processes that intervene in the determination of verbal and non-verbal behaviour of

an agent a major problem arises: how these processes interact with each other, how they influence one another. An agent, to appear believable and consistent, must respect the same expectation we would have from a human; moreover it should be consistent throughout the different facets of what makes an individual.

5.2.2 Stacy Marsella, Jonathan Gratch and Jeff Rickel, "The effect of affect: modeling the impact of emotional state on the behavior of interactive virtual humans"

The papers by Traum and Rickel and by Marsella, Gratch and Rickel cover the whole route from the spring of an emotion to the way it is expressed and influences action. They present a virtual world for the training of army members on peace-keeping missions. A realistic scenario is shown where soldiers and other people have to make important decisions, and some characters influence others also by expressing emotions. Three previous systems are integrated to reach this goal: Steve, a system where virtual humans in 3D simulated worlds interact with humans as mentors and teammates; Emile, focused on emotional appraisal, that is, on the way events in the environment clash with agents' plans and goals; IPD, that models the impact of emotional states on physical behaviour. All of these capabilities are applied to a scenario in which a troop of soldiers in Bosnia injure a boy in an accident and the boy's mother asks them not to leave her there but to help her son. The mother and soldiers are Steve agents, with sophisticated inference capacity, but also with the emotional appraisal capacities of Emile, as well as the expressive skills of IPD. At the same time, methods for representing and updating the discourse state and social obligations arising from discourse are incorporated in the system, thus providing a very sophisticated and complete system. As for the expressive side of the system, the authors aim at creating a rich, emotionally consistent body: it is not plausible, for example, that an emotionally depressed character, quite inattentive, also averting gaze, hugging himself, suddenly starts using energetic beats. To confront this problem, Marsella *et al.* propose the Physical Focus Model, inspired by Freedman (1972) and Lazarus (1991), which distinguishes: (i) body focus mode, characterised by self-focused attention and self-touching gestures, inhibited verbal behaviour and pauses, typical of depression and guilt; (ii) transitional focus mode, characterised by a closer relation to the listener, hand-to-hand or hand-to-object gestures; (iii) communicative focus mode, using the full range of communicative gestures, often exaggerated, with more attentiveness and responsiveness to the environment. The character shifts from one mode to another based on its current emotional state: the more sad or guilty, the more it shifts to body focus.

5.2.3 David Traum and Jeff Rickel, "Embodied agents for multi-party dailogue in immersive virtual worlds"

Within the same complex project, whose general frame and the aspects above are shown in detail by Marsella *et al.*, Traum and Rickel more specifically focus on the communicative aspects of conversation, by considering them as fields in a database: who is talking and by what media, what information is added to the common ground, what are the social commitments (prescribed and forbidden actions) holding among participants, and how they come to agree on them. Each of these fields is modelled in detail, and the various ways to fill them in are outlined. Traum and Rickel's dialogue model is able to simulate dialogue between two or more participants. For each discussion among agents the authors propose several conversation types, such as initiating a speaking turn, negotiating about something with other agent(s) or even getting the attention of other(s). Flashes of a sort of verbal and non-verbal lexicon are provided: for the action request-turn, you can open your mouth, raise one hand, or avoid the speaker's gaze at phrase boundaries; for assign-turn you can use a vocative or gaze at the interlocutor.

5.2.4 Dirk Heylen and Anton Nijholt, "Designing and implementing embodied agents: learning from experience"

Heylen and Nijholt draw conclusions from their experience in the creation of several embodied agents and propose some technical considerations to follow when creating embodied agent systems. Their

