International Journal of Humanoid Robotics
Vol. 3, No. 3 (2006) 1–27

World Scientific
www.worldscientific.com

HEAD GESTURES, GAZE AND THE PRINCIPLES OF CONVERSATIONAL STRUCTURE

DIRK HEYLEN

Human Media Interaction,
University of Twente, The Netherlands
heylen@ewi.utwente.nl

Received 4 November 2005

Much of the work on embodied conversational agents is concerned with building computational models of nonverbal behaviors that can generate the right behavior in the appropriate context. In this paper, we discuss, from a linguistic and a conversation theoretic point of view, how nonverbal behaviors in conversations work. We look particularly at gaze and head movements. These play a variety of functions in face-to-face interactions. We show how these functions are structured by general principles governing cooperative actions and symbolic communication.

Keywords: Nonverbal behavior; conversation; gaze; head movements.

1. Introduction

Embodied conversational agents are designed to take part in face-to-face conversations with humans. Properly engaging oneself in a conversation entails that one has internalized how to deal with the protocols and techniques that have evolved in human society and how to turn the result into linguistic action. Much has been written on the various protocols and techniques involved in having a conversation as the subject has been studied by several research traditions, including anthropology, sociology, social psychology, ethology, personality psychology, psychiatry, linguistics, anthropological linguistics, cognitive psychology, philosophy, ethnomethodology, micro-sociology, neuropsychology and psycholinguistics; a list of disciplines mentioned in Duncan and Fiske.[1] Clearly, language is not just the domain of linguists but as language involves social action, it is a matter of concern to scholars in many other disciplines. In this paper, we try to throw some light on the nature and structure of conversational protocols by focusing on two kinds of behaviors: head movements and gaze.

So by what systems of "rules" or "conventions" are face-to-face conversations organized? The interest in conversations shown by the various disciplines is evidence for the many levels on which organizational rules are defined: linguistic conventions related to lexical issues, syntax and semantics; conversational conventions such as

programs or scripts on how to enter and exit conversations or to take turns; task and specific domain conventions, and social conventions that involve knowing what is the appropriate and socially acceptable conversational move to make. Not one of these conventions functions independently from the others. For instance, the conversational conventions that regulate turn-taking also involve social parameters: various aspects of the way people relate to each other, their status, dominance and other factors play a role in whether or not (and how) someone can interrupt a speaker.

An important challenge for the research on embodied conversational agents is how to integrate these ideas, observations, and theories from the various disciplines and how to put them into rules and procedures that embodied agents can use in actual interaction. Embodied Conversational Agents (ECA) research has always been a highly eclectic business. ECA researchers borrow insights from linguistics, cognitive science, AI, cognitive psychology and social psychology. The social perspective in particular has become increasingly important in the work on embodied conversational agents in recent years, witness, for instance, the work on friendship and long-term relations with ECA's,[2] on social rapport,[3] engagement[4] and the incorporation of politeness theory in the design of tutoring agent.[5] A similar trend is visible in the work of our group on socially intelligent agents in which we have moved from implementing an embodied version of a task-oriented spoken dialogue system[6] to the design of socially intelligent agents.[7] This involves a shift in perspective. Increasingly, we have come to view language as social action. Behaviors of agents are not only designed for their specific task (providing information on the task, regulating conversational flow) but the conversation is part of a social encounter. For instance, in building an Intelligent Tutoring System (INES),[7] we made an effort to define dialogue acts using social variables that determine the kind of action that a tutor should make. A tutor has to steer and motivate the student, know when the student welcomes a hint, etc. The emotional states affected by these forms of social interaction typically involve elements and variables such as: social rewards, dependence, status, power, and face. In general one of the goals that people want to come out of social interaction is to enhance the self of each actor.[8] In the INES case, we therefore decided to incorporate the social variables into our choice of speech act primitives. The INES system was an attempt to account for the fact that the behaviors that people display in face-to-face interactions operate on multiple levels simultaneously, involving both task-level and social dimensions.

In this paper, we will show in more detail how even simple behaviors operate on multiple levels and how the levels are connected with each other in a systematic way. We take a look at particular kinds of behavior that people display in face-to-face conversations: head movements and gaze. We consider questions such as the following. How and why do people move their heads? What kinds of protocols and principles govern the gaze behavior of people involved in interaction? We first survey some of the literature that has been devoted to these questions. This will show the

many factors involved. A more systematic view arises when we look at the survey from the perspective of a single framework, a view on language as social action as it has been articulated in the work by Clark[9] amongst others. This perspective provides one way to integrate multiple views on the protocols and techniques that people use in face-to-face interactions. The question "By what systems of 'rules' or 'conventions' are face-to-face conversations organized?" was put forward in the call for papers of the Virtual Social Agents workshop organized during the AISB 2005 meetings in Hertfordshire. Parts of this paper were presented at this workshop.[10] The aim of this paper is not just to point out how behaviors such as gaze and head movements are governed by specific protocols and conventions, but to discuss in more detail the *nature* of the rules and conventions that operate in face-to-face conversation, how the general principles that govern cooperative behavior are realized in conversational acts and how, in fact, face-to-face conversations consist of protocols to generate new conventions. This will show the systematic ways in which gaze and head movements function in conversations.

In Secs. 2 and 3, we look at head movements in face-to-face conversations. We first list the kinds of movements that appear and the many functions that have been ascribed to them. Next we try to present the functions in a more systematic way, relating them to the various levels of conversational structure. In Secs. 4–6, we do more or less the same thing for gaze, but now we also point out some more general principles that govern conversational structure.

2. Head Movements

People involved in face-to-face conversations move their heads in typical ways. Who would disagree that on the whole the pattern of head movements that people display in conversations seems to differ significantly from the patterns found in non-conversational settings; when people are alone, for instance? Although this may appear too obvious to be worth stating, it is not totally insignificant because it clearly suggests that one can assume that the primary determinants of these particular displays have to do with the nature, the purpose and the organization of face-to-face conversations. If one wants to know more about the kinds of movements and movement patterns that occur and about the factors that determine these one should know more about the protocols and principles that govern face-to-face conversation.

The subject of head movements in conversation has been discussed by many researchers from a variety of disciplines. Compared though to the studies on gestures and facial expressions, head movements have received far less attention. We will first consider the way in which the movements have been described and analyzed in order to determine that the properties of the movements can play a function in the face-to-face encounters. Next we consider the various functions that have been ascribed to these movements.

2.1. *The movements*

Although it is not the major objective of this paper to look at the properties of head movements as such, it still seems appropriate to outline the various dimensions along which head movements can be described and which of these characteristics are involved in signifying processes. Birdwhistle[11] devised several coding schemes for various kinds of kinetic behaviors. He distinguished the following head movements: (i) a full nod up and down or down and up, (ii) a half nod either up or down, (iii) a small "bounce" at the end of (i) or (ii), (iv) a full side and back sweep (which may contain a nod or half nod) and (v) a cocked head.

