A model of attention and interest using gaze behavior

Christopher Peters, Catherine Pelachaud, Elisabetta Bevacqua, Maurizio Mancini
IUT de Montreuil
Université de Paris 8
{c.peters, c.pelachaud, e.bevacqua, m.mancini}@iut-univ.paris8.fr

Isabella Poggi
Università di Roma
poggi@univroma3.it

Abstract. One of the major problems of user's interaction with Embodied Conversational Agents (ECAs) is to have the conversation last more than few second: after being amused and intrigued by the ECAs, users may find rapidly the restrictions and limitations of the dialog systems, they may perceive the repetition of the ECAs animation, they may find the behaviors of ECAs to be inconsistent and implausible, etc. We believe that some special links, or bonds, have to be established between users and ECAs during interaction. It is our view that showing and/or perceiving interest is the necessary premise to establish a relationship. In this paper we present a model of an ECA able to establish, maintain and end the conversation based on its perception of the level of interest of its interlocutor.

1 Introduction

Embodied Conversational Agents (ECAs) are being used more and more in applications involving interactions with users. One of the major problems these applications face is to have the conversation last more than few second between the users and the ECAs. The reasons for such a short duration may be manifold: after being amused and intrigued by the ECAs, users may find rapidly the restrictions and limitations of the dialog systems, they may perceive the repetition of the ECAs animation, they may find the behaviors of ECAs to be inconsistent and implausible, etc. Research in several areas has been undertaken to overcome these shortcomings. But we believe that another aspect to consider is the creation of special links, or bonds, that could be established between users and ECAs. Building a relationship is linked to the notion of engagement in the conversation.

Our view is that cognitive and emotional involvement and commitment are key factors that underlie the notion of engagement. If this is the case, then for an ECA to be able to establish, maintain and end interactions, it must be endowed with mechanisms that allow it to perceive, adapt to and generate behaviors relating to attention and emotion. In this paper, we will discuss some important capabilities that we have been working on: we do not present a full Speaker/Listener model, but rather illustrate how the concepts may group together to form the core of such a model. We will also focus on two important aspects of human communication, that is interest and attention during conversation. These factors have not been considered in previous studies in the same research field.

In the next Section we will present an overview of the state of the art of studies on gaze behavior. In Section 3 we will give some definitions of engagement and we will describe its importance in Human communication. In Section 4 we will then present the steps involved in engagement detection and discuss some algorithms that can be used to detect engagement at the beginning of conversation and during interaction with ECAs.

2 State of the art

A number of studies have underlined the importance of gaze behavior in the communication process. Vertegaal et al. [25] found that gaze was an excellent predictor of conversational attention in multiparty situations and placed special consideration on eye contact in the design of video conferencing systems [26]. Colburn et al. [7] have studied the effects of eye gaze in virtual human characters and found avatars that use a natural gaze model elicit changes in viewers' eye gaze patterns. Garau et al. [11] found that when avatars used gaze behaviors related to turn-taking during conversation, they consistently and significantly outperforming random-gaze conditions in terms of participants' subjective responses. Several researches [3, 24] have been undertaken to study the effects of manipulated eye gaze on persuasion in a small group. Three users, in three remote rooms, entered in a common virtual environment where their visual representations could interact. Their gaze behavior was modified in order to augment or diminish eye interactions with the other participants.

Another research area related to our work is backchannel modelling. K. R. Thòrisson developed a multi-layer multimodal architecture able to generate the animation of the virtual agent Gandalf during a conversation with an user [23]. Gandalf recognizes information like head movements or short statements, using it to generate backchannel feedback. The Rea system [5] generates backchannel feedback each time the user makes a pause shorter than 500 msec. The feedback consists in paraverbals (e.g. "mmhmm") or head nods or short statements such as "I see". Models have also been developed for controlling gaze behavior of ECAs conversing with other ECAs. For example the models of Colburn et al. [7] and Fukayama et al. [10] are based on state machines. The first one uses hierarchical state machines to compute gaze for both one-on-one conversation than multiparty interactions while the second uses a two-state Markov model which outputs gaze points in the space derived from three gaze parameters (amount of gaze, mean duration of gaze and gaze points while averted).

