

# Performative Facial Expressions in Animated Faces

Isabella Poggi and Catherine Pelachaud

## X.1 Introduction

In face-to-face interaction, multimodal signals are at work. We communicate not only through words, but also by intonation, body posture, hand gestures, gaze patterns, facial expressions, and so on. All these signals, verbal and nonverbal, do have a role in the communicative process. They add/modify/substitute information in discourse and are highly linked with one another. This is why facial and bodily animation is becoming relevant in the construction of believable synthetic agents.

In building autonomous agents with talking faces, agents capable of expressive and communicative behavior, we consider it important that the agent express his communicative intentions. Suppose an agent has the goal of communicating something to some particular interlocutor in a particular situation and context: he has to decide which words to utter, which intonation to use, and which facial expression to display. In this work, we restrict ourselves only to the visual display of communicative intentions, leaving aside the auditory ones. We focus on facial expressions and propose a meaning-to-face approach, aiming at a face simulation automatically driven by semantic data.

After reviewing the literature on face communication and presenting some existing systems that simulate synthetic agents with talking faces, we focus on the structure of the communicative act and on the notion of the performative. Next, we introduce our model of context and we show how to express the performative by facial expression. Finally, an overview of our system and some examples are provided.

## X.2 The Relevance of Nonverbal Signals in Communication

When talking, we all move our hands, we nod, we glance at our interlocutor, we smile, we turn our head away . . . Gesture, gaze, facial expression, and body orientation all give information about what we are saying and help the interlocutor understand our attitude toward her, our emotion, and our relation to what we are saying (for instance, in emphasis or irony). They may also act as syntax markers: if a person punctuates the end of her statement with a raised eyebrow, her statement could be interpreted as a question. For example, a person is saying *Peter is going to Boston*, and she raises her eyebrow at the word *Boston* and sustains the raised eyebrow during the pause following her utterance; the sentence will be interpreted as the nonsyntactically marked question *Peter is going to Boston?* rather than the affirmation *Peter is going to Boston*. Intonation may also serve as syntactic markers: high pitch at the end of an utterance may mark a question while a statement may be indicated by a low pitch (Bolinger 1989; Pierrehumbert and Hirschberg 1987; Prevost 1996).

Several studies (Argyle and Cook 1976; Ekman 1979; Kendon 1993; McNeill 1992) have shown the importance of nonverbal signals within a conversation. They have reported the different functions of these signals. The meaning of such signals may complement the meaning of what is being said (e.g., showing a direction with the index finger while saying *He went in this street*); they may substitute words (shaking the index finger to a child to say *no*); they may modify the conversation (making a mocking face while saying *You look nice tonight*).

In particular, Bolinger (1989, 211) demonstrates how nonverbal signals may vary the sense of what is being said by interpreting a very common sentence such as *I don't know* when accompanied by different gestures:

Lips pursed: "No comment."

Eyebrows arched: "I'm wondering too."

Shoulders raised: Same.

Head tilted sideways: "Evasion."

Hands held slightly forward, palms up: "Empty, no information."

This very simple sentence may receive a number of varied interpretations. The slightest head movement, eyebrow raising may modify the sense of the utterance. The interpretation of the meaning conveyed by the speaker does need to consider both sets of signals: verbal and nonverbal.

Verbal and nonverbal signals are highly synchronized. They do not occur in a random way; most gestures occur during speech (McNeill 1992). Gestures tend to end with our speech: our hands come to a stop, and we often look at our interlocutor to signal her that she can take the speaking floor. The timing relationship between both sets of signals occurs at different levels of the discourse: change of body orientation and leg posture often happen at a change of topic discussion, while blink is often synchronized at the phoneme level. A microanalysis study of an interaction (Condon 1988) shows the organizational structure of all signals. In the microanalysis of the word *sam* (Condon 1988), it was found that several gestures happen in parallel, and gesture and speech follow the same timing pattern; during the /s/ sound, the head goes down and the eyes close; during the /ae/ sound, the head goes right and the eyes remain closed; and during the /m/ sound, the head goes up and the eyes open.

If, during a conversation, your interlocutor moves only her lips to talk but uses no other signals (no intonation to mark an accent or the end of an utterance, no facial expression, no change in the gaze direction, no hand gesture, and so on), you might soon have the impression of dialoguing with a humanoid rather than a human. Moreover, you might have a hard time understanding what she is saying, since new and important information nor end of turn is not clearly marked in her discourse; moreover no change in the direction of her gaze may soon become embarrassing. Having a person either always fixing her gaze on you or always avoiding looking at you can be a very awkward feeling.

For a sentence such as *I asked Mary to take the brown pot and put it over there*, various interpretations are possible: if no accent is indicated, the sentence could be interpreted either as *I asked Mary and not Charles to take the brown pot and put it over there* or as *I asked Mary to take the brown pot and not the black pot and put it over there*. If no pointing gesture or head direction accompanies the word *there*, it has no meaning: *there* could be anywhere. In the same way, if a listener does not give you any feedback during your speech, you will not know his reactions to what you are saying; you will not know if he understands, agrees, or is interested. It will be like talking to a wall!

Again, not displaying the correct facial expression at the right moment can be a source of misunderstanding and convey the wrong message. Marking a nonaccented word with a head nod can put the focus of the conversation on the wrong information. Suppose that in the sentence *David went to New York by car*, the voice is stressing the word *car* but a head nod occurs on the word *David*: two very different interpretations are possible, depending on which sign (verbal or nonverbal) dominates. If the verbal signal (accent on *car*) prevails, the sentence can be understood as *David went to New York by car and not by train*; but in the case where the head nod on *David* has more weight, the interpretation could be *David and not Peter went to New York by car*. In the former, the new information is the means of transportation David is using to go to New York, while in the latter it is who went to New York. Identically stressing a word and raising the eyebrow at another, one creates asynchrony between verbal and facial channels and causes difficulties in understanding the message conveyed.

These examples show the importance and the role of multimodal signals in a conversation. Signals in different modalities are intersynchronized, and their meanings need to be evaluated in the context they are emitted. We can add that deliberately stressing such a dichotomy of the channels as well as using the wrong signals is extremely difficult to do in a normal conversation.

### X.3 The Role of Face in Multimodal Communication

The face plays an important role in the communication process. A smile can express happiness, be a polite greeting, or be a backchannel signal. Some facial expressions are linked to the syntax structure of the utterance: eyebrows may raise on an accent and on nonsyntactically marked questions. Gaze and head movements are also part of the communicative process (Argyle and Cook 1976; Collett and Contarello 1987; Poggi, Pezzato, and Pelachaud 1999). The sequence of looking at the addressee and of breaking the gaze reflects the social status of the two interlocutors, their degree of intimacy, their culture, and so on. Gaze also helps regulate the flow of speech and the exchange of turns. As the literature on face communication has shown, all these facial signals can be decomposed in the following clusters based on their communicative functions (Duncan 1974; Ekman 1979; Fridlund 1994; Scherer 1980).

#### X.3.1 Affective Display

Different studies have investigated the facial expression of emotions. The facial expressions of seven universal prototypes of emotions have been specified: anger, disgust, fear, happiness, sadness, surprise (Ekman 1982), and embarrassment (Castelfranchi and Poggi 1990; Keltner 1995). The expressive patterns of these emotions may well be universal; yet each different culture has a set of norms, called "display rules" (Ekman 1982), which prescribe if and when, out of contextual convenience, an emotion is supposed to be utterly expressed, or else masked, lowered, or intensified.

#### X.3.2 Syntactic Function

Frowning and eyebrow raising co-occur with accents, emphasis, pauses, questions, and so on. Nods may punctuate an emphatic discourse (Ekman 1979; Cassell, Torres, and Prevost 1999).

#### X.3.3 Dialogic Function

Facial expressions and gaze are part of the signals involved during the exchange of speaking turn when sender and addressee change roles in the conversation (Duncan 1974; Cassell, Torres, and Prevost 1999). They help in regulating who takes the floor, keeps it or asks for it. They provide cues to the addressee on when to ask for the turn thus avoiding that the addressee interrupts the sender without waiting for his speaking turn. Turn-taking system refers to how people negotiate speaking turns (Duncan 1974). When taking the floor the sender turns his head away from the addressee as to concentrate on what she is going to say. At some particular moments of her speech (co-occurring often with the completion of a grammatical clause) the sender might check how the addressee might show his involvement in the conversation by gazing at the sender, nodding, smiling, emitting a vocalization of agreement such as /mhm/, or asking for clarification. The sender may signal her desire to handle the speaking turn by turning her head toward the addressee, finishing any arm gestures, and assuming a more relaxed position (Duncan 1974).

#### X.3.4 Social Attitude Function

Facial expression conveys the sender's social attitude or relationship to the addressee. A raised chin is a signal of dominance, while a bent head is one of submission.