The conversational character RUTH,[12] allows the same general head movements. The head can nod up and down, rotate horizontally left and right and tilt at the neck from side to side. Furthermore, it can bring the whole head forward or backward. Some of the movements can be combined:

D: Nod downward
U: Nod upward
F: Bring the whole head forward
B: Bring the whole head backward
R: Turn to model's right
L: Turn to model's left
J: Tilt whole head counterclockwise (around nose)
DR: Nod downward with some rightward movement
UR: Nod upward with some rightward movement
DL: Nod downward with some leftward movement
UL: Nod upward with some leftward movement
TL: Tilt clockwise with downward nodding
TR: Tilt counterclockwise with downward nodding

Each of these behaviors is loosely associated with a generic function (see Sec. 2.2). The question is, though, whether each occurrence of a movement this way is the right unit of analysis. Furthermore, also other features of the movement may be used as signifiers of some content. Iwano *et al.*[13] analyzed the head movements in a natural dialogue and movements during a cooperative problem-solving task. Movements were classified according to whether they were horizontal, vertical or inclined and whether they were large or small. Combinations of Inclination-Vertical and Inclination-Horizontal were also noted to have special significance.

These classifications only distinguish between the changes in head positions but there are also other features of movements that may be significant. Hadar and colleagues,[14] for instance, also looked at different properties of the movement such as velocity and amplitude. Also Smid *et al.*[15] took the speed with which certain movements are executed into account. Head orientations, speed and amplitude of movements are all basic features of the movement that play a role in distinguishing

between different types of movements and they may each contribute in their own way to how a movement is interpreted.

Besides considering the features of head movements that have to be taken into account, another important question is how to segment the stream of movements into meaningful units. Typically, nods and sweeps are movement patterns that are considered to be significant units in this respect. Similar to Birdwhistle, Graf *et al.*[16] found the following typical patterns in their corpus: (i) nods: abrupt swings of the head with a similarly abrupt motion back; (ii) nods with an overshoot at the return, and (iii) abrupt swings of the head without the back motion. When looking at syntagmatic relations, i.e. the way segments combine, other properties may become important. For instance, the cyclicity of head nods and shakes of listeners with respect to their difference in communicative function was considered in the work of DiCarlo *et al.*[12]

The timing with respect to other signals may also bear significance. Several authors (see below) have looked at the relation between head movements and speech but also the relation between head movement and facial expressions are of interest. For instance, the so-called embarrassment smile[17] is typically associated with a downward movement of the head.

Having identified some of the meaning-bearing parameters of head movements, the next question to turn to is what and how do head movements signify.

2.2. *The functions*

Based on the literature on head movements, one can put together quite an extensive list of functions and determinants of head movements during conversations. With the head movements of RUTH, the authors associate a list of generic functions:

D: General indicator of emphasis
U: Indicates a "wider perspective"?
F: Indicates the need for "a closer look"?
B: Emblem of being "taken aback"?
R: Indicates there is more information?
L: Indicates there is more information?
J: Indicates expectation of engagement from partner?
DR: Combines meaning of D and R
UR: Combines meaning of U and R
DL: Combines meaning of D and L
UL: Combines meaning of U and L
TL: Indicates contrast of related topics
TR: Perhaps indicates contrast of related topics

Looking closer at this table, one can see that they relate to aspects of information structure, expressive reactions to messages and interaction management functions.

Summarizing several sources,[18–27] one can find that head movements have many "functions." Head movements are used to: (i) signal yes or no, (ii) signal interest or impatience, (iii) enhance communicative attention, (iv) anticipate an attempt to capture the floor, (v) signal the intention to continue, (vi) express inclusivity and intensification, (vii) control and organize the interaction, (viii) mark the listing or presenting of alternatives, (ix) mark the contrast with the immediately preceding utterances.

Furthermore, synchrony of movements may (x) communicate the degree of understanding, agreement, or support that a listener is experiencing.

Greater activity by the interviewer (e.g. head nodding) (xi) indicates that the interviewer is more interested in, or more emphatic toward, the interviewee, or that he otherwise values the interviewee more.

Head movements serve as (xii) accompaniments of the rhythmic aspects of speech.

Typical head movement patterns can be observed marking (xiii) uncertain statements and (xiv) lexical repairs.

Postural head shifts (xv) mark switches between direct and indirect discourse.

Considering that people move their head to change what they want to look at, one could say that the functions of gaze behavior may indirectly count as the determinants of head movements. Gaze behavior has been observed to play a role in (xvi) indicating addresseehood, (xvii) effecting turn transitions, and (xviii) the display of attentiveness.

A typical gaze pattern occurs (xix) when doing a word search.

Gaze (xx) may reflect the social status.

Looking away (xxi) is used to avoid distraction, to concentrate, (xxii) to indicate one does not want to be interrupted.

One looks to the other in order (xxiii) to get cues about mood and disposition of the other, (xxiv) to establish or maintain social contact.

Gazing away (xxv) may reflect hesitation, embarrassment or shyness.

Furthermore, gaze is used (xxvi) to locate referents in abstract space, and (xxvii) to request listeners to provide backchannels.

What this list shows is that simple behaviors such as head movements can have many functions and are determined by many variables. The actions may have a clear semantic value, may find their use in managing the conversational process, be expressive of the mental state of the speaker or hearer (their mood, emotions, personality, or cognitive processing) and relate to interpersonal goals and attitudes. We will briefly discuss some of the specific couplings of behaviors and functions, grouping some of the functions together in more general classes adapting the classification of movements made by McClave[27] in head movements and speech, conversation management, discourse functions, cognitive processing and prepositional functions. We add one other group of functions those related to gaze. In Sec. 3, we make an attempt to map these functions within a general framework that takes language to be a form of joint, social action.

2.2.1. *Head movements and speech*

The relation between head movements and speech has been investigated in many papers. Early work by Dittman[21] and Kendon,[26] provided many observations, related to the timing of certain head movements of speakers with respect to the speech. One of the observations by Dittman was "that there is a 'significant' but not very close relationship between speech rhythm and body movement. Both hesitations in speech and body movements tend to appear early in phonemic clauses and, in addition, movements tend to follow hesitations wherever they may appear in clauses." Postural shifts of the head were claimed to indicate encoding difficulties in many cases; a phenomenon related to difficulties in cognitive processing. This indicates that the relations between head movements and speech indicate a deeper relation between head movements and the reasons that determine the speech phenomena: cognitive processing in case of hesitation, for instance.

Hadar and colleagues[14,24] have also studied the motoric aspects of head movements during speech. Parallel to the relationship of hand gestures to speech, it appears that the head moves almost constantly during speech whereas it remains mostly motionless during pauses and while listening. They also found a correlation between head movements and loudness of the speech: "rapid head movements were accompanied by primary peaks of loudness." As a large proportion of head movements is synchronized with speech features such as loudness or pitch, they can be seen as prosody markers in the visual domain. We refer to Graf *et al.*[16] for example, for a more detailed investigation of this relation. In this way, the head movements serve similar functions such as to mark prominence, for example, and so they play an important role in information structuring processes.

Bernieri and Rosenthal[19] write: "The astonishing finding in the literature, however, is not that our body is synchronized with our verbal utterances but that our body tends also to coordinate with the verbal utterances of anyone we happen to be listening to at the time." According to Hadar *et al.*[24] approximately one fourth of all head movements by listeners occur synchronously with the speaker's speech. Why do listeners do this? The typical associations that exist between head movements and speech point out several components of interactional organization in which head movements play a role. These are basic production processes (shown by the head movements at hesitation points, for instance), information structure (prominence, rhythm, stress) and synchronization processes between the participants in the conversation. Synchronization is essential to joint activities.

2.2.2. *Conversation management*

As the list of functions of head movements and gaze above shows, head movements seem to play an important role in managing the interaction, i.e. in turn-taking and backchanneling processes. McClave notes that "the 'speech-preparatory' repositioning of the head before the start of talk can simultaneously signal the assumption of a turn or the intention to continue and as such is a part of conversational

management." In the same vein, Hadar *et al.*[14] determined that postural shifts co-occurred most significantly between sentences or clauses that were associated with assuming or yielding a turn.