3 From engagement to interaction

Engagement is viewed, by Sidner et al. [22], as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake". In our terms [8, 18], it could be defined as "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction". Engagement is generally linked to (possibly caused by) interest, which could be defined as an emotional state linked to the participant's goal of receiving and elaborating new and potentially useful knowledge. Engagement and interest in their turn are a cause of attention: if I am interested in the topic or the

persons involved in an interaction, I engage in the interaction and pay attention to its topics and participants.

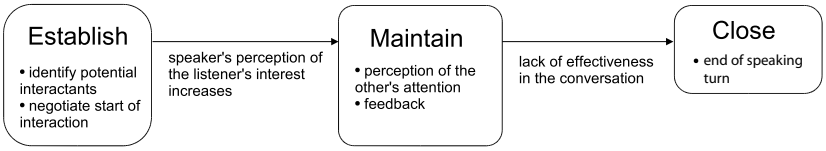

Fig. 1. Diagram of interaction phases

Actually, communication is an activity involving two (or multiple) partners. There is a Speaker (or, more generally, a Sender, since he may communicate through both verbal and non-verbal means) who wants to have a Listener (or Addressee) receive some information, and to do so he produces communicative signals. But for communication to be successful, the Addressee, in his turn, must use his resources of attention, perception and intelligence, to understand what the Sender is trying to communicate. So, it would be pointless for a Sender to engage in an act of communication only to discover that his prospective Addressee does not intend to use or does not have resources to understand what he is communicating. In other words, communication is not worthwhile without the Addressee's engagement.

In Human communication there are at least two moments in which it is important for the Sender to assess the Addressee's interest and engagement in conversation: first, at the moment of starting a communicative interaction and, second, when the interaction is going on, just to see whether the Addressee is following, understanding, concerned in, agreeing with what the Sender is saying. So there is, first, a need to assess the possibility of engagement in interaction and then, a need of checking if engagement is lasting and sustaining conversation. In case of lack of Addressee's engagement the Speaker may decide to close the conversation (see Figure 1).

3.1 Addressee's capabilities

In the construction of more and more intelligent, interactional and human-like Agents, both these moments can be reproduced and the capacities held by Human conversationalists should be implemented in ECAs.

When a Sender produces his communicative signals, for communicative interaction to go on, the Addressee must go through a number of steps:

Attention. The Addressee must pay attention to the signals produced in order to perceive, process and memorise them. Of course, attention (at least intentional attention) is made possible by engagement: if for the Addressee the goal of interacting

with the Sender, or the goal of getting information about that topic, has a very low value, he will not pay much attention to what the Sender is communicating. In the same vein, attention is a pre-condition of all subsequent steps, which are, thus all dependent on initial engagement.

Perception. The Addressee must be able to perceive the signals produced by the Sender, while not being impaired either by permanent perceptual handicaps or by transitory noise.

Comprehension. The Addressee must have the cognitive capacities for literal and non-literal comprehension: he must on the one side know the meaning attached to each signal - for example, he must master the linguistic (lexical and semantic) rules of the language used by the Sender and on the other side he must have the inferential capacities to understand the indirect meanings implied by the Sender, the structure of his discourse, his said and unsaid goals, and so on.

Internal reaction. Once the Addressee has processed the signal and extracted the meaning of what the Sender said, he might have internal reactions of a cognitive and emotional kind: for example, he may find what the Sender said to be unbelievable, or he may feel upset by it, or amused or so. This reaction, we point out, is not yet a communicative reaction, it is just the cognitive or emotional consequence induced by the Sender's says.

Decision whether to communicate the internal reaction. Whatever the internal reaction occurred in his mind, the Addressee can make one out of three possible decisions:

1. sincere communication: he may decide to communicate how he really feels, that is, to communicate in his turn to the Sender his real internal reaction. For example, if he can't believe what the Speaker is saying he can shake his head or if something in what the Sender said made him angry, he can frown to show his anger;
2. deceptive communication [9, 6]: he may decide to communicate an internal reaction different from the real one. For instance, he is really bored by the teacher's discourse, but he nods and raises eyebrows to show interest;
3. omission: he may decide not to manifest his internal reaction at all.

This decision whether to communicate internal reactions may be driven by a number of factors, among which the consequences of this communication, the social relationship with the Sender, his capability to comprehend and/or accept the Addressee's reaction (see [20]).

