Engagement Capabilities for ECAs

Engagement Capabilities for ECAs

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 Roma Tre
poggi@uniroma3.it

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 faced is to have conversations last more than a few seconds 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. Engagement is viewed, by Sidner et al. [26], as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake". In our terms [11, 21], 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".

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 sender/addressee model, but rather illustrate how the concepts may cluster together to form the core of such a model, and briefly describe some of our works aimed at implementing some parts of the whole model. In particular we concentrate in this paper in a backchannel model that witness for the Addressee's engagement.

In Section 2 we will consider some of the aspects required for a complete model. We will then present the steps involved in engagement 3. We will provide a definition of backchannels signals and a description of their properties 4.2. We will also discuss those capabilities on which we have thus far focused our research efforts in Section 4.

2. Computational Domains

ECAs are entities endowed with dialogic and expressive capabilities. They are used in interactive systems in which they can communicate with users. Being involved in a communicative process involves not only the generation of signs but also the perception of signs.

In a conversational setting, speakers and listeners are active and synchronized participants. The speaker conveys his goal to communicate through verbal and nonverbal means. An ECA, when being a speaker, ought to be able to convey communicative behaviors expressively; and when it is a listener, just as the Human listener perceives this communication [21], the ECA-listener should be able to detect, perceive and interpret behaviors.

These human-like communicative and conversational capabilities cover qualities over three main computational domains (see Figure 1): perception, interaction and generation.

Within the **perception domain**, the ECAs have the capability to perceive the users, objects and events in the real world, or, identically in the virtual world, to perceive other virtual agents, events and objects [20]. They are able to perceive attentional and emotional signals .

Within the **interaction domain**, ECAs have the capability to interpret the perceived signals. While in the perception domain, ECAs are endowed with the function of

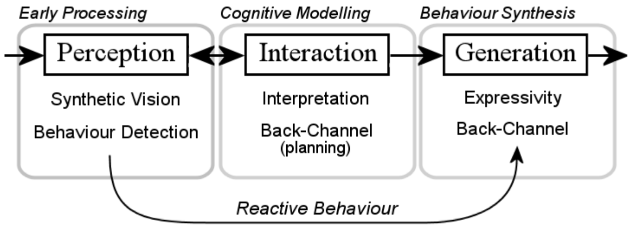

Figure 1. Overview of the model for both speaker and listener, with key capabilities listed.

processing and filtering the signals, in the interaction domain, they are able to interpret and be aware of the signals. Within the interaction domain, both speaker and listener exchange signals. The listener sends feedback to the speaker on how she understands, agrees, or even feels sympathy for what the speaker is saying. The speaker looks for these feedbacks; they allow him to adapt his discourse and communicative style by, for example, rephrasing what he has just said if the listener has not understood something, by emphasizing, or perhaps, by modulating what he says [26]. Adaptation does not involve solely feedback management, it can be done through the imitation of behaviors [3], by showing empathy [18], by changing discourse plan [15].

In the **generation domain**, an ECA should be able to display expressive synchronized visual and acoustic behaviors.

Capabilities of ECAs, such as being friendly or emphatic, may cover one or more of these computational models. In particular, the modelling of engagement covers the three main domains. The agent should show attentive behavior and awareness when an interaction starts or should start (see Section 4.1), identically when an interaction ends or should end. Keeping the engagement alive requires for the speaker to show expressive behaviors as well as to adapt to the listener's feedback; for the listener it requires to show feedback on what is being said. In this paper we viewed the modelling of feedbacks on two levels: the cognitive level (see Section 5) that involves the modelling of the mental states of the interactants as well as of their goal of communicating, and the reactive level where behaviors of an agent is triggered based on behaviors of the other agent (see Section 5.1). In the next section, we specify which human-like qualities an ECA should have to be able to interact with a user or with another agent.

