Computational Model of Listener Behavior for Embodied Conversational Agents

Elisabetta Bevacqua

DISSERTATION.COM

Boca Raton

Computational Model of Listener Behavior for Embodied Conversational Agents

Dissertation.com
Boca Raton, Florida
USA • 2010

ISBN-10: 1-59942-335-9
ISBN-13: 978-1-59942-335-7

Special Thanks

My most special thanks go to my supervisor, Catherine Pelachaud. I am very grateful for her precious help and indispensable guidance. Thank you, Catherine, for not having given up on me even in the hardest times, for having believed in me when I did not. Thank you for having been a guide not only in work matters but also in life. I am very glad that our ways met, I feel deeply lucky for that.

Many sincere thanks to Maurizio Mancini, for all that he taught me about programming and debugging. For his friendship, infinite patience and support, for all his precious advice and for having forced me to start my PhD manuscript.

I am very thankful to all my colleagues in Greta's Team for their support and precious friendship. Particularly, thanks to Sylwia Hyniewska and Marion Tellier for helping me in analyzing the statistical data and to Magalie Ochs who taught me how to prepare a presentation worth of this name. Thank you, my friends, for having spent days in reassuring and encouraging me before the final defense. It meant a lot to me.

I would like to thank also Louis-Philippe Morency for his collaboration in connecting Watson to Greta and Samer Al Moubayed and Malek Baklouti for helping me in connecting Greta with the software developed at the eNTERFACE Summer School.

Thanks to all the participants in the perceptive tests.

Last but not least, I want to thank Nick Campbell, Stefan Kopp, Dirk Heylen, Isabella Poggi, Jean-Claude Martin and Marta Olivetti Belardinelli, the members of my jury, for the interest shown in my work and for their encouraging words and useful advice.

Contents

Chapter 1

Introduction

1.1 Context

In recent years, computer science developers have focused on a new type of Human-Computer Interface to increase the quality of communication between humans and computers. Studies show that people tend to interact with computers as if they were humans [Nass et al., 1997]. Unconsciously, they apply social rules even if they believe that such an attribution is not appropriate. For such a reason the new type of humane-machine interfaces consist of systems that are able to simulate human-like interactions. This level of consistency could be reached using humanoid artifacts able to apply that rich style of communication that characterizes human conversations. Called Embodied Conversational Agents (ECAs), these humanoid interfaces can be able, on the one hand, to simulate human verbal and non-verbal behavior (like speaking in natural language, performing gestures, displaying facial expressions, shifting their gaze and moving their head like humans do in everyday life) and, on the other hand, they can "understand" what humans say,

interpret their non-verbal signals and use all this information to decide how to react and respond.

The first ECA systems aimed to reproduce entities with a humanoid aspect endowed with basic dialogic and expressive capabilities. They could produce speech along with some non-verbal behaviors, like facial expressions and gestures. Researchers have focused on the importance of the ECA's behavior while speaking, implementing models for the generation of an always richer set of verbal and non-verbal signals. However, recently, researchers have seen that ECAs able to behave appropriately while listening are equally important [Gratch et al., 2006, Cassell and Thórisson, 1999]. While communicating with others, people's goal is to transmit, both consciously and non-consciously, some information from their mind to the others' minds. To reach such a goal, they need to know if the other party is listening, understanding and paying attention to what they are saying. Listeners do not assimilate passively all the speaker words, but they actively participate in the interaction providing information about how they feel and what they think of the speaker's speech. Listeners emit expressions like "a-ah", "mmh" or "yes" to assure grounding, nod their head in agreement, smile to show their liking, frown if they do not understand and shake their head to make clear that they refuse the speaker's words. According to the listener's behavior, the speaker can estimate how his interlocutor is reacting and can decide how to carry on the interaction: for example by interrupting the conversation if the listener is not interested or re-formulating a sentence if the listener showed signs of not understanding. As a consequence, to become more credible and usable, ECAs must fit well also in the role of the listener. Many researchers have attempted to model this type of behavior in ECAs, trying to determine *when* the agent should provide the information and *which* verbal and non-verbal signals it should produce [Beun and van Eijk, 2004, Cassell and Thórisson, 1999, Cathcart et al., 2003, Edlund and Heldner, 2006, Fujie et al., 2004, Gratch et al., 2007, Heylen et al., 2004, Kopp et al., 2008, Morency et al., 2008, Pelachaud, 2005, Thórisson, 1997, Ward and Tsukahara, 2000]. All of these works have obtained positive results, even if reproducing the behavior displayed by listener during an interaction is a quite hard task. For example, listeners' signals strongly depend on several factors, like the content of the speech, the relationship between participants, the