group of researchers have built a virtual environment (web-based 3D VRML world) inhabited by several "intelligent" agents that can converse face to face among themselves. Many of the projects developed by the research group take part of the educational program. That means they have to deal with students that need to learn the system architecture in a short time in order to embed their project in the 3D world easily. After setting the problem of their research group, Heylen and Nijholt illustrate some of the problems they are facing by presenting some projects. The first one is Karin. Karin is a simple 3D cartoon-type face that is embedded in a natural language information system (SCHISMA). This dialogue system is able to inform users about theatre performances and make reservations. Karin is able to dialogue with a user using TTS as speech output and by specifying codes for non-verbal behaviour. The 3D model of Karin has been created within 3D Studio as has a set of facial expressions. A major emphasis of this project was to respect the real-time constraint for the dialogue to be as natural as possible. The second project described in the paper is Gina, a part of a software engineering assignment where a group of 4 to 5 students has to work together in a 10-week project. Gina is a very modular system, with 4 main components: GinaFace, GinaMuscleController, GinaTTs and GinaChat. The muscle controller engine, GinaMuscleController, maps tags specifying for non-verbal behaviours into facial parameters; the graphics engine, GinaFace, then generates the appropriate animation, and the tags are provided by GinaChat. GinaTTS maps the text output into synthesized speech but it also computes the appropriate visemes that are passed to GinaFace. From this project, the authors draw the conclusion that to ensure a better synchronization between modules it is better to be system-dependent. Using existing standards allows for a better portability of the system and its reuse in other projects. Moreover, for web-based applications, in order to respect as much as possible the real-time of an application, it is better to keep software requirements on the client side at a minimum.

5.3 Social relationship

The social relationship between agent and user is taken into account by Guerrin *et al.* and by Prendinger and Ishizuka; a topic which is relevant in determining how to assess, interpret and affect responsiveness to the context in which the interaction takes place. Traum and Rickel's dialog model is able to simulate dialog party between two or more participants. For each discussion among Agents the authors propose several conversation types, such as initiating a speaking turn, negotiating about something with other Agent(s), or even getting the attention of other(s).

5.3.1 Frank Guerrin, Kaveh Kamyab, Yasmine Arafa and Ebrahim Mamdani, "Conversational sales assistants"

The work presented by Guerrin at the workshop is part of a larger project, SoNG (PortalS of Next Generation), which foresees the use of 3D virtual places populated with believable embodied agents. As a step towards such portals, Guerrin and his colleagues describe a system where an agent, a shopkeeper, is able to interact with a user, to take the initiative by presenting the user with objects from its virtual shop or, for example, by pointing at objects in its shop. An interesting aspect in their system is the modelling of the context awareness which gives the 3D agent the capacity to draw inferences about the user's preferences and to build up a user model. This capacity allows him not only to adapt his behaviour to each user's preferences but also to foresee the user's need. In the notion of context awareness the authors have also embedded the notion of social role: an agent is conscious of his social role (for example of being a shopkeeper) and would behave as his social role shapes him to do. Their system works as follows: the user may type plain text or may select pre-defined gestures for their avatar; the shopkeeper agent has to interpret these verbal and non-verbal behaviours by also by considering contextual information such as the domain of the conversation or the relation between the current communicative and previous ones. Having interpreted the user query the agent plans his communicative acts, and verbal and non-verbal communicative acts are synchronized through a CML (Character Mark-up Language) script which is then rendered to produce MPEG-4 animation parameters to animate the agent's face and body.

5.3.2 Helmut Prendinger and Mitsuru Ishizuka, "Communicative behavior of socially situated agents"

Another paper that takes social role into account is by Prendinger and Ishizuka, who, in modelling affective communication, argue for the distinction between the emotional state, which is the result of reasoning about the events and the agent's goals, standards and attitudes, and the emotion expression, the result of reasoning on the agent's emotional state while also taking into account the agent's personality or mood and the social context. If someone crashes your computer, you might get angry and tend to express you anger; but what if it was your boss who did it? Will you express your anger anyway? To model the emotional state Prendinger and Ishikuza propose an affective reasoner based on Ortony (1988) and Elliott (1992), while modelling in addition the intensity of the emotion through logarithmic combination. To deal with the emotion expression, they propose a filter program that takes into account personality and conventional practices. As to the former, the program attributes to the modelled agent a numerical quantification of personality dimensions, like extroversion or agreeableness; for the latter, it takes into account behavioural constraints associated with the agent's social role (responsibilities, rights, duties, prohibitions and possibilities) and communicative conventions. They claim, first, that emotional expression is based on the agent's social awareness, which encompasses consideration of possible roles, ordered according to a power scale that attributes a rank to each agent, the social network the agent is in and the social distance between agents. Second, communicative conventions hold according to which communicating negative emotions may introduce disharmony in the interaction by threatening the other agent's public face. Therefore, before expressing emotions, based on its own social awareness of the values of social power and social distance, the agent computes the threat this expression could cause to the other agent, and if the computed value is high it is likely to suppress the emotion expression. Filter rules are presented that combine the computed threat with the agent's personality and give as a result the likeliness with which it will express its emotion. The authors also consider social dynamics factors, that is, the fact that interaction may change over time due to i) reciprocal feedback loop, where an agent's use of a polite style may lead the interactant to do the same, and to ii) social feedback, whereby social distance and power relations may change with recurrent interaction.