3. Communication as an act of influence

According to the model we adopt [8], Communication is an activity through which two or more Agents aim at influencing each other. Any time a Speaker (or, more generally, a Sender, since we communicate through both verbal and non-verbal means) communicates something to a Listener (or Addressee), he does so to have the other do something: to ask the Addressee to adopt the Sender's goal, to help him achieve it. In any speech act (or communicative act) we perform, we pursue goals of at least two different kinds. On the one hand, some "central" goals that we ask the other to fulfil - to obtain information with a question, give information with an assertion, have someone do something with a request - ; but we also have some "side" goals that we do not mention explicitly but that are in some way relevant to the explicit goals of our sentences; they are called system constraints by Goffman [14], and control goals by Castelfranchi and Parisi [8]: the goals of knowing if our interlocutor is following, understanding, if he is willing to do what we ask him to... In a word, the goals of knowing if all the preconditions for our act to be successful are fulfilled. Actually, what are the "central" and the "control" goals of a communicative act depends on the kind of act it is. For example, since a question's central goal is to get an information, if the Sender puts a question, to adopt the question's central goal means for the Addressee to provide that information (this is why conversational acts are generally organized into so-called adjacency pairs [25]). But if the Sender is making a request, whose central goal is to have the Addressee do something, if the Addressee only tells him he did not hear his sentence, he is only adopting his control goal. More: if one is expressing his opinions, telling him you agree may simply be an adoption of his control goals; but if he is making a proposal, an act of agreement or disagreement is just what he is asking.

3.1. From engagement to interaction

Communication is a two-partners activity. To succeed, it requires that the Addressee use his resources of attention, perception and intelligence, to understand what the Sender is trying to communicate. 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 his resources to understand what he is communicating. In other words, communication is not worthwhile if 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 first the Sender needs to assess the possibility of engagement in interaction; and then, to check if engagement is lasting and sustaining conversation.

Of these two phases, in the former (that we may call *establish phase*) the prospective Sender must decide by himself whether to engage in conversation, by assessing the prospective Addressee's level of attention (see Section 4.1); in the latter (*maintain phase*), the Sender can be helped in maintaining the interaction and making it effective by the Addressee's backchannel (see Section 4.2): the signals produced by the Addressee to make the Sender aware if he is really paying attention, listening to, understanding and agreeing with what is being said.

In this work we focus our efforts on both moments of the check for the Addressee's attention. In the establish phase the Sender can detect the Addressee's attention in order to decide whether or not to engage in a conversation. In the maintain phase the Sender detects and interprets gaze behavior of the Addressee, as just one example of backchannel signals.

3.2. Six steps for the Addressee

For communicative interaction to go on, the Sender produces his communicative signals while 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 memorize 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 unbelievable what the Sender said, 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 In order to respond to the Sender's communicative acts, and then to accept engaging in and maintaining the interaction, the Addressee must decide whether to do what the Sender is requesting. Since, as we saw, the Sender has "central" ad "control" goals, the Addressee may adopt both, or only the control goals, or none. For example, if the Sender has put a question, the Addressee will provide the requested answer if he wants to adopt the Sender's central goal; he may tell he didn't understand the question, if he decides only to adopt his control goals; if he does not want to engage in the interaction, he may just get away or keep silent, [22]; again, if he is very much centered on his own problems, he may make a digression on an unrelated topic, or finally he may perform another communicative act that is related to the preceding ones, but takes the initiative: his communicative goal is more driven by his internal reactions than by the goal of adopting the previous Sender's goals. This decision may be determined by a number of factors, among which the Addressee's internal emotional reactions about the Sender's communicative act, his goals, his social relationship with the Sender, and so on.

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.

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;

2. in the course of interaction, to monitor the level of engagement of the Addressee and the effectiveness of the interaction.

Thus far, we have focused our efforts on both these capabilities that will allow ECAs to be aware of, interpret and generate important behaviors related to engagement.

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 [23]. 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 [14].

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 [16], 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 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], which stores information on whether the agent thinks the other has seen it and what their level of interest is. These facilitate the process of inferring the intention of the other to interact: in our model, a high level of interest is linked with 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 target also wants to interact. A more complete model would also need to consider other aspects, such as the context, in order to establish such theories.

4.2. Backchannel

A fundamental aspect of a model for the engagement is the definition of the behavior of Addressee. In a conversation the interlocutors must provide signals in order to make the Sender aware that they are really paying attention, listening, understanding and furthermore that they are agreeing or not. That is, the interlocutors are called to inform the Sender about the smooth flowing of the processes necessary to communicative interaction: attention, perception, comprehension and internal reactions. These signals are called backchannels.