listener's goals, beliefs, emotional state and cultural heritage.

1.2 Problem definition

According to Gumperz [Gumperz, 1982] communication takes place when a response is elicited. A sentence cannot be considered as a real communicative act if it does not generate an effect, eliciting an answer from the person who is listening. In conversation the speaker relies on signals emitted by the listener to know if he is listening or not, understanding or not, agreeing or not, etc. This informs the speaker on the success or failure of the communication and helps him to decide how to carry on with the interaction.

Many studies have been conducted on the behavior displayed by the listener during conversations in order to understand and define the typical dynamics that people use. These studies highlighted that listeners show their participation through linguistic and gestural signals, called "backchannels" by Yngve [Yngve, 1970], "minimal responses" by Coates [Coates, 1993], "receipt tokens" by Heritage [Heritage, 1984]. These signals help to support the interaction itself and to go on [Goodwin, 1981, McNeill, 1992]. Allwood et al. [Allwood et al., 1993] studied listeners' behavior from a linguistic point of view. They considered listeners' signals as mechanisms which allow interlocutors to exchange information about the four basic communicative functions: contact, perception, understanding and attitude. In this thesis, to refer to signals provided by listeners, we adopt the term backchannel proposed by Yngve [Yngve, 1970]. We define backchannels as acoustic and non-verbal signals provided during the speaker's turn to exchange information about communicative functions [Allwood et al., 1993]. Backchannels are emitted in a non-intrusive way, that is without interrupting the speaker's speech.

Verbal and non-verbal information sent by the speaker is firstly perceived by the listener and then processed through a more aware evaluation [Kopp et al., 2008]. Different types of backchannel signals can be emitted during both stages of the reception of the speaker's message. At the first stage, a non-conscious appraisal of the perceived information can generate automatic behavior, for example to show lack of contact or perception. At a second stage, the more aware evaluation, involving memory, understanding and other cognitive processes, can generate signals to show understanding and other attitu-

dinal reactions, like acceptance or rejection, belief or disbelief and so on. In this work we are interested in implementing a system able to generate both types of backchannel and to refer to them we adopt the terms proposed in [Kopp et al., 2008]: we call *reactive* the behavior derived from perception processing and *response* the more aware behavior generated by cognitive processing.

Several researchers [Allwood, 1976,Ekman and Friesen, 1969,Poggi, 2007] have noticed that people perform communicative acts with different levels of awareness and intentionality both while speaking and listening. Listeners can emit signals without awareness, non-consciously: they react instinctively to the speakers behavior or speech, generating signals at a very low level of control. For example, the interlocutor can non-consciously yawn, showing unintentionally that he is bored and not interested. Listeners can also decide to consciously emit a signal in order to show their reaction to the speakers speech and even act with the intent to influence the speaker's behavior; for example, making him repeat his words when these have not been understood. In this work we are interested in creating a system able to generate listener's signals that are emitted at a very low level of awareness. We do not take into account the agent's intention to consciously provide a backchannel signal; in our system backchannels are emitted unintentionally. As we will see in Chapters 4 and 7, the triggering of a backchannel signal depends solely on the speaker's acoustic and non-verbal behavior.