But there is a function of facial expression that has not yet been systematically investigated in research on face communication and that we are going to deal with in our work: it has something to do with attitude expression, but it is, more precisely, the facial expression of the performative of a speech act or of any communicative (not necessarily verbal) act. Anytime we communicate something to other people we perform a communicative act, that includes something we are speaking of (its propositional content) and why we are speaking of that (our performative or communicative intention) (see Section X.6). Our question is how the communicative intention of a speaker in performing one's communicative acts is communicated through facial expression, and how this can be simulated in animated faces.

#### X.4 Multimodal Systems

As with human-human communication, it seems that human-computer communication benefits from the use of multimodality. Humans use facial expressions as well as gaze, head, arm, and hand gestures to communicate. Computers may use pen, speech recognition, speech synthesizer, facial feature tracking, 3-D agents, graphics, video, and so on. All these different cues coming from different modalities should be integrated to improve the interaction.

**X.4.1.1 Modality Synergy** Different studies have shown that the redundancy of audio and visual signals can improve speech intelligibility and speech perception (Bolinger 1989; Hadar et al. 1983; Magno Caldognetto and Poggi 1997; Schwippert and Benoit 1997). For example, an accent can be marked by any one of the following signals: the voice pitch, a raised eyebrow, a head movement or a gesture, or a combination of these signals. At the same time, looking at a face while talking improves human perception (Benoit 1990; Massaro and Cohen 1990; Summerfield 1992). People, especially those who are hard of hearing, make use of gesture information to perceive speech. Similarly, speech recognition performance when combining the audio and visual channels is higher than when using only one channel (Risberg and Lubker 1978).

**X.4.1.2 Different Modalities, Different Benefits** Signals from visual and audio channels complement each other. The complementary relation between audio and visual cues helps in ambiguous situations. Indeed, some phonemes can be difficult to distinguish on the basis of sound alone (e.g., /m/ and /n/) but easily differentiated visually (/m/ is done by lip closure while /n/ is not) (Jeffers and Barley 1971).

**X.4.1.3 Adaptability** Speech is the product of several activities: the configuration of the vocal cords, larynx, and lungs as well as the movement of the lips and tongue. That is, the visual and audio channels are associated in speech: the ear hears the sound while the eye sees the lip and tongue movements. These channels are the most common ones. Hearing impaired, blind people may use other speech channels (visual and audio) such as the tactile one to get information (Benoit et al. to appear) to compensate the loss of information from the other channels. Blind people use their sense of touch to understand spoken or written language, relying on, for example, the Braille or the Tadoma methods (users feel with their hands the speaker's articulators).

**X.4.1.4 Naturalness** Speech is a very natural means of communication among humans; we have been using it since we were young. We learn to speak but also to express our emotions, beliefs, attitudes, and so on with our body, eyes, face, and voice. Every day we converse with others using verbal and nonverbal signals to exchange our ideas, give orders, directions and so on.

#### X.4.2 Multimodal Artificial Systems

Artificial systems may take advantage of multimodality as human communication does (Benoit et al. to appear; Blattner and Dannenberg 1990; Suhm 1998).

**X.4.2.1 Modality Synergy** Interfaces can benefit from modality synergy on both the input and output sides of the system since it will integrate the different aspects of each modality. Information on input and output sides may be conveyed redundantly and/or complementary to increase interpretation and display accuracy respectively. For example, lately, automatic speech recognition systems combine information from acoustic and visual channel to augment their recognition rate (Adjoudani and Benoit 1996; Meier, Stiefelhagen, and Yang 1997). Recognition results for normal audio condition (no noise added) may reach

99.5% success when information on both channels are integrated while the results for the visual channel only is 55% and for the acoustic channel only is 98.4% (Meier, Stiefelhagen, and Yang 1997).

**X.4.2.2 Different Modalities, Different Benefits** Combining modalities enables the interface to take advantage of the combination of benefits from each modality. For example in some applications, having the choice of using different modalities as input such as speech, pointing gestures and menu may help the interaction: a command is easier/faster to speak than to choose from a menu while selecting an object directly by pointing at it on the screen is easier than to describe it verbally.

**X.4.2.3 Adaptability** The user may have the choice of performing the same task through several modalities (keyboard, speech, pointing gestures) and then select the modality that is best adapted to her at the moment of the action. For example in car navigation, speech input is more adequate to ask for a particular direction since the hands must be on the wheel and may not be used to select a direction on a menu. Interfaces may offer different modalities to users with disabilities, adapting modalities to suit each person's needs.

**X.4.2.4. Naturalness** Much effort has been made to make computer interaction more natural and less constraining to the user, thereby making her more at ease. This is specially valid in the case where human-human communication aspects are transmitted to human-computer interfaces. Several interfaces have been built where a user can dialogue with a 3-D synthetic agent (Cassell et al. 1994; Cassell et al. 1999; Nagao and Takeuchi 1994; Thórisson 1997) aiming at creating a natural conversation setting.

In the case of speech multimodal systems — that is, multimodal systems with speech as one of their components — several studies have been undertaken. In this chapter, we concentrate on interfaces using speech associated with a conversational agent.

## X.5 Animated Faces

Multimodal speech systems and animated agents have been created for the personalization of user interfaces (Cassell et al. 1994; Chopra-Khullar and Badler 1999; Pelachaud and Prevost 1994; Rist, André, and Müller 1997) and for pedagogical tasks (Badler et al., chap. XX; Lester et al., chap. XX; Rickel and Johnson, chap. XX), exhibiting nonverbal behaviors such as pointing gestures, gaze, and communicative facial expressions. The links between facial expression and intonation (Pelachaud and Prevost 1994) and between facial expression and dialogue situation (Cassell et al. 1994; Cassell et al. 1999; Thórisson 1997) have been studied, and a method to compute automatically some of the facial expressions and head movements performing syntactic and dialogic functions has been proposed. This has made it possible to create faces exhibiting natural facial expressions to communicate emphasis, topic, and comment through eyes and eyebrow movements, while also performing these functions through voice modulation (Cahn 1989; Pelachaud and Prevost 1994; Prevost 1996).

In particular, to simulate face-to-face conversation with a user in real time, Nagao and Takeuchi (1994) categorize facial expressions based on their communicative meaning, following Chovil's (1991) work. The system is able to understand what the user is saying (within the limits of a small vocabulary) and to answer the user. The synthetic agent speaks with the appropriate facial expression: for example, the head is nodding in concert with *Yes* and a facial shrug is used as an *I don't know* signal. *Ymir* (Thórisson 1997) is an architecture to simulate face-to-face conversation between the agent *Gandalf* and a user. The system takes as sensory input hand gesture, eye direction, intonation, and body position of the user. *Gandalf*'s behavior is computed automatically in real time. He can exhibit context-sensitive facial expressions, eye movement, and pointing gestures as well as generate turn-taking signals. Nevertheless, *Gandalf* has limited capacity to analyze the discourse at a semantic level and therefore to generate semantically driven nonverbal signals. *Rea*, the real estate agent (Cassell et al. 1999) is able of multimodal conversation: she

can understand and answer in real time. She moves her arms to indicate at a particular element in the image, to take the turn. She uses gaze, head movements, and facial expressions for functions such as turn taking, emphasis, and greetings as well as for back channel to give feedback to the user speaking to her. Even though Rea is able to do very sophisticated behaviors, she does not exhibit nonverbal behaviors for performative.

PPP Persona (André, Rist, and Müller 1998), a 2-D animated agent, has been created for the personalization of user interfaces. This agent is able to present and explain multimedia documents and to select which material to present to the user. He or she exhibits nonverbal behaviors such as deictic gestures and communicative facial expressions. In this work, the emphasis is on the discourse generation that also includes information on the relation between text and the images that illustrate it. Later on the authors enhanced their system to include animated characters that expose multimedia information to the user (André et al., chap. XX).

Noma and Badler (1997) developed tools to create a virtual human presenter based on *Jack*, an animated agent system. The tools allow the user to specify gesture, head movement, and other nonverbal behaviors within the text that the presenter should make. A set of markup elements was developed to describe the different gestures accompanying the presenter's speech. The animation is then performed synchronized to speech in real time. The specification of the markup within the text is done manually, not automatically.

The *Olga* project (Beskow 1997) integrates conversational spoken dialogue (i.e., speech recognition and natural language understanding), 3-D animated facial expressions, gestures, lip-synchronized audiovisual speech synthesis, and a direct manipulation interface. In the human-human communicative process, verbal and nonverbal signals are active. One chooses the appropriate signal to display from an internal state, goal to achieve, and mental state but also from the context where the conversation is taking place, the interlocutors, and the relationship with the interlocutor.