Goodwin[23] shows that many backchannels by hearers are responses to speakers' nonverbal requests for feedback in the form of up-and-down nods. Listeners recognize and respond to these requests in a fraction of a second.

2.2.3. *Discourse functions*

Kendon[26] notes that the particular patterns of movement vary according to the discourse function of the utterance. For example, in his corpus the speaker's head position during a parenthetical remark contrasted with that during statements that "move the substance of the discourse forward" (p. 193). Kendon points out a recurrent pattern for most locutions made by the subject who's behavior he is studying. "At the beginning of each of X's locutions, the head is held either erect and central, or it is held erect and cocked somewhat to the right. As the locution ends, the head is tilted forward or lowered and, in several cases, it is either turned or cocked to the left." The exceptions to this pattern, Kendon argues, have to do with a different discourse function of the locutions. "Of the exceptions, locutions 14 and 16 are parenthetical insertions, locution 4 represents a locution begun again as a correction for locution 3. In this case, it ends with a lowered head. Locution 1 is a 'temporizer' or 'floor acceptance' signal."

Related to such markers of discourse function, McClave groups several functions of head movements as "narrative." The first function is that of marking switches from indirect to direct discourse, marked with a new orientation of the head. The second function concerns the expression of mental images of characters. An example from her corpus was someone moving her head downward iconically when quoting someone talking to someone smaller. These functions mark the status or function of a discourse fragment. The third function McClave categorizes as "narrative" is deictic and concerns the referential use of space. She also notes a typical kinetic pattern when items in a list or alternatives are presented. "Characteristically, the head moves with each succeeding item — often to a contrasting position."

So, head movements indicate differences in the status, function and organization of discourse segments. In this sense, they work as a kind of discourse marker.

2.2.4. *Cognitive processing*

When a speaker utters a word or words and immediately rejects this as inappropriate and repairs, the repair is typically preceded or accompanied by head movements (most common: lateral shakes, often small lateral tremors). This relation is reflected in the correlation between certain head movements and speech production (hesitations) as we indicated above. The "thinking face," described in Goodwin and Goodwin,[28] which involves a turn away from the addressee and a distant look in the face, is a stereotypical expression to signal thinking.

2.2.5. *Propositional functions*

Some head movements have a symbolic meaning. Nods are used to signal affirmation and head shakes signal negation in many cultures. McClave (o.c.) points out that head movements can also express other semantic concepts such as intensification and inclusivity. Intensification is conveyed by head shakes and lateral movements co-occurring with words such as "very," "a lot," etc. These are considered by Goodwin and Goodwin (o.c.) as prototypical assessment markers. Inclusivity is expressed by a lateral sweep co-occurring with concepts of inclusivity with words such as "everyone" or "anything." Uncertainty, marked verbally by phrases such as "I guess," "I think," etc, are kinesically marked by "lateral shakes whose trajectories may be quite contained."

The propositional function of most head movements has mainly to do with expressing a particular propositional attitude relating to certainty and credibility. Instead of being a marker of cognitive processing, these movements are signals expressing an attitude towards the contents of the cognitions.

2.2.6. *Gaze*

Head movements may result from an attempt to gaze towards or away from an interlocutor or an attempt to obtain gaze. In this way, the various factors that determine gaze behavior may also be responsible for changes in head-orientation. Gaze has various functions in social interaction. Argyle and Cook[5] give an extensive description of the functions of gaze (see Sec. 4). When one compares this list of functions and determinants of head movements to those that have been assigned with gaze patterns, one can easily show some overlap, or see that the same class of functions is addressed. Some of the functions[a] mentioned in this and other sources are the following.

(i) *Conversation management.*
   (a) Speakers look to obtain immediate feedback on the reactions of listeners.
   (b) Listeners look to supplement auditory information by visual cues.

(ii) *Gaze and speech.* Shifts of gaze are systematically coordinated with the timing of speech, and help with synchronizing.

(iii) *Gaze and cognitive processing.* Speakers tend to look away to avoid distraction — particularly at the planning face of an utterance.

Gaze is also involved in signaling interpersonal attitudes (people look more at those they like, people high in dominance look more in competitive situations, people high in affiliative needs look more in a cooperative situation, negative attitudes may be signaled by looking away). Gaze is said to be a cue for intimacy. This interpersonal dimension seems less significant for head movements as such.

[a]Some of these patterns have been used in implementations of embodied conversational agents and robots.[29,30]

This survey of functions and determinants of head movements (still incomplete) shows the variety of factors that are involved. Head movements convey propositional information, they play a role in managing the interaction, are tightly connected with the prosody of speech and they express interpersonal attitudes as well. How can we integrate all these elements into a view on interaction and the design of embodied conversational agents? For this, we have to integrate a linguistic perspective that deals with the syntax and semantics of utterances as well as the organization of conversations and with a social and psychological perspective.

3. Communication as Social Action

The number of functions of head movements is bewildering at first sight. One way to get a better sense of the function of these behaviors is to consider from a more abstract point of view the nature of conversation and language use by making explicit the underlying modules and principles that govern these actions. Particularly, when we take the view that language is a form of social, interpersonal action, one can come to a deeper understanding of the many aspects involved in such simple behaviors as head movements. In this section we will summarize — because of its eclectic nature — Clark's study of language use[9] and hook this up with the various functions of head movements listed above. The major premises of Clark's view that are important to us are the following:

- Language fundamentally is used for social purposes. We engage in interaction to do something together or to have the other do something for us. This creates social obligations.
- Language use is a species of joint action and successful execution of joint actions requires certain cooperative procedures.
- Language use always involves speaker's meaning and addressee's understanding. Language is communication: information exchange through natural and non-natural signs.
- People need closure on all their actions. Given that conversational action is a form of joint action in which contributions of participants are directed at others and synchronized with others, this means that people try to ground what they do together or that the signals of closure of an action means are provided by the interlocutor.
- Many actions, and particularly linguistic actions, come in hierarchies (people do things by doing other things).
- Grounding should occur at all levels of communication.

To explain the head movement behaviors and their functions, we rely on a couple of key concepts from Clark's perspective on language. These are the notion of conversation as joint action on different levels, the notion of tracks in conversation and the various ways in which meanings arise.

3.1. *Joint actions*

An important thing to keep in mind when considering behaviors of participants in conversation, is that they are participatory actions that are part of a joint activity carried out by the participants together. In order for such an action to succeed the participatory actions must be coordinated. "What makes an action a joint one, ultimately, is the coordination of individual actions by two or more people. There is coordination of both *content*, what the participants intend to do, and *process*, the physical and mental systems they recruit in carrying out those intentions." (Ref. 9).

One of the ways in which actions are to be coordinated is in their temporal dimension. This involves the synchronization of actions. This, in turn, requires that each participant closely monitors the actions of the other and that participants provide feedback of understanding. It is clear from the above that head movements play an important part in signaling such aspects of the joint activity on various levels: from signaling addresseehood and attention, to interest and even agreement. They are central to the *grounding* process that establishes that actions of the others have been identified, recognized and understood.

3.2. *Action ladders*

The diversity in determinants of head movements is not surprising given that a lot of things happen in conversations at the same time. People do things by doing other things. Coordination works at all these levels simultaneously. Clark distinguishes four levels. A communicative act consists of a person A performing some physical action that counts as a signal for something else. Because communicative actions are "joint actions" they are mirrored by actions of participant B.