A particular form of backchannel is the mimicry of the speaker's behavior. By mimicry we mean the behavior displayed by an individual who does what another person does [Van baaren, 2003]. This type of behavior has been proven to play quite an important role during conversations. First of all, when present, it makes the conversation run more smoothly [Chartrand and Bargh, 1999], and helps to regulate the conversational flow. For example, listeners often mirror speaker's postural shifts at the end of a discourse segment and this helps the exchange of speaking turns [Cassell et al., 2001].

All the studies above show that the behavior performed by listeners during an interaction is a fundamental part of human-human communication. To investigate how this type of behavior can be replicated in ECAs, we propose the implementation of a system that generates listener's behavior for an ECA while interacting with a user. The

virtual agent will be able to perform backchannel signals according to both the user's acoustic and other non-verbal behaviors and what the agent "thinks" about the speaker's speech. In this work, we use the term "think" to refer to the agent's representation of the speaker's speech. The ECA will be endowed with a repertoire of backchannel signals. Moreover, the agent will be able to display three types of backchannel: response, reactive and mimicry. Finally, to provide a first approach to the problem of the repetitivity of the agent's behavior, the frequency of emission of backchannel signals will vary according to "user's estimated interest level" [Peters et al., 2005].

1.3 Objectives and means to achieve them

In the previous Section we propose to implement ECAs able to fit well in the role of the listener while interacting with a user. We aim to achieve the following objectives.

- At a conceptual level we want to model the listener's behavior, providing a platform to make people interact with a virtual agent. Such a platform will allow us to gather data and to study human behavior during an interaction. We aim to get a better understanding of human-machine interactions, particularly how to increase the users' feeling of engagement and how to make their conversation with a virtual agent more satisfying and useful.

- At the implementation level we want to build a real-time platform that can generate listener's behavior for a virtual agent while interacting with a user. To sustain human-like conversations, agents must react fast, complying with the speed of human-human interactions. As a consequence, a real-time platform is mandatory.

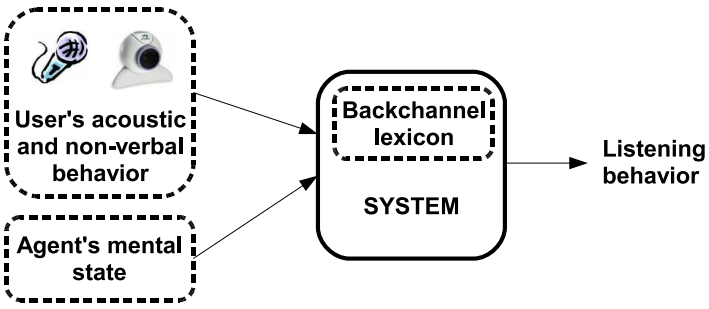

Figure 1.1: Conceptual diagram of the system.

The means to achieve our objectives are described in the conceptual diagram in Figure 1.1. In this diagram, we represent the connection between the user's acoustic and visual behavior and what the agent aims to communicate through its backchannel signals. We call the latter information the "agent's mental state", that is what the agent thinks about the user's speech. The white box, that represents the system, receives as input the user's behavior and the agent's mental state. It combines them in order to determine *when* the agent should display a backchannel and *which* signal (or combination of signals) it should perform. In this work, we aim to provide an implementation of a system that computes backchannel signals that follows the definition reported in Section 1.2. The diagram of Figure 1.1 shows our general framework and sums up the five main goals of this thesis:

- *Connecting external softwares to obtain information on user's behavior*: we aim to provide a simple solution to connect with our system several softwares that are able to analyze different features of the user's behavior. For example, head movements, facial expressions, acoustic features and so on. Data is sent to our system in HTTP format through a TCP/IP connection.

- *Defining a backchannel lexicon*: a repository that associates a set of non-verbal signals to each communicative function that the agent aims to convey through a backchannel. This lexicon has been determined through perceptive studies that analyzed both non-verbal signals performed by a virtual agent and the meaning they convey.

- *Representing the agent's mental state*: we aim to provide a representation schema of the agent's mental state. This representation will help the system to select the appropriate signal to display from within the backchannel lexicon: for example, to select a head shake when the agent disagrees with the user or a smile when the agent likes the user's speech.