Takeuchi and Naito (1995) introduced the notion of "situated facial displays." Situatedness means that the system not only follows its internal logic but is also affected by external events. External events include reactions of users, arrival of a new user interacting with the system, actions done by one of the users, and so forth. These events are perceived using a vision module that detects when users enter its field of view and track users' gaze and head behavior. As in Nagao and Takeuchi's system (1994), facial displays, called "actions" here, are computed based on their communicative meaning, and their choice depends on the internal logic of the system. On the other hand, "reactions" correspond to behaviors invoked by the system as a reaction to a new external event: the 3-D agent will turn his head fast to look at the user moving in front of him. The facial animation module outputs the 3-D face model.

In their system, Takeuchi and Naito consider only external events (users' movement and position). As we show below, our method differs from their system. Indeed, in our definition of context, we propose to take into account the social relationship of the interlocutors (here, machine/human) and their personality, along with the goal the speaker has in mind to communicate. We think that the choice of words, body postures, facial displays, and gaze behaviors is highly dependent on who our interlocutor is and what our relationship is to him or her.

Another difference from other systems is that in most of the mentioned agents, face simulation is triggered by written or intonational input: for instance, if a part of a written sentence is marked "comment," or a spoken phrase carries an emphatic stress, the system triggers a synchronized eyebrow raising; the visual signal is therefore directly connected with an audio or graphic output. The challenge of the system we designed (see also Poggi and Pelachaud 1998) is to generate a complex message that coordinates auditory and visual signals on the basis of underlying cognitive/semantic information. Our aim is to construct an expressive face starting with a meaning-to-face approach rather than a voice-to-face approach — that is, a face simulation directly driven by semantic data.

## X.6 The Notion of the Performative in Speech Act Theory

The notion of the performative dates back to the very beginning of speech act theory. Ever since his first formulation, Austin (1962) stressed that every sentence has a performative aspect, since it performs some action. In fact, any sentence performs a locutionary, an illocutionary, and a perlocutionary act: it is an act of doing something physically, but *in* being uttered (*in* locution), it also performs a social action, and through this (*per* locution), it may also have some effects on the other. The illocutionary force of the sentence (the type of action it performs) is its performative, and it can be made explicit verbally by performative verbs (such as *I assure*, *I promise*, or *I command* . . .) or performative formulas (such as *thanks* or *please*). Searle (1969) clearly states the notion of the speech act as an act including a propositional attitude and a propositional content, one formed in its turn by the act of referring to and the act of predicating.

Among the theoretical models in the line of speech act theory, the view of performatives we present here is based on a model of social action and communication in terms of goals and beliefs (Castelfranchi and Parisi 1980; Conte and Castelfranchi 1995; Parisi and Castelfranchi 1976). By matching Austin's intuition that a sentence is an action with Miller, Galanter, and Pribram's (1960) cybernetic model of action, which holds that any action is regulated by a goal, Parisi and Castelfranchi (1976) claim that every sentence has a goal — that is, in every sentence, the speaker aims at having the hearer do some action, or answer some question, or assume some information. The goal of the sentence is its meaning, and it is made up of a performative and a propositional content.

Moreover, sentences as well as actions may be hierarchically ordered in plans; any sentence has a goal but it may also have one or more supergoals, goals superordinate to the literal goal of the sentence and inferable from it; and the supergoal may be either idiomatized or not. In the former case, the inference to understand is automatic, as in Searle's (1975) indirect speech acts: *Can you pass the salt?* is no more a question but a request. In the latter, the supergoal is to be inferred anew from context and shared knowledge, somehow as a Gricean implicature (Grice 1975): if I tell you *Pass me the salt*, it may mean "Please help me to cook fast" or "Don't put too much salt in your dish; watch your blood pressure." Finally, the goals and supergoals of sentences in sequence may be hierarchically ordered and make discourses: a discourse is a plan of speech acts that all, directly or indirectly, aim at a final communicative goal (Parisi and Castelfranchi 1976).

## X.7 Communicative Act Theory

In the model we present here (Poggi 1991, n.d; Poggi and Magno Caldognetto 1997), communication holds any time that an agent *S* (a sender) has the goal of having another agent *A* (an addressee) get some belief *b* (a belief about *S*'s beliefs and goals); in order to reach this goal, *S* produces a signal *s* that *S* supposes is or may be linked in both *S*'s and *A*'s minds to some meaning *m* (the belief about one's goals and beliefs which *S* has the goal of communicating).<sup>1</sup>

To produce a signal in order to the goal of communicating some meaning is to perform a communicative act. A communicative act is then the minimal unit of communication, and it can be performed via linguistic devices (thus being a speech act proper), or via gestural, bodily, facial devices. Communicative acts are then a superset of speech acts: speech acts are those communicative acts that are performed through verbal language, while communicative acts in general may be performed through any kind of signal: a dress, a perfume, a strike, a slap, a kiss, a drawing.

A communicative act is an action performed (but also, sometimes, a morphological feature exhibited) by a sender through any (biological or technological) device apt to produce some stimulus perceivable by an addressee, with the goal that the addressee acquire some belief about the sender's beliefs and goals. Therefore, by saying that communication holds only with a goal of letting the other know something, we do not mean that a communicative goal is necessarily a conscious intention, that is, a goal one is aware of.

Given our general cybernetic notion of a goal as a regulatory state (as any state that determines behavior), among communicative goals we include also unconscious goals — say, a neurotic symptom that tries to "tell" me there is something wrong with me — or even biological goals — say, blushing — that may be viewed (Castelfranchi and Poggi 1990) as a involuntary apology for transgressing a social norm or value, aimed at preventing the group's aggression. Various levels of intentionality are then possible in communication, and all have to be included in a notion that encompasses even animal communication. Within human communicative behavior, the existence of different levels of intentionality is particularly clear in nonverbal communication. Some types of gestures or facial expressions are produced at a high level of awareness, so much that subsequently one can remember the specific gesture or expression produced, while others, especially those produced in the flow of discourse, are often not self-aware gestures or expressions, and one could not remember them precisely. This is also the case because these communicative signals are produced and perceived in a modality other than the acoustic one; that is why, as we have seen, they may occur at the same time as the verbal signal.

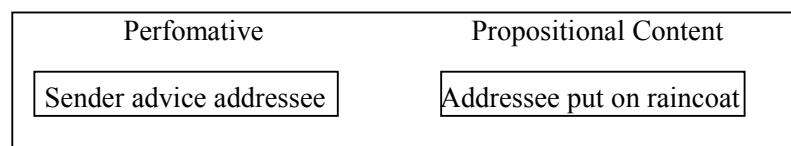
A communicative act has two faces: a *signal* (the muscular actions performed or the morphological features displayed — say, vocal articulation for speech acts, and facial expression, hand movements, or gaze for nonverbal communicative acts) and a *meaning* (the set of goals and beliefs that *Ai* has the goal to transfer to *Aj*'s mind). The signal part of a communicative act may be represented formally by facial actions, phonetic articulations, intonation contours, and so on; the meaning part may be represented in terms of logical propositions that we call "cognitive units," declarative representations of semantic primitives, by which all kinds of semantic content, including communicative intentions, word meanings, and emotions, may be expressed (see below).

#### X.8 The Meaning Side of a Communicative Act

The meaning of a communicative act includes a *performative* and a *propositional content*. In the sentence "Put on your raincoat," the propositional content includes what the sender *S* is referring to (the addressee and the raincoat) and what the sender is predicating about the referents (the action of putting on); the performative is the type of social action that *S* is performing toward *A* in mentioning the propositional content (fig. 1). Here the performative is one of advice — that is, the goal of having the addressee do some action which is good for him or her.

Figure 1

Communicative Act



*The structure of a communicative act*

That a communicative act is made up of both a performative and a propositional content is shown by the fact that we may have different communicative acts that share the same performative but have different propositional contents, as well as different communicative acts where the same propositional content is the object of different performatives. Thus, I may advise you to put on your raincoat or to study artificial intelligence; moreover, I may suggest you, order you, or implore you to put on your raincoat.

As Austin and Searle have shown, different classes of performatives can be distinguished; in our model, we distinguish three "general types of goal," very broad classes of performatives that differ for the type of action they request from the addressee. Performatives like *order*, *command*, *implore*, *propose*, *offer*, and *advise* all have the goal that the addressee do some action, and we class them as requests; *ask*, *interrogate*, and the like make up the class of questions; *inform*, *warn*, and *announce* belong to the class of informative acts.

#### X.9 The Signal Side of a Communicative Act

In performing our communicative acts, we can use verbal or nonverbal signals, or a mix of the two. Suppose I want to advise my friend to put on his raincoat and provide the justification that it's raining. I can convey both communicative acts using only verbal signals and say *Put on your raincoat. It's raining* (see fig. 2, line 1). But I could also use a mixed (verbal and nonverbal) discourse: for instance, tell him to put on his raincoat and at the same time point out the window and show him the rain, using a gesture or simply a head nod or gaze (line 2). Finally, I could rely completely on nonverbal discourse: hand him the raincoat and show him out of the window (line 3).