> Joint[A executes behavior t for B to perceive — B attends perceptually to behavior t from A]
> Joint[A presents signal s to B — B identifies signal s from A]
> Joint[A signals to B that p — B recognizes that A means that p]
> Joint[A proposes joint project w to B — B considers A's proposal of w]

Head movements from listeners provide feedback to speakers on all these levels. A listener orients his head to the speaker to obtain more information from the speaker's facial expressions but thereby he is also signaling attention. Without additional signals this can be a sign of understanding and even agreement to the proposal (or joint project) put forward by the speaker. This will be discussed in more detail in the section on gaze below.

Communicative actions are designed to get the audience to do things on the basis of their understanding of what we mean. Illocutionary acts have their origins in social practices. As Argyle points out "each person in an encounter is trying to manipulate the other person, in order to attain his own goals" but on the other hand, we have to take the goals of the other person in mind as well. Holtgraves[31] puts it as follows:

> "Not only is language use an action, it is simultaneously an interpersonal action. By interpersonal action I mean that what we do with language — the actions that we perform (e.g. a request) — have implications for the thoughts and feelings of the involved parties, as well as the relationship that exists between them. Our words are typically addressed to other people, and people are not abstract entities devoid of feelings, goals, thoughts, and values. People's language use — how they perform actions with language — must be sensitive to these concerns. We cannot always say exactly what we mean because we generally do not want to threaten or impose on or criticize our interlocutors."

Language is not just a form of joint action designed for the neutral exchange of information. Even exchanging information involves an attempt to change what the other believes. Giving up what one thinks is right, agreeing with what someone else is saying is not a neutral act from a social psychological point of view. In most conversations there is even more at stake for the interlocutors. People use conversations to argue, to negotiate deals or as a prelude to getting more intimate. It should therefore not come as a surprise that people may be offended when an interlocutor apparently does not pay attention by turning his or her head away.

3.3. *Tracks*

Clark distinguishes two lines of talk in conversations. The primary track is concerned with "official business," i.e. what the conversation is about. The second track concerns talk (or elements of talk) in the background: talk about the communication itself. Moreover, Clark, remarks that these tracks are orthogonal to the distinction in levels. "The communicative acts in track 2 are used for managing conversation at all four levels of action. When people nod, smile, or say 'uh huh' during another's utterance, they are saying 'I understand you so far,' a signal in track 2 to help achieve closure at level 3." It is immediately obvious from the list of functions of head movements above that many pertain to this second track, for instance those head movements that play a role in floor and turn-taking.

3.4. *Ways of meaning*

Besides looking at the components on which behaviors work, it is also important to look at their "mode" of operation. This will be discussed in more detail in the section on gaze below. As far as head movements are concerned, Hadar *et al.*[24] have considered the question of how the listener's movements signify. Movements that anticipate an action by the listener typically function as cues and signals. On the one hand, they resemble the general movement pattern at the initiation of speech and as such they can anticipate a turn claim precisely in being part of the initiation of speech. But also they often urge the termination of the other's speech. In this sense, they act as a signal for the other: "I (=listener) want you to stop talking." "Yes/No"

movements, on the other hand, are said to operate as symbolic, conventional signals. When one takes a closer look at the various functions of head movements one can also categorise them with respect to the way they mean: whether they are cues, signals, symbols, icons, indices (deictic use).

The various modes of semiosis become apparent when one looks again at the verbs that are used in the list of functions of head movements. A variety of expressions have been used to characterize the nature of the function: *signal, enhance, anticipate, accompany, express, control, communicate, indicate*, as the following selection of functions shows.

- Signal yes or no, interest, impatience, the intention to continue
- Enhance communicative attention
- Anticipate an attempt to capture the floor
- Accompany the rhythmic aspects of speech
- Express inclusivity, intensification, uncertain statements, lexical repairs
- Control and organize interaction
- Communicate the degree of understanding/agreement or support by synchrony of movements
- Indicate interest, empathy

We will have more to say about the semiotic processes in communication with respect to gaze in particular in Sec. 4.

3.5. *Summary*

When one turns to the literature on head movements in conversation, one is at first faced with a bewildering list of functions and determinants of all the kinds of head gestures that people display during conversations. To get a grasp on the protocols that determine how people move their heads in face-to-face interactions, it is useful to take a step back and consider in more depth how conversations work. The basic principles that govern conversation as a joint activity and a form of social action can explain most if not all of the patterns of head gestures one may observe. In the next sections, we will take a closer look at gaze behaviors to get one level deeper in the analysis, uncovering some underlying principles that are at work in conversations and that help to explain why and how these behaviors relate to so many functions.

4. Gaze

In the previous sections, we already mentioned a number of functions associated with gaze behaviors. Gaze has been observed to play a role in indicating addresseehood, the display of attentiveness, effecting turn transitions and in requests for backchanneling. Typical gaze patterns occur when doing a word search or when one shows a thinking face. Gaze behavior may reflect the social status of the participants. Looking away is often used to avoid distraction, to concentrate, or to

indicate one does not want to be interrupted. It has been said that one looks to another person in order to get cues about mood and disposition and to establish or maintain social contact. Gazing away may also reflect hesitation, embarrassment or shyness. Gaze can also be used to locate referents in abstract space. As with head movements, many of the functions are related to interaction regulation on various levels; from turn management to interpersonal relation management. Using gaze people establish and maintain contact. The list also shows that both cognitive and social processes play an important role, for instance in the case of impression management (cf. Fukayama *et al.*[32] for an interesting study of the effects of some simple gaze parameters on impression formation).

As with head movements, we will consider the question how it is possible that the literature on gaze suggests so many functions and show how basic elements of conversational structure can explain the various functions in a more principled way. We will attempt to show why gaze is so special in the role it has in conversations.

As with head movements, one of the first factors determining the multitude of functions has to do with the fact that conversation is a type of joint activity that proceeds on multiple levels. On each of these, gaze plays a role as a coordination device. The power of gaze to play a role in "engagement/disengagement," in showing liking or the role it plays in the social "bonding" process is build on the more primitive functions of lower levels. Although the gaze behaviors of participants in a conversation have been studied in quite some detail, it is interesting to take a closer look at the patterns that have been described and to relate those observations to primary processes in conversation explaining in part the relations between the various functions of gaze by the principles that underlie it. In the previous sections, we have already identified several functional components and levels on which head movements operate. Gaze operates on most of these as well.

Gaze behavior has been said to reflect the social status of the participants or to indicate interpersonal attitudes such as liking. How is it possible that such a simple behavior can carry these kinds of functions? In the following subsections, we point out how general semiotic and conversational processes together with some principles of the multi-level organization of conversations which we touched upon in the previous sections can help us to explain this to a considerable extent.

4.1. *Paying attention*

The first, basic determinant of turning one's head in the direction of another participant has to do with optimizing *perception*. When listening, turning the head to the speaker allows one to hear and see the speaker better. When speaking, turning one's head to the listener allows one to be better heard and seen. All of the various auditory and visual signals can be perceived better, those related to what is being said as well as those that relate to cues about mood and disposition of the other participant if interlocutors look at each other.

Listeners are not merely passively absorbing signals but are sending out signals themselves as well, mainly through what is called backchanneling[33] which are used to ground the actions. Therefore, the same reasons that motivate speakers to look towards listeners — to be perceived better — motivates the listeners to look towards speakers, in addition to the reason mentioned before. And *vice versa*, the same reason that motivates listeners to look towards speakers — to perceive the signals better — motivates the speakers to look at the listeners to perceive the backchannels better. So, the primary reasons for looking at each other could be said to be as follows:

(i) L looks at S to see S better.
(ii) S looks at L so L can see S better.
(iii) L looks at S so S can see L better.
(iv) S looks at L to see L better.