Generation. Once he decided to communicate (either sincerely or deceptively) his internal reaction, the Addressee should be able to display expressive synchronized visual and acoustic behaviors.

All of these processes, however, must not necessarily occur at a completely aware level: in some cases the Addressee may be aware of the fact and the ways of their occurrence, but in many cases they are quite automatic. For example, both the decision to exhibit a signal of comprehension and its generation may be quite unreflected. In any case, though, the occurrence of these processes (at least up to the process of Decision, if not the Generation process) is quite necessary for one to conclude that the Addressee is engaged in the conversation.

4 Detecting engagement before and during interaction

As we mentioned, the issue of detecting engagement in a prospective or actual Addressee is mainly relevant in two stages of an interaction:

1. at the start, when the Sender must decide if it is worthwhile to start an interaction, and does so on the basis of how possibly engaged/engageable he sees a prospective Addressee: this is what we call establish phase.
2. in the course of interaction, to monitor the level of engagement of the Addressee and the effectiveness of the interaction: this is the maintain phase.

Of these two phases, in the former (establish phase) the prospective Sender must decide by himself whether to engage in conversation, by assessing the prospective Addressee's level of interest and attention; in the latter (maintain phase), the Sender can be helped in doing so by the Listener's backchannel. During conversation, in fact, the interlocutors generally produce some signals in order to make the Speaker aware if they are really paying attention, listening to, understanding and agreeing with what is being said. That is, the interlocutors often inform the Sender about their engagement and about the smooth flowing of the processes necessary to communicative interaction: attention, perception, comprehension and internal reactions. The signal providing such information, when performed by the interlocutor without a speaking turn, are called backchannel [1], and they are performed in different modalities: by paraverbals (like mmhmm, oh), facial expression, head movements, gaze [17].

In this work we focus our efforts on both moments of the check for the Addressee's interest and attention. First we propose an algorithm for the establish phase through which the Sender can detect the Addressee's attention in order to decide whether to engage or not in a conversation. Second, we propose an algorithm for the maintain phase aimed at detecting and interpreting those backchannel signals of the Addressee that are provided through eye-gaze.

4.1 Perception of attention

Attention is a vital, if not fundamental, aspect of engagement. Indeed, it is doubtful that one could be considered to be engaged to any great extent in the absence of the deployment of attention. There are many facets of attention that are of relevance to engagement. Attention primarily acts as the control process for orienting the senses towards stimuli of relevance to the engagement, such as the Speaker or an object of discussion, in order to allow enhanced perceptual processing to take place. In social terms, the volitional deployment of attention, manifested as overt behaviors such as gaze and eye contact, may also be used for signalling one's desires, such as to become or remain engaged [21]. Therefore, the perception and interpretation of the attentive behaviors of others is also an important factor for managing ECA engagements in a manner consistent with human social behavior.

This capability focuses on social perception and attention in the visual modality geared towards the opening of an engagement. We model engagement opening as something that may start at a distance and may not initially involve an explicit commitment

to engage, such as the use of a greeting utterance. In this way, the opening of the engagement may consist of a subtle negotiation between the potential participants. This negotiation phase serves as a way to communicate the intention to engage without commitment to the engagement and has the purpose of reducing the social risk of engaging in conversation with an unwilling participant [12].

Algorithm 1 Updates **perceived gaze information and calculates interest level**
Input:
World database $database$
Sensory memory $STSS$
Short-term Memory STM

UPDATEVISUALPERCEPTION*(database, $STSS$, STM)*
$VisualSnapshot(STSS)$ //capture visual snapshot percepts into sensory memory
$STSS.ExtractAgentPercepts(personPerceptList)$ //Detect intentionality
$Resolve(personPerceptList, database)$ //Resolve person percepts list with database
for each person in $personPerceptList$ **do**
//calculate Direction of Attention
$eyeDir \leftarrow Direction(eye, me)$
$headDir \leftarrow Direction(head, me)$
$bodyDir \leftarrow Direction(body, me)$
$CalculateAttentionLevel(eyeDir, headDir, bodyDir)$ //calculate Attention level
$eyeContact \leftarrow Direction(eyeDir, myEyeDir)$ //detect Mutual Attention
$STM.AddEntry(personPercept, AL, eyeContact)$ //add information to short term memory

CALCULATEINTERESTLEVEL*(agent, $timeInterval$)*
//get interest level over time interval
$IL \leftarrow STM.Integrate(agent$,all AL's over $timeInterval)$