In fact, when we perform communicative acts through different modalities, we have to "decide" (meaning a decision usually not at a high level of consciousness and intentionality) what to communicate by verbal and nonverbal signals; moreover, since signals may use different modalities and may then be simultaneous, we may have to decide whether to use the two modalities sequentially or simultaneously, and in this case whether to convey the same message by two signals in the different modalities (hence being redundant) or to provide different information in the two modalities.

For instance, by using the two modalities simultaneously, I could convey the same communicative act by both the verbal and the nonverbal signal, at the same time saying "it's raining" and pointing out the window (line 4). Or else, I could simultaneously convey different but congruent communicative acts: say "*Put on your raincoat*" and point at the rain (line 5).

Figure 2

	CA1 S ADVISE A THAT A PUT ON RAINCOAT	CA2 S INFORM A THAT IT RAINS	
v ..... nv	Put on your raincoat	It's raining	(1) sequential
v ..... nv	Put on your raincoat	Point window	(2) sequential
v ..... nv	Hand raincoat	Point window	(3) sequential
v ..... nv		It's raining Point window	(4) simultaneous
v ..... nv	Put on your raincoat	Point window	(5) simultaneous

*Verbal and nonverbal sequential and simultaneous communicative acts*

Now, this division of labor among verbal and nonverbal signals may hold not only at the level of combinations of communicative acts, like discourse or conversation, but also at the level of a single communicative act.

Take again the advise to a friend to put on his raincoat: within this single communicative act, I may convey the performative verbally, by a performative verb or formula, or else simply through intonation or facial expression (fig. 3); and in this case, since the signals use different modalities, performative and propositional content may be conveyed at the same time.

Figure 3

	Performative S ADVICE A	Propositional Content A PUT ON RAINCOAT	
v ..... nv	I advice you	To put on your raincoat	(1) sequential
v ..... nv	Advice expression	Put on your raincoat	(2) simultaneous

*Verbal and nonverbal sequential and simultaneous performative and propositional content in the same aommunicative act*

The decision on how to distribute the verbal and the nonverbal will depends on a consideration of the available modality, but also on the cognitive ease of production and processing of signals (in describing an object, a gesture may be more expressive than a word), on the fact that some signals may be more easily devoted to communicate some kinds of meanings than others (emotions are better told by facial expression than by words), and on their different appropriateness to different social situations (an insulting word may be less easily persecuted than a scornful gaze). Moreover, there may be metacommunicative constraints that, say, lead to the use of redundancy (both the verbal and nonverbal signals) when information to convey is particularly important or when it needs to be particularly clear.

#### X.10 The Generation of Performatives in Communicative Acts

How can we simulate the generation of a performative of a communicative act?<sup>2</sup> Our idea is that the performative is not generated all at once, but through subsequent specification of the general type of goal of the communicative act, whose cognitive structure is combined with contextual information.

According to this hypothesis, then, first comes the goal, a very general goal (I want the other simply to do something, or tell me something, or believe something). But since the way I can lead the other to perform actions, tell me, or assume beliefs varies according to who the other is, what our social relationship is, and so on, I have to take all this "contextual" information into account and specify my communicative goal more narrowly, ending up with a very specific goal (a performative) tailored to the actual addressee and the social situation at hand.

In this view, the performative of a communicative act is a "context-situated interaction goal," a communicative goal where information is specified about a number of relevant interactional features. Suppose I have the general type of goal of requesting somebody to put on his raincoat because it's raining outside. I can *order* my four-year-old child to do so, but if he is sixteen I may have to *implore* him; to my boss, I may perhaps *suggest* it, while I may *warn* a friend of mine because I am worried about him. In all these cases, the sentence may be the same but my performative facial expression will be different from case to case.

In fact, in real interaction the goal of our communicative act is not only a general type of goal, but a very specific goal: among requests, we distinguish between orders, advice, implorations, proposals, suggestions; among information, we distinguish between warning, swearing, criticism, and so on. These more specific goals of communicative acts are their performatives, and they are different from each other in that each of them is semantically richer; each contains more information than the plain general type of goal. As Austin pointed out, a performative is always present in a speech act, and, we add, in a communicative act; but different from what he held, it is also always explicit, not only as it is stated through performative verbs or formulas: often, in both speech acts and communicative nonverbal acts, it is not expressed through words but through intonation or facial expression.

As we request that somebody do something, we may do it in a very bossy, polite, or empathic way, depending on the context in which the conversation takes place; and our bossiness, politeness, or empathy is made explicit, either jointly or alternatively, by words, voice, or face.

In this chapter, we investigate two points related to the generation and expression of performatives. First, we have to show how it happens that from a general goal a particular performative is specified: how it is, for instance, that from the general goal of requesting an addressee to do some action, the sender comes to specify it as a performative of, say, commanding, imploring, advising, and so forth. Second, we will see how the specified performative is exhibited through facial expression.

#### X.11 General Types Of Goals

Take these sentences:

- (1) *John, put on your raincoat.*
- (2) *John is putting on his raincoat.*
- (3) *Is John putting on his raincoat?*

These sentences exemplify the three general types of goal — request, information, and question — where, respectively, a sender wants an addressee to do some action, or to believe some belief, or to provide some information. Moreover, two types of questions can be distinguished:

- (4) *What is John putting on?*
- (5) *Is John putting on his raincoat?*

The former is a Wh-Question, where *S* wants *A* to let *S* know some new information: the latter is a Yes/No Question, where *S* wants *A* to tell whether some hypothesized information is true or not.

Below, we represent formally these four general types of goals of communicative acts in cognitive units. In our formalism, which follows Castelfranchi et al. (1998), *S* is called *A<sub>i</sub>* and *A*, *A<sub>j</sub>*; *x* is a variable, a constant or a function denoting a domain "object": specifically, *a* denotes a domain "action," *b* a domain "fact."

##### 1. Request:

*Goal A<sub>i</sub> (Do A<sub>j</sub> a)*

*A<sub>i</sub> has the goal that A<sub>j</sub> do some action a.*

##### 2. Inform:

*Goal A<sub>i</sub> (Bel A<sub>j</sub> b)*

*A<sub>i</sub> has the goal that A<sub>j</sub> believe some belief b.*

3. Ask (two types):

Wh-Question

*Goal Ai (Goal Aj (Bel Ai b))*

*Ai* has the goal that *Aj* do something in order to have *Ai* believe some belief *b*.

Yes/No Question

*Goal Ai (BW Ai (Bel Aj b))*

*Ai* has the goal that *Aj* do something in order to have *Ai* know whether some belief *b* is true or not.<sup>3</sup>

X.12 From the General type of Goal to the Performative

As we mentioned, in different communicative acts all having the same general type of goal, the different performatives are semantically richer than the general type of goal itself in that they contain additional cognitive units; therefore, the performatives can be distinguished from each other in terms of specific features, all representable in terms of cognitive units. Some of these cognitive units are particularly relevant for distinguishing within requests, some within informations, and so on. Here are the features that distinguish performatives from each other.

X.12.1 In Whose Interest Is the Action Requested or Information Provided?

In a command as opposed to a piece of advice, a relevant difference can be seen in the answers to these questions: Whose goal does the requested action serve? In whose interest is it? If I command you *Go and get me the newspaper*, then you take the newspaper is a goal of mine. But if I offer you the advice *Take the umbrella when you go out*, I am suggesting that taking the umbrella will prevent you from getting damp, which is a goal of yours — more precisely, an interest of yours, a goal you may not be aware of (Conte and Castelfranchi 1995). So,

- In a command, *Ai* wants *Aj* to do *a*, where *a*, the requested action, is in *Ai*'s interest. This means that *Ai* wants to achieve the goal *g* and believes that the action *a* will be useful in achieving it:

*Goal Ai (Do Aj a),*

*(Goal Ai g) ∧ (Bel Ai (Achieve a g)).*

- In advice, *Ai* also wants *Aj* to do *a*, but the difference from a command is that the requested action is in *Aj*'s interest. *Ai* believes that *Aj* wants to achieve the goal *g* and that the action *a* will allow *Aj* to achieve it:

*Goal Ai (Do Aj a),*

*(Goal Aj g) ∧ (Bel Ai (Achieve a g))*

X.12.2 Degree Of Certainty

Among performatives of information, a relevant difference is the degree of certainty with which the provided information is assumed by the sender: this distinguishes, for instance, claiming from suggesting (in its reading as information). Uncertainty is represented by the mental atom "uncertain."<sup>4</sup>

4. Suggest:

*Bel Ai (Uncertain (Achieve a g))*

*Ai* is not certain that the action *a* may help to achieve the goal *g*.