- *Implementing a listener behavior generation system*: to compute listener's behavior the system has to determine *when* the agent should react and *which* signal (or combination of signals) it should perform. We propose a method to trigger a backchannel according to the user's acoustic and non-verbal behavior and to select

the appropriate non-verbal signals to display according to the agent's mental state.

- *Implementing a real-time system*: to be able to interact with users in a natural way, the system must generate the agent's backchannel signals complying with the speed of human-human interactions. Backchannels are often emitted very fast. For this reason we implemented a system able to generate backchannel signals in real-time.

Three additional goals are pursued in this thesis:

- *Evaluating our listener behavior generation system*: we aim to evaluate the quality of our system by performing perceptive studies.

- *Implementing a flexible system*: our system will be easily connected to different applications that analyze user's behavior. Moreover, the backchannel signals generated as output by our system can be displayed by both virtual and/or physical agents.

- *Implementing an extensible system*: we aim to create a system that is easily extensible. New communicative functions can be added to the agent's mental state. Moreover, the process to trigger a backchannel signal can be modified to take into account some characteristics of the agent. For example, the agent's personality or culture could influence when or/and how it reacts to the user's behavior.

1.4 Contributions

The contribution of this thesis is to define, implement and test an ECA system that computes backchannel signals for a listening agent during an interaction with a user. In more detail:

- We define a representation of the agent's mental state, that describes how the agent is reacting to the user's speech. We represent such a state as the list of communicative functions the agent could want to transmit through backchannel signals. From the literature [Allwood et al., 1993, Poggi, 2007] we selected twelve communicative functions: agree, disagree, accept, refuse, like, dislike, believe, disbelieve, understand, not understand, be interested and not interested. Moreover, the agent's

mental state contains the agent's mimicry level, a threshold that describes the agent's tendency to mimic the user's non-verbal behavior. We provide an XML-based language to define the agent's mental state.

- We propose a backchannel lexicon, a repository that defines the mapping between the agent's communicative functions and non-verbal signals. Such a mapping has been found through perceptive tests where subjects were asked to assign a meaning to a set of non-verbal signals (and combination of signals) displayed by a virtual agent. We found that to convey a certain communicative function different non-verbal behaviors can be performed, for example the agent can show its agreement by nodding its head or by nodding and smiling.

- We propose a representation language to describes the rules to trigger a backchannel signal according to the user's acoustic and other non-verbal behavior. For a given user's behavior the rule specifies the probability to trigger a backchannel signal. Such a probability allows us to define agents that react differently to a user during an interaction. For example, we can define agents that have high probability to provide many backchannels and that respond especially to the user's acoustic signals; or we can define agents that do not provide many backchannels and that react mainly to the user's head movements. Probabilities could vary according to agent's personality, mood and even culture. For example, in [Maynard, 1989], Maynard showed that Japanese people tend to perform many more backchannel signals than Americans do. By using an XML-based language, one can easily modify or extend the set of rules.

- We implement a system that computes the agent's backchannel signals according to the user's acoustic and non-verbal behaviors and to the agent's mental state. The agent can provide three type of backchannel signals: reactive, response and mimicry.

- We endowed our system with two important capabilities: adaptability and modularity. The system can be adapted to work with different levels of information about the agent's mental state. This data can be unavailable, or provided by systems able to compute the agent's mental state according to the available information about

the content of speaker's speech and the agent's goals and beliefs. Being modular, the system can be easily connected to different external softwares to receive data about the user's acoustic and visual behaviors; this allows the system to take into account further user's behaviors without requiring changes in the source code. Finally, modularity is useful also on the output side. The listening behavior generated by our system can be displayed by different virtual and physical agents.

- We evaluate the system through evaluation studies. We perform two types of tests:
  - we analyze the effect of the agent's mental state on the user-agent interaction through the backchannel signals. Does the user recognize the agent's mental state (if it is agreeing or disagreeing and so on)? Does this influence the interaction?
  - we analyze the effect of mimicry of the smile during user-agent interactions. The agent smiles either to mimic the user's behavior, or to show positive feedback. Does mimicry of the smile affect the interaction?