5.4 Emotion and context evaluation

Another aspect towards the creation of a believable agent is the relevant issue of emotion, its modelling and its communicative output, which is exploited in Silva *et al*. Consideration of the context when computing appropriate behaviour is presented in the paper by Huang *et al.*

Marsella et al. propose an approach where emotion appraisal, influence of the emotion on the Agent's verbal and nonverbal behaviour, and the communicative intention of the agent are encompassed in a single and complex model. Another approach taken by Silva *et al.* is to script the text a storyteller is saying with emotion tags and other communicative tags.

5.4.1 André Silva, Marco Vala and Ana Paiva, "The Storyteller: building a synthetic character that tells stories"

The Storyteller is a first approach towards a more ambitious project: the creation of a believable storyteller able to express emotions as he develops the story, where emotions should affect the verbal and non-verbal behaviour of the storyteller. At present, the authors have built a simple storyteller whose behaviour is scripted using control tags. Several kinds of tags are available: one tag is used to define the virtual world where the storyteller is; other tags define the type of illumination of the virtual world, day or night lighting, that allows the user to create different atmospheres. Several sets of tags have been designed to shape the character of the storyteller, among which one represents his emotional state as a discrete value (e.g. happiness, surprise). Emotion affects not only the facial expressions of the agent but also his voice: the Storyteller system uses text-to-speech technology, whose voice parameters correspond to the TTS set of vocal parameters and pauses and text punctuations are also explicitly defined. The last two sets of tags represent facial expressions of emotion and gesture. The

value of each tag may be changed dynamically as the story is being developed. The conclusion of the authors is that current TTS technology does not allow complete flexibility to model affective speech: a more complex graphics engine is needed to combine gestures and have a better synchronisation scheme between lip movement and speech; but all in all the architecture and module design reached the intended purpose.

5.4.2 Zhisheng Huang, Anton Eliëns and Cees Visser, "The programmability of intelligent agent avatars"

Huang *et al.* propose a cognitive model of a soccer player based on an evaluation of the context of the match, where the player is able to compute its next move. The authors first propose a taxonomy of web agents, dividing them into 2D versus 3D, client versus server, singularity versus multiplicity. The authors concentrate then on 3D web agents. These agents are able to interface with 3D virtual worlds built using VRML. Several primitives have been implemented that enable the system to get information on the virtual world. These primitives act as sensors and effectors. Some examples are: *getPosition* of a given object, *Load* a given URL. In order to build intelligent web agents the authors developed distributed logic programming (DLP) that combines logic programming, object-oriented programming and parallelism. To test DLP the authors developed a benchmark example: soccer playing. An intelligent soccer avatar has the following properties: it cooperates with the other avatars; it has a cognitive model of the soccer game; and it knows the physics of the soccer ball. There exist two types of soccer player avatars: soccer player and soccer player user. The first is controlled by a software agent while the last is manipulated by a user. For each avatar type, different roles are available: goalkeeper, defender and so on. Each agent knows the rules of the soccer games, and it also knows some critical distances: e.g. goal distance, defined as the distance between an avatar and the goal. The application has been developed using multi-threads and therefore is able to control multiple players; therefore this application has shown that DLP and the use of multi-threads enable the creation of distributed intelligent agent avatars.

5.5 Computational model of facial expressions

Once one has represented an emotion model and a complex dialogue planner, one needs to address the problem of computing the behaviour (facial expression, gaze and gesture type, body movement) the agent should display.