X.12.3 Power Relationship

One more difference among requests, particularly clear when comparing, say, commands, advice, and implorations, is the power relationship holding between sender and addressee. In a command, *Ai* calls up to one's power on *Aj* and shows a willingness to take advantage of it; this implies that if *Aj* does not fulfill the request, then *Ai* could retaliate. In imploring, *Ai* acknowledges *Aj*'s power over *Ai*; while in advising, even if having power over *Aj*, *Ai* claims that he or she does not want to take advantage of it, thus leaving *Aj* free to do the requested action or not.

5. Command:

$(Goal\ Ai\ (Do\ Aj\ a)) \wedge (Goal\ Ai\ (Bel\ Aj\ (Power-on\ Ai\ Aj)))$

6. Implore:

$(Goal\ Ai\ (Do\ Aj\ a)) \wedge (Goal\ Ai\ (Bel\ Aj\ (Power-on\ Aj\ Ai)))$

#### X.12.4 Type of Social Encounter

Our talk is different in a familiar context (say, to a friend) or in a service encounter (say, to a clerk); hence, performatives differ in how formal or informal the relationship between sender and addressee is and, more generally, according to the kind of social relationship they have with each other, whether motivated by instrumental or affective goals. *Forgive* and *excuse*, for instance, seem to be more linked to forgetting a person's faults against, respectively, ethics versus etiquette rules; in the same vein, *inform* seems more formal than *tell*.

#### X.12.5. Affective State

Many performatives contain information about some actual or potential affective state of *Ai*'s. In a peremptory order, *Ai* shows that he or she might be angry if *Aj* did not fulfill *Ai*'s request; in a warning, *Ai* reveals worrying about *Aj*'s good.

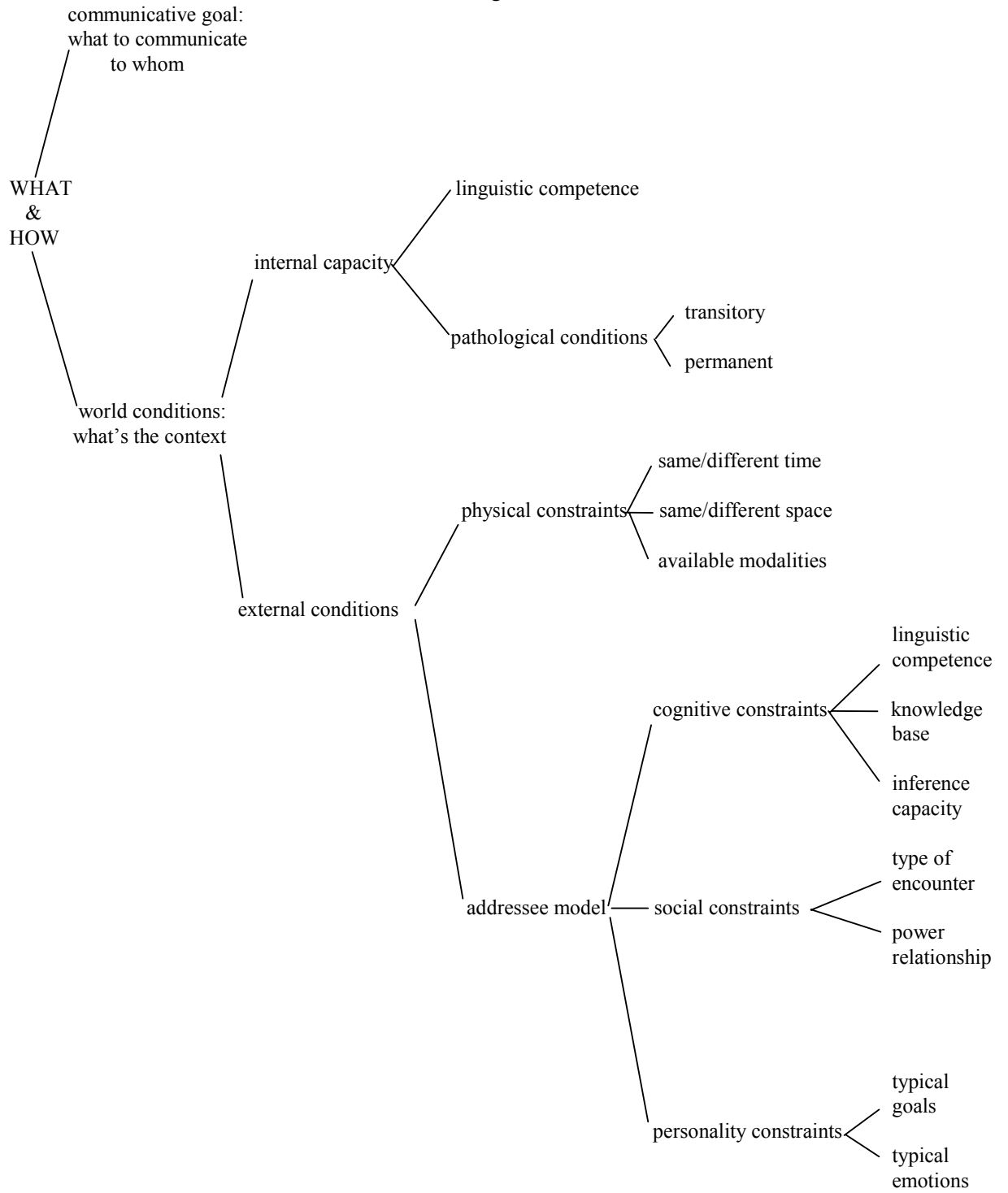
In the system we present here, information that specifies the general type of goal as a performative, and then outputs it through expressive devices, comes in at two different stages. Information about interest, certainty, power relationship, and social encounter (degree of formality) determines the mental state of the specific performative, while affective state only comes in at the expressive stage, resulting in enhancement or deintensification of the expression, or in the addition of affect displays to it.

Our hypothesis is, in fact, that information that specifies the general type of goal of a communicative act is not part of our communicative goal from the beginning, but it comes from consideration of context. A sender "decides" which specific performative to use in his or her sentence (or other nonverbal communicative act) on the basis of the social situation, the social relationship to the addressee, and the addressee's cognitive, affective, and personality factors.

#### X.13 A Model of Context

As any action of our life, communicative behavior is determined by both our goals and the world conditions at hand — that is, by the context (see fig.4).

Figure 4



*A model of context*

So, at the start, a sender has a global idea of a communicative act in mind, but he or she has to specify how to convey it by taking into account the communicative possibilities at hand (Poggi and Pelachaud 1998). On the one hand, the sender has to consider her internal capacities, which include her linguistic competence (say, she may be a foreigner who has not mastered the language completely) and possible transitory or permanent pathological conditions (such as slips of the tongue or aphasia). On the other hand, external conditions exist: physical constraints and the sender's model of the addressee.

As for physical constraints, the sender will take into account whether communication is face to face or at a distance, whether the addressee is simultaneously present in the same spatiotemporal situation, and what the available modalities are: only acoustic (say, on the phone), only visual (through a window or across a road), or both. This may determine, for instance, whether to communicate a performative through intonation and words or facial expression. Finally, the model the sender has of the addressee includes his or her cognitive, social, and personality constraints. Cognitive constraints are the addressee's linguistic competence, knowledge base, and inference capacity: they account for why, for instance, we speak slowly to a tourist, because we think he does not understand our language well; why we explain things more at length to students lacking background knowledge; or why we explicate obvious causal links to children or dull people, assuming them to be inferentially slower. Among the social constraints we consider are the type of social encounter we are engaged in, our power and status relationship to the addressee, and personality factors (see Ball and Breese, chap. XX; Nass, Isbister, and Lee, chap. XX).

Now, the physical constraints and the constraints of the addressee's linguistic competence and inference capacity are particularly relevant in deciding whether to communicate in a visual or acoustic modality and in choosing the level of explicitness in sentences; but since we are now dealing with the facial expression of performatives, here we focus only on the contextual constraints that are relevant in choosing facial actions.

These include:

#### X.13.1 Type Of Social Encounter (Formal vs. Informal)

Suppose I ask someone to pass me the salt: if I am at the same table with a friend I may simply say *Salt*, or even only point at the salt with gaze or a chin tilt. But if I ask unknown people at another table, I may say *Could you please give me the salt?* with a shy, smiling expression. A more formal situation, in fact, determines the triggering of politeness rules that may for instance generate a smile as an equivalent of polite forms in sentences (the use of *please* or indirect requests).

X.13.2 Power Relationship between Sender and Addressee This is what determines the difference among, say, command, advice, and imploration. I may decide to display a straight and serious face if I think that the addressee is obliged to do what I want — that is, if I think I have some power over him or her.