1.5 Thesis outline

This thesis is divided into 3 parts. In the first part we provide an overview of the background concepts we refer to in this work. Firstly, we describe the studies done on the backchannel performed by people while listening during a human-human interaction. Then, we present the context in which this thesis has been developed. We provide a definition of Embodied Conversational Agents presenting an overview of other ECA systems exhibiting listening behaviors while interacting with users. Some of them analyze the user's speech content to determine when a backchannel signal should be provided, whereas others look at the user acoustic and visual behavior to drive the agent's signals. This part of this dissertation is covered by Chapters 2, 3 and 4.

The second part of the thesis describes the steps that have been realized to develop the system. In Chapter 5 we present two perceptive tests we performed to study non-verbal backchannel signals. With these evaluations we aim, on the one hand, to define a set of backchannel signals that the agent can show during an interaction with a user, and on the

other hand we want to make sure that the user is able to interpret the agents signals as "intended" by the ECA. Chapter 6 gives an overview of our system for interactive ECAs. Our agent architecture follows the design methodology proposed in [Thórisson et al., 2005] and is compatible with the standard SAIBA framework [Vilhjálmsson et al., 2007]. Finally, we present our system in Chapter 7. We focus on the generation of backchannel signals starting from the user's acoustic and visual behavior and the agent's mental state during the interaction. The system must be able to: decide *when* a backchannel signal should be emitted and select *which* communicative functions the agent should transmit through the signal. The generation of backchannel signals is done through 4 steps.

In the third part of the thesis we present in Chapter 8 a study we conducted to evaluate our system. We have performed two evaluations: the first experiment aims at studying the effect of different agent's mental states on communication. With the second test, we want to evaluate the effect of the agent's smile on interaction. The agent smiles either to mimic the user's behavior, or to show positive backchannel.

Finally, in Chapter 9, we describe two application scenarios for our system that highlight the properties of our model: the modularity and the adaptability. The applications show that we can easily connect our system with different softwares that provide information about the speaker's behavior. Moreover, the backchannel signals generated as output by our system can be displayed by both virtual and/or physical agents.

Part I

Presentation

Chapter 2

Background

2.1 Human-human communication

Communication takes place when a response is elicited [Gumperz, 1982]. Therefore, human-human communication is not a one way phenomenon, participants do not alternate their roles like in a volley match, just uttering sentences while speaking and freezing while listening. Communication is much more like a dance, in which people move together, sometime they lead and sometime they follow, waltzing towards a common goal: exchange information, coordinate actions, display emotional state, establish a relationship. But human-human communication is also a social instrument that people use in order to make interlocutors fulfill their needs, for instance, to obtain information, to request someone to do something and so on [Castelfranchi and Parisi, 1980, Poggi, 2007]. These can be called the "central goals" of communicating. Yet, beside these, the speaker has some other needs, called "control goals" [Castelfranchi and Parisi, 1980], that are implicitly mentioned and that the speaker needs to have fulfilled, like the goals of knowing

if the listener is following, understanding, willing to do what the speaker needs him to do; in short the goal of knowing if the other is attending to the conversation. On the other hand the listener has to show his/her participation in the interaction in order to push it forward and make the speaker go on. In fact, whenever people listen to somebody, they do not assimilate passively all his/her words, but they actively participate in the interaction providing information about how they feel and what they think of the speaker's speech. Listeners emit expressions like "a-ah", "mmh" or "yes" to assure *grounding*[1], nod their head in agreement, smile to show liking, frown if they do not understand and shake their head to make clear that they refuse the speaker's words. In short, while listening people assume an active role in the interaction showing above all that they are attending the exchange of communication. According to the listener's behavior, the speaker can estimate how his/her interlocutor is reacting and can decide how to carry on the interaction: for example by interrupting the conversation if the listener is not interested or re-phrasing a sentence if the listener showed signs of incomprehension.