5.5.1 Zsófia Ruttkay and Han Noot, "FESINC: Facial Expression Sculpturing with INterval Constraints"

Most of today's applications are based on a simplistic approach of display computation producing unnatural animations: expressions are defined symmetrically, no dynamism of the expression is included, all the behaviours computed by the system are linearly combined and so on. To avoid such an approach, Ruttkay and Noot have developed a constraint system that allows expressions to appear, remain on the face and disappear in a non-systematic way, but nevertheless always within pre-defined limits: for example, simultaneous expressions may differ, in their degree of symmetry, on their onset time, thus creating a more natural animation. The authors characterise each expression by 3 parameters – application, release and relaxation – which correspond to the display life of an expression; these parameters can be viewed as intervals between time values (start of application, start of release and so on). The authors use piece-wise linear interpolation between these time values that may be considered as control points of a curve; then, constraints may be defined on any parameters. Examples are the time range, time duration and relative time duration of a parameter; but also the value range, value change and relative parameter change speed. Such constraints may be specified interactively by the user when building up the animation. The resolution of multiple constraints is obtained by interval propagation that updates the domain values of the control points. This method has been extensively tested on 2D characters. The authors propose that each reader builds her own smile to test the program.

5.5.2 Aldo Paradiso and Marcello L'Abbate, "A model for the generation and combination of emotional expressions"

The work presented by Paradiso and L'Abbate is part of the European project COGITO which aims at improving consumer-supplier relationships for e-commerce. The authors employ a chatterbot called eBrain and expand it with a 3D expressive agent, an automatic expression generator and machine-learning techniques. During a dialogue session between a user and the agent, keywords from the user's typed sentences are mapped to emotions and to comments (occurring on emphasised words). The context, defined in their paper as the topic of a discourse, is represented as a set of rules that controls how the user's input is coherent with the dialogue going on. Coherence is then mapped to facial expression (e.g. facial expression of doubt may signal low coherence between successive dialogue acts). The type of user is also a parameter to control the behaviour of the agent in the chatterbot. The authors go on by deriving a computational model to combine facial expressions: using the Facial Animation Parameters (FAPs) defined by MPEG-4, they define a set of operators to combine FAPs and an algebraic model to create and control an animation. Through their operators, facial expressions may overlap with each other on the whole face or on a given region. Several computation formulae are proposed by the authors.

5.6 Evaluation

Finally, the problem of how to evaluate multimodal interaction with an embodied agent has to be faced.

5.6.1 Helen McBreen, James Anderson and Mervyn Jack, "Evaluating 3D embodied conversational agents in contrasting VRML retail applications"

McBreen *et al.* designed an empirical setting to evaluate the impact of different types of agents (male versus female, casual versus formal dressing), in several applications (a cinema box-office, a bank and a travel agency). The agents have been created using existing 3D software and were placed into a virtual environment. The agent's body follows H-Anim specification and by altering joint values it was possible to create four types of gesture for each agent, namely nodding, waving, shrugging and typing. In the experiment setting a user would engage in a real-time face-to-face conversation with one of the agents. The authors made five predictions before the experiments: (1) the users would accept as plausible the agent in the role of assistant in retail spaces; (2) users would enjoy interacting with the agent in any of the three applications; (3) the stereotyped appearance of the agent would play a role only in given applications; (4) there would be no difference when varying the gender of the agent in the users' responses; (5) the user would perceive the agent's personality as invariant over the different applications. The authors were precise that the first four predictions were supported by the results of their experiment, but even though all the agents were perceived as having similar personalities in all three retail applications, the trustworthiness of the agents differed between the applications. The authors end their paper by addressing the delicate problem of trust: does the user trust the agent, is the user ready to let the agent make some decisions and perform some transactions on their behalf?