#### X.13.3 Personality Factors of Both Sender and Addressee

Personality factors, in our model, can be seen especially in terms of the two following elements:

X.13.3.1 Typical Goals In our model (Conte and Castelfranchi 1995; Poggi 1998), personality may be viewed, at least in part, as the different importance that different people attribute to the same goal. For instance, a bossy person is one who always has the goal to impose his or her will; a very autonomous person attributes a high value to the goal of choosing one's goals freely and of not being helped by others, so he or she will not give advice; to a generous one, the goal of helping others is particularly worthy; for a proud one, the goal of not submitting to other people is very important.

X.13.3.2 Typical Emotions We often describe people's personalities linguistically, with adjectives such as *shy* or *touchy*. This means that we also classify people in terms of their higher or lower tendency to feel some emotions. So a shy person is one quite likely to feel shame; a touchy person is one who has a low threshold for feeling offended, that is, wounded in one's own image or self-image.

Now, these personality factors of both *Ai* and *Aj* determine the choice of a performative. If *Ai* is proud, he or she will never implore; if *Aj* is very touchy, *Ai* will be very soft in disagreeing.

#### X.14 Some Performatives of Request

Let us now see how some performatives can be represented in terms of cognitive units. We start from the performative of peremptory order.

• Peremptory order:

- (1) *Goal Ai (Do Aj a)*
- (2) *(Goal Ai g) ∧ (Bel Ai (Achieve a g))*
- (3) *Goal Ai (Bel Aj (Power-on Ai Aj a))*
- (4) *If (Not (Do Aj a)) then (Feel Ai (Angry Ai))*

The first cognitive unit (1) characterizes the peremptory order as a request; the second (2) mentions the feature "in whose interest" is the requested action. Action *a* is useful to achieve goal *g*, which is a goal of *Ai*. In fact, when I order you something, it is for a goal of mine, not for yours. The third cognitive unit (3) remarks the power relationship between *Ai* and *Aj*: I order when I want you to think I have power over you. The fourth cognitive unit (4) mentions the potential affective state of the peremptory order: if you are not doing what I request, I will be angry at you.

• Advice:

- (1) *Goal Ai (Do Aj a)*
- (2) *(Goal Aj g) ∧ (Bel Ai (Achieve a g))*
- (3) *Goal Ai (Not (Bel Aj (Power-on Ai Aj a)))*
- (4) *Goal Ai (Bel Aj (Can Aj (Not (Do Aj a))))*

In advice, the action requested is in the interest of *Aj* (2), and in making one's request *Ai* remarks that he or does not assume to have, or at least does not evoke, power over *Aj* (3), so that *Aj* is free to do the requested action or not (4).

• Imploration:

- (1) *Goal Ai (Do Aj a)*
- (2) *(Goal Ai g) ∧ (Bel Ai (Achieve a g))*
- (3) *Goal Ai (Bel Aj (Power-on Aj Ai a))*
- (4) *If (Not (Do Aj a)) then (Feel Ai (Sad Ai))*

When *Ai* implores, the action *Aj* is requested is in *Ai*'s interest (2), *Ai* claims (3) that *Aj* has power over *Ai* (hence, *Aj* would be free of not doing a), but if *Aj* should not do a, *Ai* would be sad (4).

• Proposal:

- (1) *Goal Ai (Do Aj a)*
- (2) *(Bel Ai (Goal Ai g)) ∧ (Goal Aj g)*
- (3) *Bel Ai (Uncertain (Achieve a g))*
- (4) *Goal Ai (Not (Bel Aj (Power-on Ai Aj a)))*
- (5) *Goal Ai (Bel Aj (Can Aj (Not (Do Aj a))))*

In a proposal, like in advice, *Ai* claims to be in a peer relationship with *Aj* (4) and therefore leaves *Aj* free not to do the requested action (5). The two most distinctive units are that the goal the action serves is supposed to be a goal of both *Ai* and *Aj* (2) and that *Ai* shows uncertain whether the proposed action is useful to that shared goal or not (3).

• Suggestion:

- (1) *Goal Ai (Do Aj a)*
- (2) *Bel Ai (Goal Aj g)*
- (3) *Bel Ai (Uncertain (Achieve a g))*

Suggestion is like an advice in that the requested action is in the interest of *Aj* (2), and it is like a proposal in that *Ai* is not completely sure that the suggested action is useful to *Aj*'s goal (3).

### X.15 The Expression of Performatives

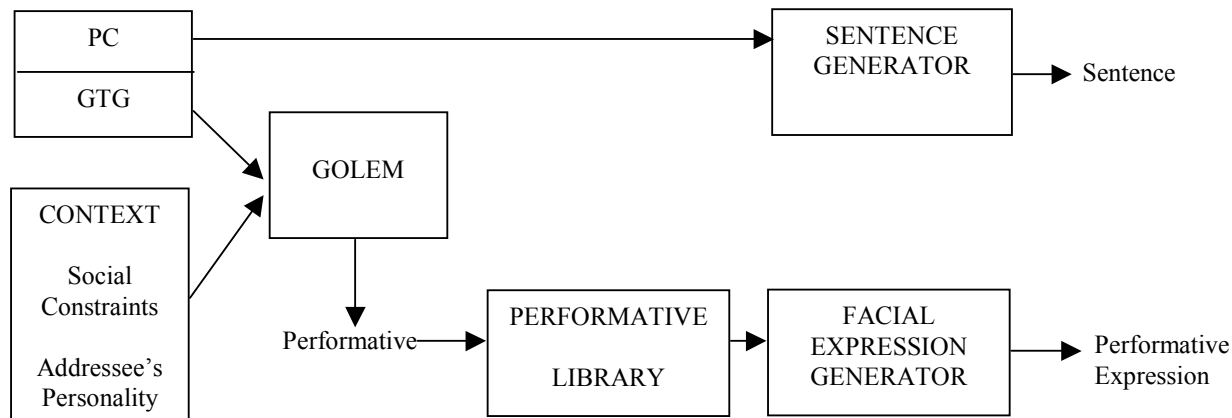
Let us now come to the signal part of a performative: how we can compute facial expression. Our hypothesis is that to each cognitive unit or cluster of them defining a performative is associated one or more nonverbal behaviors. For example, the common signal of each performative whose general type of goal is request is to "keep head right." Power relationships will be marked by a "bend head aside" when the addressee has power over the sender (ethologically, the sender displays a weak position showing her neck as a demonstration that she is not in attacking position); but in the opposite case, when the sender has power over the addressee, the sender will look down at the addressee (typical behavior of domination). Moreover, many performatives contain information about a particular affective state and may therefore be associated with a facial expression. For example, since in a peremptory order the sender shows she could be angry at the addressee in case she should not perform the requested action, this potential anger will be expressed by frowning, a typical eyebrow expression of anger. On the other hand, since in imploring the sender is potentially sad in the event that the addressee does not do *a*, potential sadness is expressed by raising the inner parts of eyebrows.

In our computer graphics system, facial expression is computed in terms of Ekman and Friesen's (1978) FACS (Facial Action Coding System), a notational system to describe visible facial actions. FACS is derived from an analysis of the anatomical basis of facial movement. Muscular actions (contraction or relaxation) are the underlying basis for changes in facial expression, and single muscular actions (or groups of related muscular actions) of the face are denoted as action units. An action unit (AU) corresponds to the minimal visible facial action. The facial model we are using is a hierarchically structured, regionally defined object (Platt 1985). The face is decomposed into regions (forehead, brow, cheek, nose, lip, for example) and subregions (upper-lip, lower-lip, left lip corner, upper lip corner). The model uses FACS to encode any basic action, so that each region corresponds to the application of particular AUs. For each facial expression, we compute the corresponding set of AUs. The final expression is obtained by adding each signal and therefore by adding each set of AUs.

### X.16 System Overview

In this section, we present an overview of our system (see fig. 5) and a detailed example.

Figure 5: System overview



The system does not have a vision module that is able to detect the user's presence and movements. It assumes the user is in front of it ready to communicate with it. The input to the system includes a propositional content and a general type of goal (GTG) (request, inform, or ask). The system takes this information as input and combines it with context information (CI). Within the model of context outlined above, the relevant information for the generation and expression of the performative includes the social constraints (type of encounter and power relationship) and the addressee's personality (typical goals and typical emotions).