According to Kendon, an individual's perceptual activity within interaction functions in two different but interrelated ways: as a means of monitoring and as a means of regulation and expression. The reasons for looking at each other that we have just identified are motivated by the need for monitoring. The expressive function either derives naturally from these by common semiotic processes (which will be detailed in the next subsection). This process will explain, for instance, how a person's gaze to a certain object can be explained as a deictic signal, directing the attention to the person he looks at, rather than a simple shift in his own focus of attention.

4.2. *Signaling attention*

In the first place, gaze functions as a means of monitoring: a way to perceive visual input. But one kind of visual input that participants in a conversation can perceive is the gaze of the other itself. The fact that $A$is looking at B combined with $B's$ knowledge that people look to perceive and attend, B can interpret the gaze of A as a **symptom** that A is looking at B because A wants to attend to what B is saying. Given that A knows that B will read $A's$ behavior in this way, A can use gaze, in turn, intentionally, not only to merely attend to B but also to **signal** his attention to B. Part of this comes about naturally because gaze is reciprocal: I can see with my eyes that you are looking at me with your eyes.[b] So from being a symptom of an act of perception, gaze becomes a signal of attention and interest. Semiotically speaking, the behavior gaze first becomes an index for paying attention, as the two are causally connected. This is also part of the reason why gaze can function as a deictic signal. If A is talking to B about something and than averts the gaze

[b]Note that gaze is special in this way. I cannot hear with my ears that you are hearing me with your ears. Touch on the other hand is highly reciprocal as well. This may explain why they are both closely tied to feelings and expressions of intimacy.

towards some object x, then this can be read not just as an index but also as a signal. Further processing involved will be explained using Castelfranchi's[34] analysis in terms of behavioral intentional communication. The mechanism of the semiotic shift allows Goodwin to analyze the function of gaze as follows.

> "It will be argued that one way in which a nonspeaking party can indicate whether he is acting as a hearer is by gazing at the speaker. Hearership can of course be demonstrated in other ways [...] A speaker can use gaze to indicate that the party being gazed at is an addressee of his utterance."

Important in this quote is the use of the words "indicate" and "demonstrate." Gazing can be seen as functioning, more or less, in the first stage as a symptom or natural sign[c] in Gricean terminology. This is a case of a more general strategy: "Signals are deliberate actions. Some are performed as parts of conventional languages [...], but any deliberate action can be a signal in the right circumstances." (Ref. 9, p. 32).

The ability to make such shifts lies at the foundation of communication in general.[34] The effect that one is showing or signaling that one is attending by paying attention through actions that can be perceived by others is built on what Castelfranchi calls *behavioral implicit communication* (BIC).

> "Usual, practical, even non-social behaviors can contextually be used as messages for communicating. Behavior can be communication without any modification or any additional signal or mark. We will call this form of communication without specialized symbols: Behavioral Implicit Communication (BIC). "Behavioral" because it is just simple non-codified behavior. "Implicit" because — not being specialized and codified — its communicative character is unmarked, undisclosed, not manifest, and thus deniable. [...] A lot of social control and collaboration monitoring and coordination, are in fact based on this form of communication and not on special and explicit messages (communication protocols)."

Castelfranchi lists the steps "in the evolution from mere practical behavior to BIC and to a conventional sign." The first step is just behavior (i). It is merely an action without any signification or communication. In our case: merely looking to observe. The second step Castelfranchi calls "signification" (ii) when an agent (x) is just acting as in (i) but another agent (y) observes this and ascribes some "meaning" to this behavior (x is looking at me — y). At this point the behavior becomes a symptom (in Clark's/Grice's terminology) for y. At this stage, there are

[c] "Language use depends on both natural signs and signals. Take natural signs. The sounds I hear mean that the radio is on. The shape of the object my friend is holding means that it is a book. The pitch of a caller's voice means that he is a man. A speaker's involuntary hesitation in uttering a word means that he probably had difficulty thinking of, choosing, or pronouncing it in time. Most things have a natural meaning, and these can be important for language use because they are all natural signs that this or that is true." (Clark, o.c.).

two possibilities: either x involved in behavior (i) is unaware of y observing him and attributing meaning to his actions or he does know that y is monitoring him. In that case the attribution of meaning to the actions of x may be a "known but unintended effect" of this behavior. This is then a case of "weak BIC." If, however, x intends his behavior to be observed by y, this counts as a case of "strong BIC." Note however, that x is not trying to "mean" anything beyond what his practical behavior involves. "*With a BIC message x intends that y recognizes x's action, and perhaps that y recognizes and understands the practical intention motivating the action.*" Applied to the case where a listener is looking at the speaker this stage is where the listener is looking with the intention that y notices that x is looking at y.[d]

Such shifts from pure self-motivated action to symptoms or cues and from symptoms to signals occur again and again[e]. In the case of gaze another function that arises in this way is the "request for backchannelling" that Goodwin writes about: as a listener sees the speaker look at him, he may look for a reason of this gaze behavior and try to find out why the speaker is monitoring him to conclude that the reason is a need for feedback. Looking at a person in general is often interpreted as a request for action. As conversation involves joint action, this has to be regulated in all kinds of ways. Participants have to show each other that they are engaged in the communicative process.

4.3. *Regulation*

As we mentioned before, Kendon notes that perceptual activity within interaction functions as a means of monitoring, regulation and expression. The first function we discussed in Sec. 4.1, the third in Sec. 4.2. The regulation function is very important as well and as Goodwin points out gaze plays an important role in ensuring regulation of the joint action.

> "These functions account in some measure for the positioning of gaze within interaction. Thus, the places where a speaker gazes at his recipient — utterance endings and phrase boundaries within the utterance — are choice points, places where the future action of the speaker is contingent on the subsequent action of his hearer. By looking at his recipient at these points, the speaker can both monitor the recipient's response and signal that a response is desired."

[d]It is important to stress that Castelfranchi does not limit BIC to what are typical, communicative, linguistic behaviors but takes any kind of behavior into account.

[e]Another, related semiotic process involves copying and imitation (pretending). The deictic case relies on the principle of copying "pay attention to what I am attending to" (to establish joint and coordinated attention). But the principle is ubiquitous, for instance, as a low voice is associated physically with large individuals it can be used to convey the impression of a large signaller. As largeness involves power, lowering a voice can be used to signal power (dominance, anger, etc.).

The patterns of gaze towards interlocutors and away have shown a strong connection with several coordination management functions or dialogue control functions: turn-taking and information structure,[f] for instance. Mutual orientation of speakers and listeners is important at the initiation of conversations. Peters,[39] who is concerned with building computational models of agents that may or may not enter into conversations, describes a model "where agents are provided with basic attributes encoding their social relations with other agents as well as their goals to engage in conversation. Agents cannot access other agents' conversational goals directly and therefore they do not know if the other agent wants to engage in conversation with them. Rather, agents are endowed with synthetic senses and perception, and must formulate their own theory on whether the other agent wants to converse, based on their perceived *level of interest* in conversing. Level of interest is determined primarily through gaze and direction of intention."