STARTCONVERSATION*(agent, $timeInterval$, $interestThreshold$)*
if $myGoal == interact$ **and**
$CalculateInterestLevel(agent, timeInterval) > interestThreshold)$ **and**
$STM[agent].AttentionProfile(timeInterval) == RISING$ **and**
$eyeContact == TRUE$ **then**
Signal start of conversation

4.2 Establish phase

In our model, a synthetic vision system allows our agent to visually sense the environment in a snapshot manner. Sensed information is filtered by social attention mechanism

that only allows continued processing of other agents in the environment. This mechanism acts as an agency or intentionality detector [14], so that only the behaviors of other agents are considered in later processing. Perception then consists of the segmentation of perceived agents into eye, head and body regions and the retrieval of associated direction information, as well as locomotion data, from an object database. Direction information is then weighted based on region, so that the eyes and regions oriented towards the viewer receive a higher weighting. This results in an attention level metric for an instant of time that is stored in a short-term memory system. Percepts from the memory system may then be integrated on demand to provide an attention profile spanning a time segment. Such a profile is useful for the interpretation of the attention behaviors of others: we link it, along with a gesture detection, to a theory of mind module [2] in order to establish the intention of the other to interact. Explicit commitments to interaction are only made when an agent wants to interact and theorises that there is a high probability that the other also wants to interact (see Algorithm 1 for an overview of the process).

4.3 Maintain phase

Now we want to focus our research on the attention and the interest of the Listener during conversation and how they affect the Speaker. Through the evaluation of the level of interest, the Speaker can perceive the effectiveness of the conversation and decide if it is high enough to maintain the interaction with the Listener or if he should close it. Regarding the assessment of the Listener's attention, in this paper we focus above all on gaze. Gaze is an especially important way of providing feedback and subtle signaling. Through it a Listener can show his level of interest and engagement. For example, a Listener that needs to disengage from a conversation may start to avert his gaze more frequently. Moreover, the more people share looking behaviors, the more they are involved and coordinate in the conversation. This may not necessarily involve mutual eye contact with the Speaker: during shared attention situations involving another object or entity, the Listener may actually signal their interest in the situation by directing their attention away from the Speaker and at the object in question [22].

Gaze behavior dynamism for speaker and listener

In a previous version of gaze model we used Bayesian Belief Networks [15] to determine the gaze behavior of a virtual agent. This model was based on statistical data reported in [4], corresponding to the annotation of body behaviors (gaze direction, head nods, back channels) of two subjects having a conversation. It was able to generate gaze from an input file containing both text and some tags taken from an XML-style language called APML [16]. An APML text contains what the Speaker will say and the meaning he aims at conveying, that is it does not specify which signals (i.e. facial expressions, gestures) have to be used.

A weakness of the model is the impossibility of simulating multi-party conversations without having to redefine transition tables needed by the Belief Network model. The tables increase exponentially in their complexity. Our current model is based on both APML input and two state machines defined using HPTS++ [13]. HPTS++ is a

definition language that provides tools for describing multi-agent systems using finite state machines. It also provides an environment to automatically manage the parallel execution of these machines and resolve conflicts between the allocation of the resources needed by each machine.