2.2 Listener behavior

Many studies have been done on the behavioral signals displayed by the listener during conversations in order to understand and define the typical dynamics that people apply. These studies have been conducted by Yngve, [Yngve, 1970], Allwood et al. [Allwood et al., 1993], Poggi [Poggi, 2007], Welji and Duncan [Welji and Duncan, 2004], Coates [Coates, 1993], Heritage [Heritage, 1984], Clark and Schaefer [Clark and Schaeffer, 1989], Schegloff [Schegloff and Sacks, 1982] and many more. In this thesis we focus on the theories proposed respectively by Yngve, Allwood et al. and Poggi, since we will base on their work our definition of the signals performed by listeners during a conversation. Such a definition, provided at the end of this Section, will be used all along this work of research.

One of the first studies about the expressive behaviors shown by people while interacting, has been presented by Yngve [Yngve, 1970]. His work focused mainly on those signals used to manage turn-change, both by the speaker and the listener. To describe this type of

[1]Grounding: the process by which partners in a dialogue gain *common ground*, that is mutual belief and shared conception [Clark and Schaeffer, 1989]

signals, Yngve introduced the term "backchannel". In this conception, backchannels are seen as non-intrusive acoustic and non-verbal signals provided during the speaker's turn. Backchannels are used to set the common ground without bringing much new information.

Allwood et al. [Allwood et al., 1993] extended Yngve's theory. They chose the term of *feedback* (originally introduced by [Wiener, 1948]) to encompass a wider set of verbal and non-verbal signals that the listener can provide. Feedbacks include not only non-intrusive acoustic and gestural signals, but also single words or short phrases like "I see", "Oh, that's great!" that can need a full speaking turn to be expressed. In particular, Allwood et al. analyzed the linguistic feedbacks from a semantic and pragmatic point of view, considering them as mechanisms which allow interlocutors to exchange information about four basic communicative functions [Allwood et al., 1993]:

- *Contact*: the interlocutor wants to and can continue the interaction,
- *Perception*: the interlocutor is willing and able to perceive the message,
- *Understanding*: the interlocutor is willing and able to understand the message,
- *Attitude*: the interlocutor wants to and can show her/his attitudinal reactions to the communicated content, that is if he believes or disbelieves, likes or not, accepts or refuses what the speaker is saying.

Within their studies, Allwood and colleagues showed that feedback behavior can convey more than one communicative function at a time, for example an interlocutor can show, through a feedback signal, that he is both understanding and agreeing. The meaning of a feedback strongly depends on the meaning and the polarity of the previous communicative act. If the locutor says "It's raining" and the interlocutor answers "no", this feedback means rejection. But if the speaker says "It isn't raining" the same feedback ("no") means acceptance.

Through feedback signals, a listener can transmit both positive and negative information [Poggi, 2007]. A speaker does not need to know only when the listener can perceive his words or understand the content of his speech, but he needs also to know when the perception decreases and when what he is saying is not understandable anymore. These negative signals help him to apply strategies to make the conversation successful again.

Like Allwood et al., Poggi sees feedback as a means used by listeners to provide information about some conditions of the of the communicative process [Poggi, 2007]: the functions of attention, comprehension, believability, interest and agreement. In addition, she defined feedback signals as a mechanism used by interlocutors to fulfill the other party's "control goals" during an interaction [Castelfranchi and Parisi, 1980, Poggi, 2007]. While talking, people aim to obtain or give information, have the interlocutors do something, in short they have some "central goals" they want to see fulfilled. But, at the same moment, people have also some side goals, called "control goals", that, even if not explicitly mentioned, are quite relevant: the goals of knowing if their interlocutor is listening, understanding, agreeing to do what they ask for, and so on. In this idea, feedback signals provide information about the fulfillment of these control goals, showing the interlocutor's level of comprehension and personal reaction.

Taking into account the theories proposed by Yngve, Allwood et al. and Poggi, we define hereafter the behavioral signals performed by listeners during a conversation. To refer to these signals we decide to adopt the term "backchannels" proposed by Yngve and we will stick to it all along this thesis.