4.2.1. Defining Backchannel As we mentioned, conversation is an interaction in which Sender try to have their central and control goals adopted by each other. In fact, as a Sender performs a communicative act, beside the act's central goal, he also has the goal to know if this act was successful. An Addressee, while becoming a Sender himself, can choose either to adopt or not to adopt the previous Sender's goals. If he does, he can adopt either his central or his control goals, while if he does not he'll choose disengagement, digression, or to take the initiative. When the Addressee adopts the control goals of the Sender, we can say that he is providing feedback (see [1] for this notion). However, the Addressee can provide feedback in many ways: he may do so by taking the floor with a whole speaking turn, for example directly by saying "I see what you mean", or indirectly with a new sentence that mirrors what the Sender has said [24], or even without full awareness or intentionality, by getting asleep or unwillingly closing his eyes during a teacher's explanation. But in some cases the Addressee consciously and deliberately, or at times automatically, provides feedback without interrupting the Sender: he does so by using a gaze, a head nod, a vocalization like "mhm". We may call feedback signals the communicative acts that fulfil the Sender's control goals by exploiting a whole turn, while we call backchannel the feedback signals that do not fill in a turn. The lack of backchannel makes a dialogue quite difficult and unsatisfying since the Sendere cannot understand if the Addressee is paying attention or not; he really needs a backchannel and, therefore, providing feedback is not only a matter of synchronization but also of politeness [4].

4.2.2. Types of backchannel Depending on the type of control goal adopted, we can distinguish backchannels of

1. attention,
2. comprehension,
3. believability,
4. interest,
5. agreement.

Further, a backchannel will be positive or negative depending on whether the Addressee communicates he is or not attentive, understanding, believing, interested, agreeing. Actually, the Addressee might pay attention but not understand, understand but not believe what the Sender is saying, and might believe but not agree, that is, he might think what is being said is true (plausible if compared to the Addressee's previous beliefs) but not right, good or acceptable (contrasting with his goals or values). A special point must be made about signals of agreement. A communicative act of agreement is not always a backchannel. According to models that view conversation as a common ground on top of which new contributions are offered [9], one could distinguish a "backward looking" from a "forward looking" engagement, with the former barely relative to an acknowledgment of what has been said, and the latter putting something new forward. But according to our model, in which the Sender is asking the Addressee to fulfill his goals (to do, to tell, to believe), an act of agreement will be a backchannel or not depending on the particular communicative act that precedes it. So, the specific backchannels produced - and whether a communicative act is to be considered a backchannel or not - depend on the type of Speech Act they respond to: a signal of agreement/disagreement will typically follow the expression of opinions, evaluations, planning, and in this case it will be seen as a backchannel; but it will count as adoption of the Sender's central goal if performed after a request for approval. In the same vein, a signal of comprehension/incomprehension will not be a backchannel after an explicit question like "Did you understand?", while it will be so during a description or an explanation. Moreover, some control goals, and hence the corresponding backchannels, may be more or less salient in different types of interaction. For instance, the symmetry/a-symmetry factor affects salience of control goals and frequency of backchannels: in primary school it is very rare for pupils to provide the teacher backchannels of agreement, given the high unbalance of interactional power [27, 17]. This also may account for the fact that some backchannels, though being polysemic, that is, possibly providing different types of information, are usually well interpreted in context. For example, a frown can give a negative feedback as to three different control goals: understanding, believing and agreeing. A frown literally means "this is difficult (to understand)", and thus it can be a backchannel of incomprehension; but it can also be an indirect polite way to communicate "this belief you are communicating does not fit with my previous beliefs, hence it is unbelievable", thus being a backchannel of believability; or it can mean "it does not fit with my opinions, hence I don't agree with what you're saying" - thus being a backchannel of agreement. Actually, the specific backchannel information that the frown provides in each situation can be interpreted on the basis of the present interaction and of the preceding speech act. A positive back-channel of comprehension, believability, and agreement, somehow correspondent to a de-intensified head-nod [6] can be performed through a light closing of the eyelids. This could be paraphrased as "ok, I am following you" (or "I believe what you're saying", or "I agree"), but with a somewhat snobbish nuance: like if saying "Though being so much more important than you, I very graciously accept to follow (believe, approve) what you are saying". This back-channel is typically used when the Addressee feels to have a higher status than the Sender, and implies it is a very gracious gift for him to lower himself and listen; so it will be used in a-symmetric interactions. Finally, during real interaction backchannel generally uses a combination of signals. For instance, to show you don't trust what is being said, a negative backchannel of believability, you can incline your head while staring obliquely and frowning to the Sender: two gaze signals combined with a head signal.