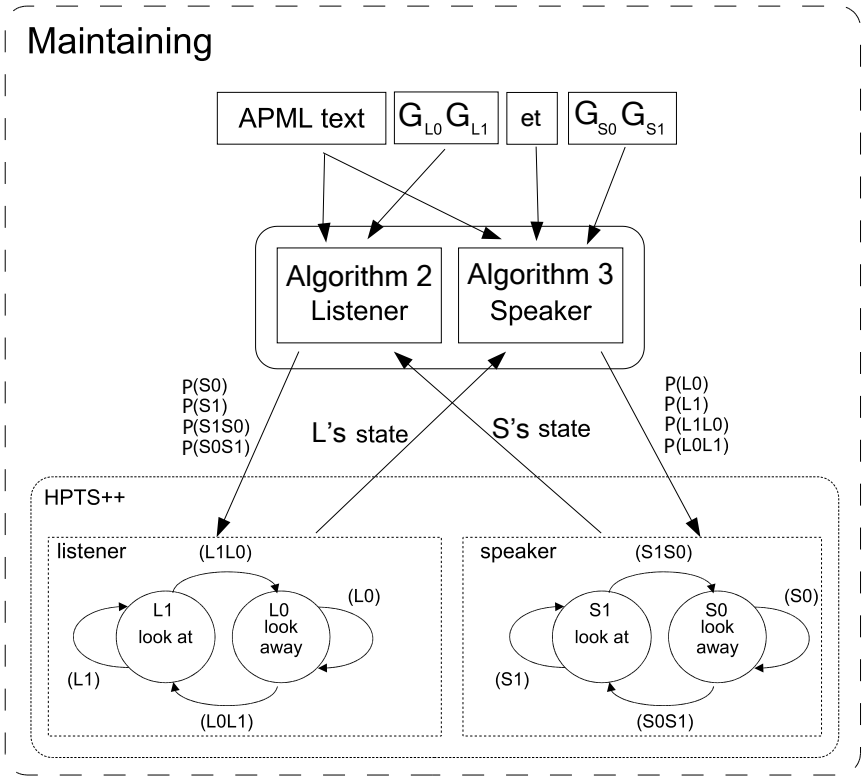

Fig. 2. Diagram of the maintaining state of the conversation

Speaker and listeners are described by state machines and their gaze behavior at a given time corresponds to the current state of the machines (see Figure 2). So, for example, to simulate multiparty conversations we need just to instantiate one state machine for each one of the participants to the conversation and let the system elaborate gaze behavior through time.

In the lower part of the diagram (the HPTS++ levels of implementation) the nodes of the state machines represent the possible gaze states of Speaker and Listener: gaze at (S1, L1) and gaze away (S0, L0). On the arcs there are the probabilities to either remain in the same state or to change state. At each time step (phoneme level) the probabilities

on the arcs may vary (see Algorithms 2 and 3). Based on these values the HPTS++ system decides if a given gaze state should hold or should transit to another one.

Algorithm 2 Computes **probabilities for HPTS++ Listener's state machine**
Input:
APML text $apml$
Max duration of Listener's look-atG_{L1} Max duration of Listener's look-away G_{L0}

while $t \leq turn\_duration$ ***do***
  $L_t \leftarrow Gaze\_of\_Listener(t)$ *//compute expected Listener's gaze direction*
  if L_t *== L0* ***then***
    $incr(P(L0))$,$incr(P(L1L0))$ *//incr the proba to remain or to transit to state L0*
    $decr(P(L1))$,$decr(P(L0L1))$ *//decr the proba to remain or to transit to state L1*
  if L_t *== L1* ***then***
    $incr(P(L1))$,$incr(P(L0L1))$ *//incr the proba to remain or to transit to state L1*
    $decr(P(L0))$,$decr(P(L1L0))$ *//decrease the proba to remain or to transit to state L0*
  //if the Listener has been in a state L0 for too long (i.e. for a duration longer than G_{L0}) then incr the proba that it will change state
  if $Time\_Listener\_in\_L0 \geq G_{L0}$ ***then***
    $incr(P(L0L1))$
  //if the Listener has been in a state L1 for too long (i.e. for a duration longer than G_{L1}) then incr the proba that it will change state
  if $Time\_Listener\_in\_L1 \geq G_{L1}$ ***then***
    $incr(P(L1L0))$
  compute gaze state of the Listener

Algorithm 2 computes Listener's gaze behavior. It takes 2 inputs, an input text with APML tags and two numbers which represent the maximum duration the Listener may consecutively look at the Speaker G_{L1} and the maximum duration the Listener may consecutively look away from the Speaker G_{L0}. Actually these parameters characterize the Listener's gaze behavior and they have been introduced in **??**. Algorithm 2, starting from APML,

determines when the Listener is expected to look at the Speaker (such as on emphasis, boundary markers, change of speaking turn, etc. see [16]). These pre-calculated behaviors are used at each time step t to determine the four probabilities for the Listener's gaze states. The algorithm considers also if a given gaze state of the Listener has not last too long (determined by G_{L1} and G_{L0} values). This allows us to avoid the state machine to remain in a deadlock state.