5.6.2 Melanie Baljko, "The evaluation of microplanning and surface realization in the generation of multimodal acts of communication"

Another paper focused on the evaluation of embodied agents is Baljko's. A communicative plan, that is, a complex illocutionary act, can be accomplished through many possible different locutionary acts. Therefore, a task in constructing agents that communicate multimodally is to state how, given a particular communicative plan, a surface realization is derived for it; this is what Baljko calls MP&SR – "Micro-Planning and Surface Realisation" (what we mentionned as the issue of multimodal signals synchronisation): a planning that will be different for NLG and for a multimodal agent. Baljko presents research for the evaluation of MP&SR in a particular domain: the modelling of human-human communication in a context where an interlocutor has an expressive communication disorder, no functional speech or writing and uses an Augmentative-Alternative Communication (AAC) device for

the generation of synthesized speech. Arguing that the criteria of evaluation generally used for NLG are not sufficient for an MP&SR module of a multimodal agent, Baljko discusses possible specific criteria for evaluating specific tools in this domain: resemblance to human behaviour, consistency across modalities and consideration of intermodal interactions and adaptation to the context. Moreover, evaluation may compare different interfaces, some with embodied conversational agents and some not, or just compare different embodied conversational agents interfaces. She presents a study aimed at the evaluation of the MP&SR in a communicatively impaired agent. A set of videotaped interviews in a job application scenario were coded as for eye gaze, gesture and speech: in one condition the AAC device was a communication board, in the other a voice-output communication aid. Coding was refined through inter-judge discussion. In future research the refined coding method should be applied to a multimodal version of the classical Referential task, where a subject has to describe an object to an interlocutor. In conclusion, evaluation of MP&SR in a speech-impaired domain should assess whether the allocation of information through different modalities is the same in a virtual impaired agent as in an impaired human, and if the capacity for surface realization may evolve in the virtual agent in the same way it does in the human, for instance by adapting to the specific interlocutor and context and hence by exhibiting varying degrees of multimodality.

Baljko designs an iterative process where evaluation is performed and its findings gives rise to the refinement of the Agent model; as iterations grow it will offer to the system a mean to compute more coordinated and communicative behaviours for a more Believable Agent.

6 Conclusion

The aim of the workshop was to establish the various steps that are needed to build a believable interactive embodied agent. From the papers presented at this workshop we can summarise that research aimed at creating BIEAs must include three phases. First, a phase of empirical research aimed at finding out the regularities in the mind and behaviour of human-like agents and at constructing models of them. Second, a phase of modelling believable interactive embodied agents, where the rules found out are formalised, represented and implemented in the construction of agents. Third, a phase of evaluation of the implemented agents, aimed at testing how they fit the user's needs and preferences and, especially in a simulative approach, how similar they look to a real human agent. The papers presented in this workshop cover most of these steps. Their themes range from the elaboration of empirical research and annotation tools to the elaboration of emotion models and dialogue managers. Papers showed the theoretical aspects of the problem as well as the implementation details. They offer a multi-faceted and broad view of the subject and provide important building blocks towards future research.

References

Andre, E, Rist, T, van Mulken, S, Klesen, M, and Baldes, S, 2000, The automated design of believable dialogues for animated presentation teams. In J. Cassell, J. Sullivan, S. P. and Churchill, E. (eds) *Embodied Conversational Characters*. MIT Press, Cambridge, MA.

Argyle, M, and Cook, M, 1976, *Gaze and Mutual gaze*. Cambridge University Press.

Ball, G, and Breese, J, 2000, Emotion and personality in a conversational agent. In J. Cassell, J. Sullivan, S. P. and Churchill, E. (eds) *Embodied Conversational Characters*. MIT Press, Cambridge, MA.

Beattie, G, 1981, Sequential temporal patterns of speech and gaze in dialogue. In Sebeok, T. and Umiker-Sebeok, J. (eds) *Nonverbal Communication, Interaction, and Gesture*, pages 297–320. The Hague, New-York.

Birdwhistell, R, 1952, *Introduction to kinesics, an annotation system for analysis of body motion and gesture*. University of Louisville.

Bok, S, 1982, *Secrets. On the Ethics of Concealment and Revelation*. Pantheon, New York.

Boyes-Braem, P and Braem, T, 1999, A pilot study of the expressive gestures used by classical orchestra conductors. In Emmorey, K. and H.Lane (eds) *The signs of Language Revisited: An Anthology in Honor of Ursula Bellugi and Edward Klima*. Laurence Erlbaum Associates.