In order to generate the complete cognitive structure of a specific performative — that is, to specify the single cognitive units that form the final representation of a specific performative — the system uses the GOLEM resolution engine (Castelfranchi et al. 1998) to infer specific cognitive units to complete the final performative representation. For example, suppose the social encounter  $A_i$  is engaged in is a formal encounter with his boss, one who attributes great importance to status relationships and who is very selfish and quite touchy. Here are the inferences that can be drawn:

- *if  $A_j$  is my boss, then  $A_j$  has power over  $A_i$  (power relationship);*
- *if  $A_j$  attributes great importance to status relationship, then the goal of looking powerful is an important goal of him (typical goals);*
- *if  $A_j$  is selfish, then an important goal of his is that actions he does are in his interest (typical goals);*
- *if he is touchy, then he is particularly keen to feel offended if people don't acknowledge his status (typical emotions).*

From these inferences, the system decides that the specific performative to express, the kind of attitude and social relationship  $A_i$  should express to  $A_j$  with its cognitive units must (a) imply that the action  $A_i$  is requesting from  $A_j$  is useful to  $A_j$ 's goals; and (b) show uncertainty about whether the requested action is really the best for  $A_j$ 's goal, and present it as just one of different possible actions.

In other words, the system concludes that the most convenient performative to express contains the following cognitive units:

- (a) *Bel  $A_i$  (Goal  $A_j$  g)*
- (b) *Bel  $A_i$  (Uncertain (Achieve a g))*

The system then goes to a performative library, looks for the performative that matches these requirements (whose representation contains the required cognitive units), and outputs the resulting performative.

In this example, the system concludes that the most adequate performative is *suggest*

- (1) *Goal  $A_i$  (Do  $A_j$  a)*
- (2) *Bel  $A_i$  (Goal  $A_j$  g)*
- (3) *Bel  $A_i$  (Uncertain (Achieve a g))*

Of course, different contextual information would trigger other inferences, and the system would choose another performative. Taking the same example as before, but with only one cognitive unit changed, we get a different result: for instance, if power relationship is

*Power-on  $A_i$   $A_j$*

( $A_i$  has power over  $A_j$ ), then  $A_i$  could show potential anger in case of nonfulfillment, and the resulting performative would be one of peremptory order.

Finally, the outputted performative is dispatched to the facial expression generator, while the propositional content is sent to the sentence generator: they finally give as outputs, respectively, a facial expression in terms of AUs and a sentence taken from a library of prestored sentences. The facial expression for a given performative is obtained by combining all nonverbal signals specified for each cognitive unit. For the moment, the combination is simply obtained by adding nonverbal signal.

Currently, personality traits are not yet implemented in our system, but in the future we foresee implementing both generators (facial expression generator and sentence generator) so they may again receive information on the context as an input.

Thus, type of encounter and power relationship as well as the addressee's personality traits may also motivate enhancement or intensification of facial expression. For instance, in imploring, I may show sadness because if I ask you something very important to me and I am in your power, I can anticipate that I will be sad if you do not fulfill my request. Moreover, in this case, if  $A_i$  thinks that  $A_j$  is particularly eager to be moved, an imploring face may be loaded with a more intense expression of sadness, provided by inner parts of eyebrows particularly raised. The intensity of facial expression is computed by changing either the intensity of each AU of the facial expression or the basic facial expression itself. For example, in *suggest* the eyebrows may be more or less raised, or they may be accompanied by a wide eye opening. These modulations may function as Ekman's display rules (Ekman 1982).<sup>5</sup>

#### X.17 Conclusion

We have proposed a way to construct an artificial agent that can express one's communicative intentions through facial expression. The agent computes the appropriate performative of one's communicative acts through consideration of the context of communication, particularly of the addressee, and then makes it explicit through facial expression.

#### X.18 Notes

1. Our notion of communication and of communicative act differs from information theory (Shannon and Weaver 1949) and semiotic models (Eco 1975; Jakobson, 1963; Pierce 1931–35) in that it is basically defined in terms of the goal of communicating. This is why we speak of a sender and an addressee instead of an emitter and a receiver: only if the sender has the goal of having the addressee know something is it the case that not simply a transfer of information, but a communicative process holds: we would not call communication, for instance, a case in which someone (say, a spy) comes to get some information that someone else had not the goal to let him or her know. Therefore, we speak of a sender and an addressee because these are intrinsically goal-based notions: a sender is defined as one who has the goal of transferring beliefs to someone else; an addressee is the one to whom some sender has the goal to transfer some beliefs (Poggi n.d.).

2. In this chapter, we do not discuss other important views of performatives and communicative actions in AI, such as Cohen and Levesque (1990; 1995) and Posner (1993).

3. *Belief-Whether (BW)*:  $(Bel A b) \vee (Bel A \text{ not } b)$ .

4. For computational simplicity, we consider at the moment only three degrees of certainty: *Bel*, *BelNot*, *Uncertain* (Castelfranchi et al. 1998).

5. One thing we did not deal with in this work is the computation of intonation. In principle, we think that it should be possible to build a system that, starting from a given cognitive material, is able to generate, not only a verbal output through a natural language sentence generator, but also the appropriate intonation and an adequate facial expression. In this work, we focused only on the visual output. On the other hand, as we mentioned, intonation is an acoustic device that, to a great extent, is equivalent to facial expression as far as the function of communicating the performative of sentences goes; in addition, it is an even more sophisticated device than the face itself in expressing performatives: in everyday communication, we can distinguish very subtly between a suggestion and an advice, a prayer and an imploration, a warning and an announcement, and this we do it just thanks to the subtle intonational differences in our voice (Ladd, Scherer, and Silverman 1986). Unfortunately, the nature of such a sophisticated device has not yet been defeated by researchers, neither by auditory nor by acoustic instrumental analysis (Scherer 1988); intonational differences have been identified among, for instance, interrogative and informative sentences

in general (Bolinger 1989; Pierrehumbert and Hirschberg 1987), but research is at its first steps in distinguishing the intonational differences within each general class of sentence (Cahn 1989; Prevost 1996): within the class of informative sentences, how is it phonetically characterized an announcement from a warning? Once research has filled in this gap, it will also be possible to construct synthetic talking faces that can mark these differences from an intonational point of view.

## References

- Adjoudani, A., and C. Benoit. 1996. On the integration of auditory and visual parameters in an HMM-based ASR. In D. G. Stork and M. E. Hennecke, eds., *Speechreading by humans and machines: Models, systems, and applications*, 461–472. Berlin: Springer-Verlag.
- André, E., T. Rist, and J. Müller. 1998. Webpersona: A lifelike presentation agent for the world-wide web. *Knowledge-based Systems* 11(1):25–36.
- Argyle, M., and M. Cook. 1976. *Gaze and mutual gaze*. Cambridge: Cambridge University Press.
- Austin, J. L. 1962. *How to Do Things with Words*. The William James Lectures at Harvard University, 1955, J.O. Urmson, ed. London: Oxford University Press.
- Benoit, C. 1990. Why synthesize talking faces? In *Proceedings of the ESCA Workshop on Speech Synthesis*, Autrans, France: 253–256.
- Benoit, C., J.C. Martin, C. Pelachaud, L. Schomaker and B. Suhm. to appear. Audio-visual and multimodal speech systems. In *Handbook of Standards and Resources for Spoken Language Systems*. The Hague: Kluwer.
- Beskow, J. 1997. A conversational agent with gestures. In *Proceedings of IJCAI '97 Workshop on Animated Interface Agents—Making Them Intelligent*, Nagoya, Japan, August.
- Blattner, M. M., and R. Dannenberg. 1990. CHI '90 workshop on multimedia and multimodal interface design. *SIGCHI Bulletin* 22(2):54–57.
- Bolinger, D. 1989. *Intonation and its uses*. Stanford: Stanford University Press.
- Cahn, J. 1989. Generating expression in synthesized speech. Ph.D. diss., Media Lab, Massachusetts Institute of Technology, Cambridge, MA.
- Cassell, J., J. Bickmore, M. Billinghamurst, L. Campbell, K. Chang, H. Vilhjalmsson, and H. Yan. 1999. Embodiment in conversational interfaces: Rea. In *CHI '99 Conference Proceedings*, 520–527, Pittsburgh, Pennsylvania.
- Cassell, J., C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. 1994. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Computer Graphics Annual Conference Series*, ACM Siggraph, 413–420.