Algorithm 3 computes Speaker's gaze behavior. It takes as input the APML text, a threshold value and two maximum duration values for gaze direction G_{S1} and G_{S0} that work similarly as for the Listener (see description of Algorithm 2) . At first the algorithm computes the expected Speaker's gaze based on APML tags (see [19]). Then the Listener's level of attention L_a is computed as a function of Listener's gaze state L_t. Then the Listener's level of interest L_i is computed as an integration over time of the attention level. At this point it is possible to look at L_i and decide to modify or not

Algorithm 3 Computes **probabilities for HPTS++ Speaker's state machine**
Input:
Effectiveness threshold et APML text $apml$
Max duration of Speaker's look-at G_{S1}
Max duration of Speaker's look-away G_{S0}

---

while $t \leq turn\_duration$ **do**
  $L_t \leftarrow Gaze\_of\_Listener(t)$
  $S_t \leftarrow Gaze\_of\_Speaker(t)$
  $L_a \leftarrow Listener\_attention$ from L_t and S_t
  $L_i \leftarrow Listener\_interest$ from L_a
  if L_i is low **then**
    $incr(P(S1))$,$incr(P(S0S1))$ //incr the proba to remain or to pass to state S1
    $decr(P(S0))$,$decr(P(S1S0))$ //decr the proba to remain or to pass to state S0
  else
    $incr(P(S1S0))$,$incr(P(S0))$ //incr the proba to remain or to pass to state S0
    $decr(P(S0S1))$,$decr(P(S1))$ //decr the proba to remain or to pass to state S1
  //if the Speaker has been in a state S0 for too long (i.e. for a duration longer than G_{S0}) then incr the proba that it will change state
  if $Time\_Speaker\_in\_S0 \geq G_{S0}$ **then**
    $incr(P(S0S1))$
  //if the Speaker has been in a state S1 for too long (i.e. for a duration longer than G_{S1}) then incr the proba that it will change state
  if $Time\_Speaker\_in\_S1 \geq G_{S1}$ **then**
    $incr(P(S1S0))$
  //if Speaker is looking at the Listener
  if S_t == S1 **then**
    if L_i is low **then**
      $decr(effectiveness)$
    else
      $incr(effectiveness)$
  if $effectiveness \leq et$ **then**
    end of the conversation, quit the algorithm
  compute gaze state of the Listener

---

the Speaker's HPTS++ probabilities and the *effectiveness* of conversation accordingly. This computation is done only when the Speaker is gazing at the Listener. The level of *effectiveness* of conversation is compared with the threshold et. If it is lower the Speaker ends the conversation. Otherwise the next Speaker's and Listener's gaze state are decided based on the probabilities just computed. Since they use different parameters (G_{S0},G_{S1}, G_{L0},G_{L1}) and algorithms, the Speaker and Listener behaviors will be different if their roles are exchanged.

5 Conclusion and Future Works

In this paper, we have presented capabilities that an ECA requires to be able to start, maintain and end a conversation. We addressed in particular the notion of engagement from the point of view of the Speaker and Listener. We have also presented our preliminary developments toward such a model.

In the future we aim at considering other modalities than gaze in our algorithms. HPTS++ allows one to have a common component for all the communicative modalities of the agent and easily define relations of coordination and synchronization between them. So it will be possible to create some new state machines for hand gestures for example or for facial expressions and let them run in parallel to generate consistent multimodal agent's behavior.

In the current state of our model, we do not consider the agents' mental and emotive states in consideration. But an effect of Listener's lower level of interest for the conversation may be to make the Speaker in a negative emotional state. Our model should not consider simply behavior information but also cognitive and emotional information of the agents.

Finally we will like to try out different ways for the Speaker to get the attention of the Listener by transgressing some communicative and social rules. Distractors could be applied such as making a strange noise, not gazing in a direction when expected [**?**].

6 Acknowledgements

We thank Stéphane Donikian and Fabrice Lamarche for letting us use the HTPS++ toolkit. We are grateful to Nicolas Ech Chafaï for discussing with us on this paper. This work has been partially funded by the Network of Excellence Humaine (Human-Machine Interaction Network on Emotion) IST-2002-2.3.1.6 / Contract no. 507422 (http://emotion-research.net/).

References

1. J. Allwood. Bodily communication dimensions of expression and content. In I. Karlsson B. Granstrm, D. House, editor, *Multimodality in Language and Speech Systems*, pages 7–26. Kluwer Academic Publishers, 2002.
2. S. Baron-Cohen. How to build a baby that can read minds: cognitive mechanisms in mind reading. *Cahiers de Psychologie Cognitive*, 13:513–552, 1994.