Brown, P and Levinson, S, 1987, *Politeness. Some Universals in Language Usage*. Cambridge University Press, Cambridge.

Calbris, G, 1990, *The semiotics of French gestures*. University Press, Bloomington: Indiana.

Carbonell, J, 1980, Towards a process model of human personality traits. *Artificial Intelligence*, **15** 49–74.
Carolis, ND, Pelachaud, C, Poggi, I, and de Rosis, F, 2001, Behavior planning for a reflexive agent. In *IJCAI'01*, Seattle, USA.
Cassell, J, 2000, Embodied conversational interface agents. *Communications of the ACM*, **43**(4) 70–78.
Cassell, J, to appear, Embodied conversation: Integrating face and gesture into automatic spoken dialogue systems. In Luperfoy, editor, *Spoken Dialogue Systems*. MIT Press, Cambridge, MA.
Cassell, J, Bickmore, J, Billinghurst, M, Campbell, L, Chang, K, Vilhjálmsson, H, and Yan, H, 1999, Embodiment in conversational interfaces: Rea. In *CHI'99*, pages 520–527, Pittsburgh, PA.
Castelfranchi, C, de Rosis, F, Falcone, R, and Pizzutilo, S, 1997, A testbed for investigating personality-based multiagent cooperation. In *European Summer School of Logic, Language and Information*, Aix-en-Provence, France.
Castelfranchi, C, de Rosis, F, Falcone, R, and Pizzutilo, S, 1998, Personality traits and social attitudes in multi-agent cooperation". *Applied Artificial Intelligence*, **12**(7–8) 649–675.
Castelfranchi, C and Poggi, I, 1998, Bugie finsioni sotterfugi. In *Per una scienza dell'inganno*. Carocci, Roma.
Chovil, N, 1991, Social determinants of facial displays. *Journal of Nonverbal Behavior*, **15**(3) 141–154.
Condon, W, and Osgton, W, 1967, A segmentation of behavior. *Journal of Psychiatric Research*, **5** 221–235.
deRosis, F and Grasso, F, 2000, Affective natural language generation. In Paiva, A. (ed) *Affect in interactions*. Springer-Verlag, Berlin.
Doenges, P, Lavagetto, F, Ostermann, J, Pandzic, I, and Petajan, E, 1997, MPEG-4: Audio/video and synthetic graphics/audio for mixed media. *Image Communications Journal*, **5**(4).
Ducrot, O, 1984, *Le dire et le dit*. Les Editions de Minuit, Paris.
Duncan, S, 1974, Some signals and rules for taking speaking turns in conversations. In Weitz, S., editor, *Nonverbal Communication*. Oxford University Press.
Ekman, P, 1979, About brows: Emotional and conversational signals. In von Cranach, M., Foppa, K., Lepenies, W., and Ploog, D. (eds) *Human ethology: Claims and limits of a new discipline: contributions to the Colloquium*, pages 169–248. Cambridge University Press, Cambridge, England; New-York.
Ekman, P, 1982, *Emotion in the human face*. Cambridge University Press.
Ekman, P and Friesen, W, 1969, The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, **1**.
Ekman, P and Friesen, W, 1978, *Facial Action Coding System*. Consulting Psychologists Press, Inc., Palo Alto, CA.
Elliott, C, 1992, *An Affective Reasoner: A process model of emotions in a multiagent system*. PhD thesis, Northwestern University, The Institute for the Learning Sciences. Technical Report No. 32.
Searle, J, 1992, *(On) Searle on Conversation*. John Benjamins Publishing Company.
Freedman, N, 1972, The analysis of movement behavior during the clinical interview. In Siegman, A. and Pope, B. (eds) *Studies in Dyadic Communication*, pages 177–210. Pergamon Press, New York.
Freedman, N and Grand, S, 1977, *Communicative Structure and Psychic Structures*. Plenum Press, New-York, London.
Fridlund, A, 1994, *Human facial expression: An evolutionary view*. Academic Press, New York.
Frijda, N and Swagerman, J, 1987, Can computers feel? Theory and design of an emotional system. *Cognitive and Emotion*, **1**(3) 235–257.
Grice, HP, 1989, *Studies in the Way of Words*. Harvard University Press, Cambridge, MA.
Hall, R, 1966, *The hidden dimension*. Doubleday, New-York.
Ingenhoff, D and Schmitz, H, 2001 (in press), Com-trans: A multimedia tool for scientific transcriptions and analysis of communication. In M.Rector, I.Poggi, and N.Trigo (eds) *Gestures, Meaning and Use*. Universidad Fernando Pessoa, Porto.
Isbister, K and Nass, C, Forthcoming, Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*.
Johnson, H., Ekman, P, and Friesen, W, 1975, Communicative body movements: American emblems. *Semiotica*, **15**(4).
Johnson, W, Rickel, J, and Lester, J, 2000, Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *To appear in International Journal of Artificial Intelligence in Education*.
Kendon, A, 1980, Gesticulation and speech: Two aspects of the process of utterance. In M.R.Key, editor, *The Relation between Verbal and Nonverbal Communication*, pages 207–227. Mouton.
Kendon, A, 1988a, How gestures can become like words. In Poyatos, F. (ed) *Cross-cultural perspectives in nonverbal communication*, pages 131–141. Hogrefe, Toronto.
Kendon, A, 1988b, *Sign Languages of Aboriginal Australia: Cultural, semioti and communicative perspectives*. Cambridge University Press, Cambridge.
Kendon, A, 1993, Human gesture. In Ingold, T. and Gibson, K. (eds) *Tools, Language and Intelligence*. Cambridge University Press, Cambridge.