4.2.3. Backchannel signals of gaze Gaze and, more generally, direction of attention, is an especially important way of providing feedback and subtle signalling. In terms of engagement, it can represent a useful way in which an Addressee may let the Sender (and other Addressees) know their level of engagement or intention to maintain engagement without the need to be especially explicit and interrupt the flow of the conversation. For example, an Addressee that needs to disengage from a conversation may still gaze at the Sender and nod, but may start to orient their body away from him in order to pave the way for their exit. Looking behaviors are also of vital importance. Mutual eye contact, that is usually associated with increased psychological arousal, establishes a special connection between Sender and Addressee where each is the object of the others attention. As pointed out in [26], the more people share looking behaviors, the more they are involved and coordinate in the conversation. Our emphasis here is that Addressees also communicate feedback on their engagement in the interaction, and can do so without the necessity to interrupt the

Sender every time. This may not necessarily involve mutual eye contact with the Sender: during shared attention situations involving another object or entity, the Addressees may actually signal their engagement in the situation by directing their attention away from the Sender and at the object in question. The key elements in our backchannel gaze model consist of duration of and direction of attention, cued by the eyes, head and body, and gaze duration (mutual or otherwise) and relate closely to those in Section 4.1. It is therefore the task of the agent providing the backchannel to try to send those signals that the receiver will perceive as being in line with their intention: if they wish to end the engagement, they should not send signals that will increase or maintain a strong attention profile in the other agent. In this way, we view gaze as fulfilling a dual role in relation to engagement: it provides a way of signalling ones own intention to engage or remain engaged with others, while at the same time, as mentioned in Section 4.1, it is used to monitor the gaze behaviors of others in order to establish their intention to engage or maintain engaged.

5. Backchannel computational models

Backchannel strongly depends on the context, the Addressee acts according to what the Sender tells and does and the feedback can be more or less intentional. Previous research on bodily communication [1] has suggested that there are different degrees of awareness and intentionality in gestures made by someone while interacting with other people.

Backchannel feedback can be unaware or conscious and to describe the possible reactions of a Addressee a single backchannel model is not enough. We need two computational models, respectively a reactive model and a cognitive model.

The reactive model generates an instinctive feedback, provided by the Addressee without reasoning. Usually, during a conversation, a lot of behavioral decisions are made by the Addressee, often in such a short time that he is not even aware of them. He reacts instinctively to the Sender's behavior or speech, generating backchannel signals unawarely.

Instead, the cognitive model generates a reasoned feedback. The Addressee consciously decides to provide a backchannel feedback in order to provoke a particular effect on the Sender or to reach a specific goal. The latter model can be very complicated and sometimes even not applicable. In fact, to elaborate reasoned reactions from an Addressee, notions about his personality and temper are needed. If we lack such information we have to look to the reactive model.

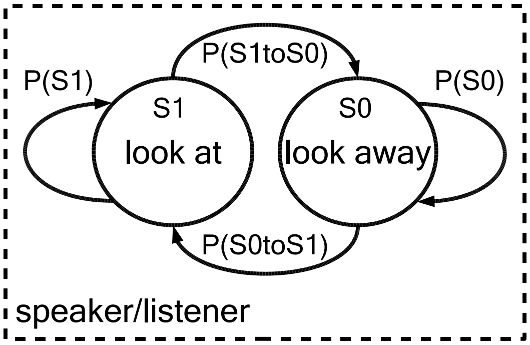

Figure 2. State machine for Speaker/Listener

5.1. Reactive model

In this paper we focus primarily on gaze. Gaze is an especially important way of providing feedback and subtle signalling. Through it an Addressee can show his level of interest and engagement. For example, an Addressee that needs to disengage from a conversation may start to advert 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 Sender: during shared attention situations involving another object or entity, the Addressee may actually signal their interest in the situation by directing their attention away from the Sender and at the object in question [26]. To simulate human-like quality, the gaze patterns of an ECA should be influenced by factors such as the general purpose of the conversation (persuasion discourse, teaching...), personality, cultural root and social relations. Not having precise information on the influence that each of these factors may have on gaze behavior, we are trying to find out a possible set of parameters that will enable one to qualitatively alter the gaze behavior itself.

In our previously developed model [19] we developed a computational model of reactive gaze behavior using Bayesian Belief Networks. We implemented this model starting from statistical data reported in [5], corresponding to the annotation of body behaviors

(gaze direction, head nods, back channels) of two subjects having a conversation.