- Cassell, J., Torres, O., and S. Prevost. 1999. Turn taking vs. discourse structure: How best to model multimodal conversation. In Y. Wilks, ed., *Machine conversation*. The Hague: Kluwer.
- Castelfranchi, C., and D. Parisi. 1980. *Linguaggio, conoscenze e scopi*. Bologna: Il Mulino.
- Castelfranchi, C., and I. Poggi. 1990. Blushing as a discourse: Was Darwin wrong? In R. Crozier, ed., *Shyness and embarrassment. Perspective from social psychology*, 230-251. Cambridge: Cambridge University Press.
- Castelfranchi, C., F. de Rosis, R. Falcone, and S. Pizzutilo. 1998. Personality traits and social attitudes in multiagent cooperation. *Applied Artificial Intelligence* 12(7-8): 649–675.
- Chopra-Khullar, S., and N. I. Badler. 1999. Where to look? Automating attending behaviors of virtual human characters. In *Proceedings of Autonomous Agents '99*, Seattle, Washington, May.
- Chovil, N. 1991. Social determinants of facial display. *Journal of Nonverbal Behavior* 15(3):141–154.
- Cohen, P. R., and H. J. Levesque. 1990. Performatives in a rationally based speech act theory. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, 79–88.
- . 1995. Communicative actions for artificial agents. In *Proceedings of the International Conference on Multi-Agent Systems*. San Francisco: AAAI Press.
- Cole, R., T. Carmell, P. Connors, M. Macon, J. Wouters, J. de Villiers, A. Tarachow, D. Massaro, M. Cohen, J. Beskow, J. Yang, U. Meier, A. Waibel, P. Stone, G. Fortier, A. Davis, and C. Soland. 1998. *Intelligent animated agents for interactive language training*. Unpublished manuscript.
- Collett, P., and A. Contarello. 1987. Gesti di assenso e di dissenso. In P. E. Ricci Bitti, ed., *Comunicazione e gestualità*. Milan: Franco Angeli.
- Condon, W. S. 1988. An analysis of behavioral organization. *Sign Language Studies* 58:55–88.
- Conte, R., and C. Castelfranchi. 1995. *Cognitive and social action*. London: University College.
- Duncan, S. 1974. On the structure of speaker-auditor interaction during speaking turns. *Language in Society* 3:161–180.
- Eco, U. 1975. *Trattato di semiotica generale*. Milan: Bompiani.
- Ekman, P. 1979. About brows: Emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog, eds., *Human ethology: Claims and limits of a new discipline: Contributions to the Colloquium*. Cambridge: Cambridge University Press.
- . 1982. *Emotion in the human face*. Cambridge: Cambridge University Press.
- Ekman, P., and W. Friesen. 1978. *Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press, Inc.

- Fridlund, A. 1994. *Human facial expression: An evolutionary view*. New York: Academic Press.
- Grice, H. P. 1975. Logic and conversation. In P. Cole and J. L. Morgan, eds., *Syntax and semantics: Speech acts*. New York: Academic Press.
- Hadar, U., T. J. Steiner, E. C. Grant, and F. Clifford Rose. 1983. Kinematics of head movements accompanying speech during conversation. *Human Movement Science* 2:35–46.
- Jakobson, R. 1963. *Essais de Linguistique generale*. Paris: Minuit.
- Jeffers, J. and M. Barley. 1971. *Speechreading (lipreading)*. Springfield, Illinois: C.C. Thomas.
- Keltner, D. 1995. Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology* 68(3):441–454.
- Kendon, A. 1993. Human gesture. In T. Ingold and K. Gibson, eds., *Tools, language and intelligence*. Cambridge: Cambridge University Press.
- Ladd, D. R., K. Scherer, and K. E. A. Silverman. 1986. An integrated approach to studying intonation and attitude. In C. Johns-Lewis, ed., *Intonation in discourse*. London/ Sidney: Crom Helm.
- Magno Caldognetto, E. and I. Poggi. 1997. Micro- and macro-bimodality. In C. Benoit and R. Campbell, eds., *Proceedings of the Workshop on Audio-Visual Speech Processing. Cognitive and Computational Approaches*, Rhodes, Greece, September 26–27.
- Massaro, D. W., and M. M. Cohen. 1990. Perception of synthesized audible and visible speech. *Psychological Science* 1(1):55–63.
- McNeill, D. 1992. *Hand and mind*. Chicago: University of Chicago Press.
- Meier U., R. Stiefelhagen, and J. Yang. 1997. Preprocessing of visual speech under real world conditions. In C. Benoit and R. Campbell, *Proceedings of the ESCA Workshop on Audio Visual Speech Processing. Cognitive and Computational Approaches*, 113–116. Rhodes, Greece.
- Miller, G. A., E. Galanter, and K. H. Pribram. 1960. *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston.
- Nagao, K., and A. Takeuchi. 1994. Speech dialogue with facial displays: Multimodal human-computer conversation. In *Proceedings of the 32th ACL '94*, 102–109.
- Noma, T., and N. Badler. 1997. A virtual human presenter. *Proceedings of the IJCAI '97 Workshop on Animated Interface Agents-Making Them Intelligent*. Morgan-Kaufmann Publishers, Nagoya, Japan.
- Parisi, D., and C. Castelfranchi. 1976. Discourse as a hierarchy of goals. Working chapters, Università di Urbino.

- Pierce, C. C. 1931–35. *Collected chapters*. Cambridge: Harvard University Press.
- Pelachaud, C., and S. Prevost. 1994. Sight and sound: Generating facial expressions and spoken intonation from context. In *Proceedings of the 2<sup>nd</sup> ESCA/AAAI/IEEE Workshop on Speech Synthesis*, New Paltz, NY:216–219.
- Pelachaud, C., N. I. Badler, and M. Steedman. 1996. Generating facial expressions for speech. *Cognitive Science* 20(1):1–46.
- Pierrehumbert, J., and J. Hirschberg. 1987. The meaning of intonational contours in the interpretation of discourse. *Technical memorandum, AT&T Bell Laboratories*.
- Platt, S. M. 1985. *A structural model of the human face*. Ph.D. diss, Dept. of Computer and Information Science, University of Pennsylvania, Philadelphia, PA.
- Poggi, I. 1991. La comunicazione. In R. Asquini and P. Lucisano, eds., *L'italiano nella scuola elementare. Aspetti linguistici*. Florence: La Nuova Italia.
- . 1998. A goal and belief model of persuasion. *6th International Pragmatics Conference*, Reims, France, July 19–24.
- . N.d. *Multimodal communication. Hands, face and body*. Forthcoming.
- Poggi, I., and E. Magno Caldognetto. 1997. *Mani che parlano. Gesti e Psicologia della comunicazione*. Padova: Unipress.
- Poggi, I., and C. Pelachaud. 1998. Performative faces. *Speech Communication* 26:5–21.
- Poggi, I., N. Pezzato, and C. Pelachaud. 1999. Gaze and its meanings in animated faces. In P. McKeivitt, ed., *Language, Vision and Music. Proceedings of the CSNLP-8*, Galway, Ireland, August 9–11.
- Posner, R. 1993. Believing, causing, intending: The basis for a hierarchy of sign concepts in the reconstruction of communication. In R. J. Jorna, B. van Heusden, and R. Posner, eds., *Signs, search, and communication: Semiotic aspects of artificial intelligence*. Berlin: De Gruyter.
- Prevost, S. 1996. Modeling contrast in the generation and synthesis of spoken language. In *Proceedings of ICSLP '96: The Fourth International Conference on Spoken Language Processing*. Philadelphia, PA.
- Risberg, A., and J. L. Lubker. 1978. Prosody and speechreading. *Technical Report Quarterly Progress and Status Report 4, Speech Transmission Laboratory*, KTH, Stockholm, Sweden, 1978.
- Rist, T., E. André, and J. Müller. 1997. Adding animated presentation agents to the interface. In *Proceedings of Intelligent User Interface*, 79–86.
- Scherer, K. R. 1980. The functions of nonverbal signs in conversation. In R. St. Clair and H. Giles, eds., *The social and physiological contexts of language*, 225–243. Hillsdale, N.J.: Erlbaum.

———. 1988. *Facets of emotion: Recent research*. Hillsdale, N.J.: Erlbaum.

Schwippert, C., and C. Benoit. 1997. Audiovisual intelligibility of an androgynous speaker. In C. Benoit and R. Campbell, *Proceedings of the ESCA Workshop on Audio Visual Speech Processing. Cognitive and Computational Approaches*. 81–84. Rhodes, Greece.

Searle, J. R. 1969. *Speech acts*. Cambridge: Cambridge University Press.

———. 1975. Indirect speech acts. In P. Cole and J. L. Morgan, eds., *Syntax and semantics: Speech acts*. New York: Academic Press.

Shannon, C. E., and W. Weaver. 1949. *The mathematical theory of communication*. Urbana: Illinois University Press.

Suhm, B. 1998. Multimodal interactive error recovery for non-conversational speech user interfaces. Ph.D. diss., Dept. of Computer Science, Karlsruhe University, Germany.

Summerfield, Q. 1992. Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London* 335:71–78.

Takeuchi, A., and T. Naito. 1995. Situated facial displays: Towards social interaction. In *Proceedings of ACM CHI '95- Conference on Human Factors in Computing Systems*, 1:450–455.

Thórisson, K. R. 1997. Layered modular action control for communicative humanoids. In *Computer Animation'97*, Geneva: IEEE Computer Society Press.