People do not just enter into conversations without a reason, merely to be able to pay attention to each other. On the one hand, to guarantee the success of joint action the actions should be well coordinated, i.e. synchronized and grounded. But getting involved in joint action with someone else, also involves performing actions for the benefit of the other, or having the other perform actions for your benefit; devoting all kinds of resources: time, attention, energy and taking up or refusing the various projects that the other proposes: requests to change your mind, to answer questions, or what have you.

4.4. *Intimacy*

In "Some functions of the face in a kissing round" Adam Kendon[40] analyzes in detail how a young woman on a park bench regulates the actions of her amorous partner solely by her head and facial movements. It shows the power of head movements, amongst other signals, to control interpersonal attraction. Factors such as dominance, embarrassment, the appropriate equilibrium of intimacy and various emotional characteristics have been said to play a part in gaze behavior.[23] How is this possible?

If two people look at each other, it does not automatically mean that they want to initiate a conversation. It certainly does not automatically mean that they are engaged in flirting or a kissing round. Gaze does not automatically mean that people are busy creating a bond. People look somewhere because they want to see something. But if two people look at each other, this will lead them both to think "why does the other look at me" which may lead to the further question "what does the other person want from me?" This can be a good thing or a bad thing. Such an appraisal of the situation can be based on interpreting other signals, for instance by different kinds of smiles.[g] The screen shots from Pelachaud's Greta in Fig. 1 were

[f]Information structure is part of a general concern with audience design as it deals with structuring an utterance taking into account the state of knowledge of the other participant.
[g]There are a lot more things to say about the relation between gaze and smiles.[1,17]

Fig. 1. Screenshots from Greta made by students investigating different kinds of simles.

made by students investigating different kinds of smiles. It is interesting to see in this context, how differences do not merely relate to the mouth or the orbicularis oculi muscles but also at the direction of gaze. In the embarrassment smile, the gaze is directed downwards. In the flirtatious smile, the head is turned sideways. In the embarrassment case, one wants to avoid the gaze of the other. One does not want to be noticed. Flirting is an interesting case because of the ambivalence: attraction combined with rejection. On the one hand it is an invitation as one gazes towards the other (from the corners of the eyes); on the other it is a rejection (head turned away).

Conversation is a form of social action. People initiate conversations because they want to, they do not just happen. People may want to enter into the conversation for various reasons, one being just to talk, but in all cases they have the desire that the other conversational partner does something as well, listening being one of the actions. In most cases, there is other business to conduct as well. A conversation as a whole can be seen as a complete project that the person initiating the conversation proposes to the conversational partner. They have all to do with having the other do something. Social, interpersonal variables such as affiliation and dominance thus play an important role. Social obligation management is an intrinsic part of conversation, witness the standard apologies that one makes when starting a conversation: "Sorry to bother you," "Excuse me, can I ask you a question?" Argyle puts it as follows:

> "During the encounter itself, A is concerned with eliciting certain responses from B, or with establishing and maintaining some relationship with B. In order to do this, A needs continuous information about B's reactions to his own behavior, so that the can modify it if necessary.

> *A* may simply want *B* to like him, or he may have other quite personal motivations with regard to *B*, or *A* may want *B* to learn, buy, vote, or respond in terms of mainly professional goals which *A* has. In either case *A* needs to know what progress he is making with *B*. He may be concerned with *B*'s attitudes towards himself, with *B*'s emotional state, with *B*'s degree of understanding, or with other aspects of *B*'s response."

According to the social skill model of conversation that is elaborated in Argyle (o.c.), "each person in an encounter is trying to manipulate the other person, in order to attain his own goals." As one is aware of being the object of intentions, perceptions and attitudes of the others present, an important goal of interaction is self-presentation. Argyle remarks: "However, in order to take account of concern with the other's point of view, this use of an imaginative cognitive model of the other, some addition seems necessary to the social skill model itself." Self-presentation means presentation of the self for someone else and requires the capacity of a person to take another's perspective.

Interpreting gaze behaviors, in general, as signal of attention (or interest etcetera) assumes the cognitive ability to understand others as intentional agents. It is therefore not surprising that gaze plays an important role in the "reading the mind in the eyes" test devised by Baron-Cohen and others.[36] This test[h] is intended to show how well someone can put himself into the mind of another person and tune in to their mental state. This ability is "the main way in which we can make sense of or predict another person's behavior."[37] It is clear that it plays an important role in the semiotic processes described in the previous subsections (witness the attention to these concepts in the papers by Castelfranchi[34] and Peters[39]). In general, gaze serves as an invitation and acceptance to engage in joint actions.

5. Gaze and the Organization of Conversation

In this section, we will look at theories of conversational processes, touched upon in the previous section, to present the various aspects of conversation that were used in the account of the variety of functions of gaze behavior in the previous section.

Multiple levels. Austin already distinguished between actions on different levels such as the locutionary act and the illocutionary act. This refers to the kind of constitutive relations that abound in conversational activity: people do something by doing something else (illocution: the act performed "in" locuting), by turning their heads and eyes in a certain ways lovers may invite each other for another kissing round. As we saw before, in his analysis of language use, Clark[9] proposes to distinguish four essential levels that relate to how signaling and communication work. At the first level, a certain behavior is executed with the goal that it is perceived by someone else. This behavior is intended as a signal that has to be

[h]In the test a participant is presented with a set of 25 pictures of the eye region of faces of actors and actresses and is asked which of two words describes best what the person is thinking or feeling.

identified as such by the other (level 2) and recognized as meaning what the signaler wants it to mean (level 3).[i] The actions are intended to bring about some kind of change. At level 4, Clark[9] talks about proposing and considering joint projects:

> "[B]y performing an illocutionary act [...] I am proposing, suggesting, posing, or putting forward a project for us to carry out jointly — namely, that I get you to sit down. Now, getting you to sit down is another thing I can't do by myself. It is a joint action that I am projecting for the two of us to do, and that requires us to coordinate our actions. I will call this joint action a *joint project*. Joint projects have two parts. In my terminology, the speaker *proposes* a joint project, and the addressees *take* it *up*."

Another important feature of conversational activity that was stressed in *Using Language* and that is apparent from the description of the levels is that conversation involves a joint activity, i.e. an activity that people undertake together. Their participatory actions only make sense in relation to each other. For instance, the act of speaking is designed with the complementary act of listening in mind. Jointness occurs on all levels as we have shown above.

The fact that communicative actions are joint actions has important consequences. The participatory actions of the individual participants need to be coordinated properly in order for the conversation to go smoothly. People have to coordinate on (i) the kind of action and (ii) the timing of the action. Also,

> "One reason joint activities are complicated is two or more people must come to mutually believe that they are participating in the same joint activity."

Signaling "addresseehood" and "speakerhood" is part of signaling that one is involved in a joint activity and what role one is taking. By gazing one not only attends to the other but one also shows that one is attending.

Many of the actions that people undertake are designed for coordination implicitly. They have a conventional form that allows participants to coordinate with each other. But there may also be explicit cues that directly address the coordination explicitly. Utterances and other signals are operative on two tracks: the "official business track" and the "communication itself track." Gaze, being an essential part of conducting the interaction or communication, becomes (by means of the familiar semiotic processes) an ideal signal for coordination as part of the communicative track, such as turn management. We noted before that gaze can do this easily because of its double role of functioning on several levels. Recall the discussion about behavioral intentional communication. In that case, the execution of the behavior and the fact that it is being observed by itself constitutes a significant process, without being mediated by a prior conventionalized signal.

[i]Note how this involves similar steps as those presented from moving to simple gaze behaviors to the interpretation of attention *signals*. In Section 4.2.