Kendon, A, 1995, Gestures as illocutionary and discourse structure markers in southern Italian conversation. *Journal of Pragmatics*, **23** 247–279.

Kendon, A, 1997a, Gesture. *Annu. Rev. Anthropol.*, **26** 109–128.

Kendon, A, 1997b, On gesture: Its complementary relationship with speech. In Siegman, A. and Feldstein, S. (eds) *Nonverbal behavior and communication*, pages 65–97. L. Erlbaum, Hillsdale, 2nd. edition edition.

Knapp, M, 1972, *Nonverbal Communication in Human Interaction*. Holt and Rinehart and Winston and Inc.

Kreidlin, G, 2001 (in press), The dictionary of Russian gestures. In C.Mueller and R.Posner (eds) *The Semantics and Pragmatics of Everyday Gestures*. Berlin Verlag Arno Spitz, Berlin.

Laban, R and Lawrence, F, 1974, *Effort: Economy in body movement*. Plays, Inc, Boston.

Lazarus, R, 1991, *Emotion and adaptation*. Oxford Press.

Lester, J, Stuart, S, Callaway, C, Voerman, J and Fitzgerald, P, 2000, Deictic and emotive communication in animated pedagogical agents. In J. Cassell, J. Sullivan, S. P. and Churchill, E. (eds) *Embodied Conversational Characters*. MIT Press, Cambridge, MA.

Magno-Caldognetto, E and Poggi, I, 1995, Creative iconic gestures: some evidence from aphasics. In Simone, R. (ed) *Iconicity in Language*, pages 257–275. John Benjamins, Amsterdam.

Martinho, C, Machado, I and Paiva, A, 2000, A cognitive approach to affective user modeling. In Paiva, A., editor, *Affect in interactions*. Springer-Verlag, Berlin.

McNeill, D, 1992, *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago.

Mehrabian, A, 1969, Significance of posture and position in the communication of attitude and status relationships. *Psychological Bulletin*, **71**(5).

Morris, D, 1977, *Manwatching*. Cape, London.

Morris, D, Collet, P, Marsh, P and O'Shaughnessy, M, 1979, *Gestures. Their origins and distribution*. Cape, London.

Ortony, A, Clore, G and Collins, A, 1988, *The Cognitive Structure of Emotions*. Cambridge University Press.

Paiva, A (ed) (2000, *Affective Interactions. Towards a New Generation of Computer Interfaces*. Springer, Berlin.