Currently we are beginning to use a different model based on state machines defined using HPTS++ [12]. In this model the Speaker (or Sender) and the Listenet (or Addressee) gaze behavior is described as shown in Figure 2. The agent can either be in a state *S1* in which he looks at the other agent or in a state *S0* in which he looks away. Some important aspects of this kind of modelling include for example the possibility to simulate several situations without defining complex transition tables as it is needed for the Belief Network model. For example if we want to calculate the gaze of a single speaker talking to multiple listeners we will just instantiate one state machine for each one of the participants to the conversation and let the system elaborate gaze through time. Moreover the HPTS++ implementation layer will be the common ground component for all the communicative modalities of the agent and will enable one to easily define relations of coordination and synchronization between them. At the implementation level of the proposed HPTS++ model the probabilities *P(S0)*, *P(S1)*, *P(S0toS1)* and *P(S1toS0)* that determine the transitions (and so the gaze behavior) inside the state machines will be dynamically calculated from the same statistical data in [5] and by considering for each machine the actual state of all the other machines. For example, if we want to simulate a shy agent who glances very rapidly at the interlocutor, we will low the probability of events S1 and S0toS1 respect to the probability of S0 and S1toS0.

6. Related work

Past researches on ECAs have provided a first approach to the implementation of a backchannel model. 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 [28]. Gandalf can show and recognize information like head movements or short statements, using it to perceive and generate backchannel feedback.
Another backchannel model was proposed by Cassell [7] and adopted in Rea. Rea 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 a short statements such as “I see”.

Other models have been developed for controlling gaze behavior of ECAs conversing with other ECAs. For example the models of Colburn et al. [10] and Fukayama et al. [13] 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).

Most of the models provide a reactive feedback, the virtual agents do not act on the base of an inner reasoning about what the speaker is saying or what they want to achieve, but only on particular behaviors of the speaker that usually in real world induce the listener to provide a backchannel feedback.

7. Conclusion

In this paper, we have presented capabilities an ECA requires to be capable of starting, maintaining and ending 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.

8. Acknowledgements

We are indebted to Candace Sidner and anonymous reviewer for their insightful comments, which contributed to clarify some points of this paper.

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. J. Blanchard and L. Canamero. Using visual velocity detection to achieve sychronization in imitation. In *Workshop Imitation in AISB'05*, 2005.

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

[5] 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.

[6] J. Cassell. Embodied conversational interface agents. *Communications of the ACM*, 43(4):70–78, April 2000.

[7] 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.

[8] C. Castelfranchi and D. Parisi. *Linguaggio, conoscenze e scopi*. Il Mulino, Bologna, 1980.

[9] H.H. Clark and E.F. Schaefer. Contributing to discoruse. In *Cognitive Science 13*, pages 259 – 294, 1989.

[10] 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.

[11] R. Conte and C. Castelfranchi. *Cognitive and Social Action*. University College, London, 1995.

[12] Stéphane Donikian. Hpts: a behaviour modelling language for autonomous agents. In *AGENTS '01: Proceedings of the fifth international conference on Autonomous agents*, pages 401–408, New York, USA, 2001. ACM Press.

[13] 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.

[14] E. Goffman. *Forms of Talk*. Oxford: Blackwell, 1981.

[15] J. Gratch and S. Marsella. Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proceedings of the 5th International Conference on Autonomous Agents*, Montreal, Canada, May 2001.

[16] A.M. Leslie. The perception of causality in infants. *Perception*, 11(2):173–186, 1982.

[17] F. Massimi. Il linguaggio verbale e non verbale degli alunni come feed back per l'insegnante. Unpublished Degree Thesis, Universita' Roma Tre, 2004.

[18] A. Paiva, J. Dias, D. Sobral, S. Woods, and L. Hall. Building empathic life-like characters: the proximity factor. In *Workshop on Empathic Agents, AAMAS'04*, 2004.

[19] C. Pelachaud and M. Bilvi. Modelling gaze behavior for conversational agents. In *proceedings of the IVA 2003 conference*. Springer LINAI Series, 2003.

[20] C. Peters. Towards direction of attention detection for conversation initiation in social agents. In *Workshop Social Presence Cues for Virtual Humanoids in AISB05*, 2005.

[21] I. Poggi. *Mind, hands, face and body. A goal and belief view of multimodal communication*. To be published, Forth.

[22] I. Poggi, C. Castelfranchi, and D. Parisi. Answers, replies and reactions. In M.Sbisa H.Parret and J.Verschueren, editors, *Possibilities and Limitations of Pragmatics*.

[23] 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.

[24] Carl R. Rogers. *Client-centered Therapy*. Boston: Houghton Mifflin Company, 1951.

[25] H. Sacks, E.A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50:696–735, 1974.

[26] 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.

[27] A. Stefanini. Il back-channel nell'interazione in classe. Unpublished Degree Thesis, Universita' Roma Tre, 2001.

[28] K. R. Thórisson. *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills*. PhD thesis, MIT Media Laboratory, 1996.