> Joint[A gazes at B for B to perceive — B gazes at the gaze from A]
> Joint[A shows that he intends to gaze at B — B identifies the gaze of A as a signal that A wants to attend to B]

Composition. Another aspect of the logic of actions in conversations is that most joint actions are composed out of a sequences of smaller actions:[7]

> "Most joint activities get realized as sequences of smaller actions, many of which are themselves joint actions. Playing the quartet divides into sections, or phases, each of which divides into subsections and subphases, and so on. What emerges is a hierarchy of joint actions. [. . .]
> Synchrony of action requires coordination on the entry and exit times to each phase. To achieve synchrony, the participants must be able to project both times from what went before. They should be helped whenever the times are: (1) good reference points – jointly salient moments in time; and (2) easy to project from the previous phases. The participants achieve continuous synchrony."

It is therefore not surprising that timing of gaze relates to important phases in conversations:

> "The looks of the speaker toward the hearer occur at the ends of phrases. At points of hesitation, the speaker looks away from his recipient, gazing back at him when fluent speech is resumed."

Closure. Another important aspect of coordination and jointness of action is that participants have to find out whether their actions have succeeded. Donald Norman refers to this as the need for closure "agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it" (cited in Ref. 9, p. 222). Because of the nature of conversation as a form of joint action, where actions by participants are intended to effect changes in other participants, the evidence needs to be provided by the others, which requires such actions as backchanneling. Actions may only succeed if the other participant performs a complementary action. If the others do not perform this complementary action this may be taken as a cue that the initial action did not succeed. This motivates the need for gaze in the backchannel.

5.1. *Summary*

It is not surprising that gaze behavior carries so many functions. The nature of conversations as a *joint activity* in which different people have to *attend* to each other and *coordinate* their actions performs the basis for the importance of gaze. People engaged in a conversation have to look at each other to *monitor* each other's actions because they have to be coordinated, demanding synchronization. By a common *semiotic process* these simple behaviors become communicative in their

own right. Jointness of action operates on *multiple levels*. Each action is directed at the other, from the presentation of signals to the social and interpersonal actions that are carried out through the more basic levels and on each of these the principle of closure operates The importance of gaze percolates from the basic level of monitoring to these higher levels as well. An important factor is the ability of participants to interpret the behaviors of the others in terms of the intentions that lie behind them.

6. Implications for ECA Research

Traditional spoken dialogue systems abstract away from many processes found in natural, face-to-face conversations. They are turn-based, task-oriented and make use of limited input and output modalities. The work on Embodied Conversational Agents has been trying to move away from these limitations in several ways. First, by extending the communicational signaling to other modalities taking into account facial expressions, gestures and also posture and gaze, for instance. Secondly, by not only taking a task-centered approach to conversation but by also paying attention to emotion, personality and the social context of interaction as is witnessed by the work on rapport, engagement, long-term relations, politeness, and social intelligence.[3–5] A third theme that has received some attention as well is the move away from turn-based systems towards continuous interaction. These concerns all relate to modeling more and more of the intricacies of natural conversations.

Gaze is one of the behaviors that have been investigated quite extensively in mediated communication and embodied conversational agents research. The algorithms that have been implemented are mainly of two kinds. The gaze behavior may be captured, totally or in part, either by statistical models or it may be captured by simple rule-based systems. References 32 and 42 for instance, are good examples of statistical and probabilistic models, whereas Poggi *et al.*[43] present an analysis of several common gaze behaviors in terms of BDI-like constructs. The approach taken in this paper is to propose a kind of dictionary of gaze behaviors with definitions presented as formulas of this BDI-framework. In Ref. 29 and our own work,[30] we implemented a simple gaze rule associated with turn-taking behavior. Gaze behavior plays an important role in human–human conversation affecting many dimensions of interactive behavior. From our research on gaze in an earlier study, we came to the conclusion that it is not possible to separate the various functions and determinants of gaze. Focusing on the gaze behaviors typically associated with turn-taking behaviors, we did not only find differences in conversational smoothness, but also in the way the virtual character was perceived and appreciated. We concluded that fiddling with one or more parameters of the gaze behavior designed to optimize a particular function will automatically have repercussions for other functions. This lead us to the question of how exactly the functions relate to one another.

Most implementations of gaze models in the ECA community have relied on the conversational and psychological literature to define the rules that govern the gaze

behavior of the ECA, often combined with data obtained for the particular setting. The potential drawback of these data-sources is that the rules and patterns that are found in one study are specific to the situation that has been investigated. For instance, the fact that one study might say that in dyadic conversations listeners look more at speakers than *vice versa* may be reversed in a situation where somebody is giving directions to some other person using a map. In the latter kind of situation, a listener will attend to the map during most of the time, and the speaker will look at the listener to check whether he is paying attention. Similarly, we may suspect people from telling a lie because they do not dare to look into our eyes. However, when people are really trying to convince others that they are telling the truth, they may look at the other on purpose to avoid the appearance of lying. Reversals in the patterns of gaze may thus be due to the specifics of the setting and task, or to the effect of being consciously in control of the gaze behavior or not. There are several other cases were a particular opposite setting in a variable results in opposite behaviors. To give just one example, it has been observed that people high in dominance look more in competitive situations whereas people high in affiliative needs look more in a cooperative situation.

Although data driven and simple rule-based approaches have proven to be quite effective, in the future we hope to be able to build more principle-based models of conversational interaction that can generate and interpret meaningful gaze behaviors. The work by Peters (o.c.) can be seen as an example of a principle-based approach. The analysis presented in this paper partly serves as a preliminary investigation for this endeavor.

A more principled account should start from the basic functions that gaze serves and combine this with the underlying mechanisms governing conversation to deduce the various concrete behaviors. Gaze is the prime means for basic functions in language: monitoring. The basic form of communication starts when some action is observed by someone else. The duality and mirroring involved in language — actors being both senders and receivers at the same time — is matched nicely with the reciprocity of observation allowed by visual perception. Because of this, the recursion this gives rise to (I see that you see that I see that you see that I see you...) together with the in-born tendency to attribute intentions to human actions makes us read a lot of things in each other's eyes. What is important about the availability of such processes in conversation is that not all "protocols" and "conventions" are given beforehand but that they arise and get constituted within the interaction. Whereas Castelfranchi points out the general principle behind the shifts for all kinds of behaviors, we would like to add the special status that gaze plays in its double role. On the one hand gaze fulfills the role of "observation" and on the other it can constitute the behavior that is being observed. Given this reflexive status and the fact that it reciprocal, gaze plays an important role in regulating interaction, in establishing synchrony and creating new protocols.

One of the important implications of this view on the functional organization of behaviors in conversations is that by devising algorithms for the generation of

communicative behaviors one will have to take into account the complete picture, i.e. all the levels of action involved in communication, the complex but systematic ways in which they are related, and the principles behind this organization.

Acknowledgments

This work arose out of considering head movements and gaze behavior in the context of the Humaine Network of Excellence.