Payratò, L, 1993, A pragmatic view on autonomous gestures: A first repertoire of Catalan emblems. *Journal of Pargmatics*, **20** 193–216.

Pennebaker, J, 1982, *The psychology of physical symptoms*. Springer-Verlag, New York.

Picard, R, 1997, *Affective Computing*. MIT Press, Cambridge.

Poggi, I, 2001, The italian gestionary. Meaning representation, ambiguity, and context. In C.Mueller and R.Posner (eds) *The Semantics and Pragmatics of Everyday Gestures*. Berlin Verlag Arno Spitz, Berlin.

Poggi, I and Mango-Caldognetto, E, 1996, A score for the analysis of gestures in multimodal communication. In L.Messing, editor, *Proceedings of the Workshop on the Integration of Gesture and Language in Speech, Applied Science and Engineering Laboratories*, pages 235–244, Newark and Wilmington, Del.

Poggi, I and Magno-Caldognetto, E, 1997, *Mani che parlano. Gesti e Psicologia della comunicazione*. Padova: Unipress.

Poggi, I, Cirella, F, and Zollo, A, In Prep., Touch as a communicative behavior.

Poggi, I and Mastropasqua, M, 1999, Multimodal communication and the conductor's face. In *The Eighth International Workshop on the Cognitive Science of Natural Language Processing (CSNLP-8) "Language, Vision and Music"*, Galway, Ireland.

Poggi, I and Pelachaud, C, 1998, Performative faces. *Speech Communication*, 26:5–21.

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

Poggi, I, Pezzato, N and Pelachaud, C, 1999, Gaze and its meaning in animated faces. In *The Eighth International Workshop on the Cognitive Science of Natural Language Processing (CSNLP-8) "Language, Vision and Music"*, Galway, Ireland.

Posner, R and Serenari, M, 2001 (in press), The emergence of gestures from body movements. In Rector, M., Poggi, I. and Trigo, N. (eds) *Gestures, Meaning and Use*. Universidad Fernando Pessoa, Porto.

Prendinger, H and Ishizuka, M, 2001, Social role awareness in animated agents. In *Proceedings of the 5th International Conference on Autonomous Agents*, Montreal, Canada.

Rimé, B, 1987, Le partage social des emotions. In Rimé, B and Scherer, K (eds) *Les emotions*. Delachaux et Niestlé, Genève.

Rimé, B and Schiaratura, L, 1991, Gesture and speech. In Feldman, R and Rimé, B (eds) *Fundamentals of Nonverbal Behavior.* Cambridge University Press, Cambridge, New York.

Romagna, M, 1999, L'alfabeto dei gesti. parametri formazionali nel lessico dei gesti simbolici usati dagli udenti in italia. Tesi di laurea, Università Roma Tre.

Scheflen, A, 1973, *Body Language and Social Order.* Prentice-Hall, Inc.

Scherer, K, 1988, *Facets of emotion: Recent research*. Lawrence Erlbaum Associates Publishers.

Sowa, T and Wachsmuth, I, 2001 (in press), Coverbal iconic gestures for object dscriptions in virtual environments: an empirical study. In Rector, M, Poggi, I and Trigo, N (eds) *Gestures, Meaning and Use*. Universidad Fernando Pessoa, Porto.

Sparhawk, C, 1978, Identificational features of Persian gestures. *Semiotica*, **24** 49–86.
Stokoe, W, 1978, *Sign language structure: An outline of the communicative systems of the American deaf*. Linstock Press, Silver Spring.
Tomkins, S, 1982, *Affect, Imagery, Consciousness*. Springer-Verlag, 1982.
Trappl, R and Petta, P (eds), 1997, *Creating personalities for synthetic actors: Towards autonomous personality agents*. Spinger.
Tumarkin, P, 2001 (in press), On a dictionary of Japanese gesture. In Rector, M, Poggi, I and Trigo, N (eds) *Gestures, Meaning and Use*. Universidad Fernando Pessoa, Porto.
Yan, H, 2000, Paired speech and gesture generation in embodied conversational agents. Master's thesis, MIT Media Laboratory.