3. A.C. Beall, J. Bailenson, J. Loomis, J. Blascovich, and C. Rex. Non-zero-sum gaze in immersive virtual environments. In *Proceedings of HCI International*, Crete, 2003.
4. J. Cappella and C. Pelachaud. Rules for responsive robots: using human interaction to build virtual interaction. In Reis, Fitzpatrick, and Evangelisti, editors, *Stability and change in relationships*. Cambridge University Press, New York, 2001.
5. J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjlmsson, and H. Yan. Embodiment in conversational interfaces: Rea. In *CHI*, Pittsburgh, PA, April 15-20 1999.
6. C. Castelfranchi and I. Poggi. Bugie finsioni sotterfugi. In *Per una scienza dell'inganno*. Carocci, Roma, 1998.
7. M. F. Cohen, R. A. Colburn, and S. M. Drucker. The role of eye gaze in avatar mediated conversational interfaces. In *Technical Report MSR-TR-2000-81*. Microsoft Corporation, 2000.
8. R. Conte and C. Castelfranchi. *Cognitive and Social Action*. University College, London, 1995.
9. P. Ekman. *Telling lies*. New York: Norton, 1985.
10. A. Fukayama, T. Ohno, N. Mukawa, M. Sawaki, and N. Hagita. Messages embedded in gaze of interface agents — impression management with agent's gaze. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 41–48, New York, USA, 2002. ACM Press.
11. M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M.A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the conference on Human factors in computing systems*, pages 529–536. ACM Press, 2003.
12. E. Goffman. *Forms of Talk*. Oxford: Blackwell, 1981.
13. F. Lamarche and S. Donikian. Automatic orchestration of behaviours through the management of resources and priority levels. In *Proceedings of Autonomous Agents and Multiagent Systems (AAMAS'02)*, Bologna, Italy, July 15-19 2002. ACM.
14. A.M. Leslie. The perception of causality in infants. *Perception*, 11(2):173–186, 1982.
15. C. Pelachaud and M. Bilvi. Modelling gaze behavior for conversational agents. In *proceedings of the IVA 2003 conference*. Springer LINAI Series, 2003.
16. C. Pelachaud, V. Carofiglio, B. De Carolis, and F. de Rosis. Embodied contextual agent in information delivering application. In *First International Joint Conference on Autonomous Agents & Multi-Agent Systems (AAMAS)*, Bologna, Italy, July 2002.
17. C. Peters, C. Pelachaud, E. Bevacqua, M. Mancini, and I. Poggi. Engagement capabilities for ecas. In *AAMAS'05 workshop Creating Bonds with ECAs*, 2005.
18. I. Poggi. *Mind, hands, face and body. A goal and belief view of multimodal communication*. To be published, Forth.
19. I. Poggi and C. Pelachaud. Signals and meanings of gaze in animated faces. In P.McKevitt, S.Nuállain, and C.Muhlvihill, editors, *Language, Vision and Music*. John Benjamins, Amsterdam, 2001.
20. I. Poggi, C. Pelachaud, and B. De Carolis. To display or not to display? towards the architecture of a reflexive agent. In *Proceedings of the 2nd Workshop on Attitude, Personality and Emotions in User-adapted Interaction. User Modeling 2001*, Sonthofen (Germany), 13-17 July 2001.
21. I. Poggi, C. Pelachaud, and F. De Rosis. Eye communication in a conversational 3d synthetic agent. In *AI Communications*, volume 13, pages 169–181. IOS Press, 12 2000.
22. C. L. Sidner, C. D. Kidd, C. Lee, and N. Lesh. Where to look: A study of human-robot interaction. In *Intelligent User Interfaces Conference*, pages 78–84. ACM Press, 2004.
23. K.R. Thórisson. Layered modular action control for communicative humanoids. In *Computer Animation'97*, Geneva, Switzerland, 1997. IEEE Computer Society Press.

24. M. Turk, J. Bailenson, J. Blascovich, and R. Guadagno. Multimodal transformed social interaction. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, 2004.
25. R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 301–308, New York, NY, USA, 2001. ACM Press.
26. R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung. Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 521–528, New York, NY, USA, 2003. ACM Press.