References

1. S. Duncan and D. W. Fiske, *Face-to-Face Interaction* (Lawrence Erlbaum Associates, 1977).
2. T. Bickmore, *Relational Agents: Effecting Change through Human-Computer Relationships* (MIT 2003).
3. T. Bickmore and J. Cassell, Social dialogue with embodied conversational agent in eds. J. van Kuppevelt and L. Dybkjaer and N. Bernsen, *Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems* (Kluwer Academic, 2005).
4. C. Sidner, C. Lee, C. Kidd and N. Lesh, Explorations in engagement for humans and robots, in *IEEE RAS/RSJ Int. Conf.on Humanoid Robots* (2004).
5. W. L. Johnson, P. Rizzo, W. Bosma, S. Kole, M. Ghijsen and H. van Welbergen, Generating Socially Appropriate Tutorial Dialog, *Affective Dialogue Systems* (2004), pp. 254–264.
6. A. Nijholt and D. Heylen, Multimodal communication in inhabited virtual environments, *Int. J. Speech Technology* **5**, 343–353 (2002).
7. D. Heylen, M. Vissers, H. J. op den Akker and A. Nijholt, Affective feedback in a tutoring system for procedural tasks, eds. E. André, L. Dybkjaer, W. Minker and P. Heisterkamp, *ISCA Workshop on Affective Dialogue Systems* (Springer-Verlag, Berlin, 2004), pp. 244–253.
8. M. Argyle, *Social Interaction* (Atherton Press, New York, 1969).
9. H. Clark, *Using Language* (Cambridge University Press, 1996).
10. D. Heylen, *Challenges ahead. in. Proc. AISB Symp. on Social Virtual Agents* (2005).
11. R. Birdwhistle, *Kinesics and Context — Essays on Body Motion Communication*, University of Pennsylvania Press, Philadephia (1970).
12. D. DiCarlo, M. Stone, C. Revilla and J. J. Venditti, Specifying and animating facial signals for discourse in embodied conversational agents, *Computer Animation Virtual Worlds* **15**(1), 27–38 (2004).
13. Y. S. Iwano, Kageyama, E. Morikawa, S. Nakazato and K. Shirai, Analysis of head movements and its role in spoken dialogue, *ICSLP 96* (1996).
14. U. Hadar, T. J. Steiner, E. C. Grant, F. Clifford Rose, Kinematics of head movements accompanying speech during conversation, *Human Movement Science* **2**, 35–46 (1983).
15. K. Smid, I. S. Pandzic and V. Radman, Autonomous speaker agents, in *Proc. CASA 2004*, MIRALab, University of Geneva (2004) pp. 259–266.
16. H. P. Graf, E. Cosatto, V. Strom and F. J. Hang, Visual prosody: Facial movements accompanying speech, in *IEEE Conf. Automatic Face and Gesture Recognition* (2002) pp. 381–386.
17. P. Ekman, *Telling Lies* (Norton, New York, 2001).
18. M. Argyle and M. Cook, *Gaze and Mutual Gaze* (Cambridge University Press, 1976).
19. M. Argyle, R. Ingham, F. Alkema and M. McCallin, The different functions of gaze, *Semiotica* (**7**), 32–49 (1972).

20. F. B. Bernieri and R. Rosenthal, Interpersonal coordination: Behavior matching and interactional synchrony in, eds. R. S. Feldman and B. Rimé, *Fundamentals of nonverbal behavior* (Cambridge University Press, 1991), Chap. 11, pp. 401–431.
21. A. T. Dittmann, The body movement-speech rhythm relationship as a cue to speech encoding, *Studies in Dyadic Communication*, eds. A. Siegman and B. Pope, (Pergamon, New York, 1972), pp. 135–151.
22. N. Freedman, The analysis of movement behavior during the clinical interview, in eds. A. Siegman and B. Pope *Studies in Dyadic Communication* (Pergamon, New York, 1972) pp. 153–175.
23. C. Goodwin, *Conversational Organization* (Academic Press, NY, 1981).
24. U. Hadar, T. J. Steiner and F. Clifford Rose, Head movement during listening turns in conversation, *J. Nonverbal Behavior*, **9**(4), 214–228 (1985).
25. J. Heritage, Current developments in conversation analysis, in eds. D. Roger and P. Bull *Conversation: An Interdisciplinary Perspective* (Multilingual Matters Ltd, 1989) pp. 21–47.
26. A. Kendon, Some relationships between body motion and speech: an analysis of an example in A. Siegman and B. Pope, *Studies in Dyadic Communication* (Pergamon, New York, 1972) pp. 177–210.
27. E. Z. McClave, Linguistic functions of head movements in the context of speech, *J. Pragmatics* **32**, 855–878 (2000).
28. M. H. Goodwin and C. Goodwin, Gesture and coparticipation in the activity of searching for a word, *Semiotica* **62**, 51–72 (1986).
29. J. Cassell, O. Torres and S. Prevost, turn taking vs. discourse structure, *Machine Conversations* 143–154 (1999).
30. D. Heylen, I. van Es, B. van Dijk and A. Nijholt, Experimenting with the gaze of a conversational agent, in eds. J. van Kuppevelt, L. Dybkjaer and N.O. Bernsen *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems* (Kluwer Academic Publishers, 2005).
31. T. M. Holtgraves, *Language as Social Action* (Lawrence Erlbaum Associates, Mahwah, London New Jersey (2002)
32. A. Fukayama, T. Ohno, N. Mukawa, M. Sawaki and N. Hagita, Messages Embedded in Gaze of Interface Agents. Impression management with agent's gaze, *CHI 2002* (2002), pp. 41–48.
33. V. H. Yngve, On getting a word in edgewise, *Papers from the Sixth Regional Meeting*, Chicago Linguistics Society (1997), pp. 567–578.
34. C. Castelfranchi, ToM and BIC: Intentional behavioral communication as based on the theory of mind, in *Proc. AISB Symp. on Social Virtual Agents* (2005).
35. J. L. Austin, *How to do Things with Words* (Clarendon Press, Oxford, 1962).
36. S. Baron-Cohen, Jolliffe, Mortimore and Robertson, Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger Syndrome, *J. Child Psychology Psychiatry* **38**, 813–822, (1997).
37. S. Baron-Cohen, The 'reading the mind in the eyes' test revised version: A study with normal adults, and adults with Asperger Syndrome or High-Functioning Autism, *J. Child Psychiatry Psychiatry* **42**, 241–252 (2001).
38. J. J. Ohala, Ethological theory and the expression of emotion in the voice, in Proc. 4th *Int. Conf. on Spoken Language Processing*, Philadelphia (1996).
39. C. Peters, Foundations of an agent theory of mind model for conversation initiation in virtual environments, in *Proc. AISB Symp. on Social Virtual Agents* (2005).
40. A. Kendon, Some functions of the face in a kissing round, *Semiotica* **15**(4), (1976).

41. J. Cohn, L. I. Reed, T. Moriyama, J. Xiao, K. Schmidt and Z. Ambadar, Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles, in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG'04)*, (2004) pp. 129–138.
42. S. P. Lee, J. B. Badler and N. I. Badler, *Eyes alive, ACM Trans. Graphics* **21**(3), 637–644 (2002).
43. I. Poggi, C. Pelachaud, and F. de Rosis, Eye communication in a conversational 3D synthetic agent, Special issue on behavior planning for life-like characters and avatars, *AI Commun.* **13**(3) 169–181 (2000).

Dirk Heylen studied philology at the University of Antwerp. He received his Ph.D. degree in Computational Linguistics from the University of Utrecht in 2000. Before this, he had been working in several European projects on machine translation and corpus analysis. He is currently Assistant Professor in the Human Media Interaction group of the University of Twente. His main interests are the study of human face-to-face interactions with the aim of understanding the underlying principles that will enable building virtual copies. He takes a broad view on communication, looking at both verbal and nonverbal behavior on the one hand and extending the linguistic perspective with a more global view considering language as social action.