> ***Definition of backchannel signal:*** *a backchannel is a verbal or non-verbal signal performed by the listener during an interaction to exchange information about the communicative functions: contact, perception, understanding and attitude. A backchannel is provided during the speaker's turn without attempting to take the floor.*

According to Yngve's definition, backchannel signals can be seen as a subset of the feedback signals proposed by Allwood et al. In fact, they are provided without taking the turn, instead Allwood's feedback contains also all those signals of a listener that can be provided in a full speaking turn. In particular, in this thesis, among all the attitudinal reactions showed by a listener, we decide to focus on agreement, belief, acceptance, liking and interest.

2.3 Level of awareness and intentionality

Several researches, that have studied the behavior shown by people during an interaction, have noticed that people perform communicative acts with different degrees of *awareness* and *intentionality* both while speaking and listening [Allwood, 1976, Ekman and Friesen, 1969, Poggi, 2007]. Allwood et al. [Allwood et al., 1993] proposed three levels of awareness for listeners:

- *indicative*, at this level the listener provides backchannel signals that he is not aware of,
- *displayed*, at this level the listener performs consciously backchannels but it is not necessary that the speaker recognizes and interprets them,
- *signaled*, at this level backchannel signals not only are emitted consciously but they are provided with the aim of influencing or modifying the interaction, the listener wants the speaker to perceive and interpret them.

Listeners can emit signals without awareness, non-consciously: they react instinctively to the speakers behavior or speech, generating backchannels at a very low level of control. For example, the interlocutor can non-consciously yawn showing that he is bored and not interested. The listener can show backchannels at the appropriate moment even without really attending to the speech content [Thórisson, 2002, Bavelas et al., 2000]. Listeners can also decide to consciously emit a signal in order to show their reaction to the speakers speech. Sometimes the level of awareness is even stronger and the listener is not only conscious of emitting a signal but he deliberately chooses a specific one to provoke a particular effect on the speaker, for example: the listener decides to stare at the speaker to show disbelief or surprise expecting a confirmation by the speaker. In this work we are interested in creating a system able to generate listener's signals that are emitted at the first level of awareness. We do not take into account the agent's intention to consciously provide a backchannel signal; backchannels are emitted with a low level of intentionality. As we will see in Chapters 4 and 7, the triggering of a backchannel signal depends solely on the speaker's acoustic and non-verbal behavior.

2.4 Types of backchannel signals

2.4.1 Reactive and response

Verbal and non-verbal information sent by the speaker is firstly perceived by the listener and then processed through a more aware evaluation [Kopp et al., 2008]. Backchannel signals can derive from both stages of the reception of the speaker's message. At the first stage, a non-conscious appraisal of the perceived information can determine an automatic type of behavior, that, for example, shows lack of contact or perception. At a second stage, the more aware evaluation, involving memory, understanding and other cognitive processes, can determine a type of signals that shows understanding and attitudinal reactions, like acceptance or rejection, belief or disbelief and so on.

To refer to these two types of backchannel we adopt the terms proposed in [Kopp et al., 2008]: we call *reactive* the type behavior derived from perception processing and *response* the type of behavior determined by cognitive processing. We consider reactive backchannels as a reaction to the perceived speech. For example, these backchannel signals can show lack of contact and perception, as when the speaker is too far or he speaks too softly. Response backchannels convey information about what the listener "thinks" of the speaker's speech (if it agrees or not, if it understands or not, etc.). Through this type of backchannel listeners provide information on their beliefs, intentions and emotional state related to the conversation. To emit response backchannels a more aware evaluation of the content of the speech is needed.

In this thesis we are interested in implementing an ECA system able to display both types of backchannel. We aim at providing a conversational agent with the ability of showing a varied listener's behavior that comprehends several appropriate backchannel signals. One of the reason is that human behavior is very rich and varied; we think that to create believable virtual agents, such a richness should be reproduced. Moreover, we need this dichotomy between reactive and response backchannel signals for implementation purposes, we want our algorithm able to trigger appropriate backchannel signals even if information about what the agent "thinks" of the speaker's speech is not provided.