

Social Agents: the first generations

Dirk Heylen, Mariët Theune, Rieks op den Akker, Anton Nijholt
Human Media Interaction
University of Twente

{heylen,theune,infrieks,anijholt}@cs.utwente.nl

Abstract

Embodied Conversational Agents can be viewed as spoken dialogue systems with a graphical representation of a human body. But the embodiment is not the only difference. Whereas Spoken Dialogue Systems are mostly focused on computing the linguistic dimensions of communication, conversational agents are conceived as intelligent agents that have an identity, a persona. Thus, cognitive modeling is often more involved in ECAs including the modeling of emotion. Whereas spoken dialogue systems are focused on the task, virtual humans are also equipped with social skills involved in interaction. This can take various forms. In this paper we review some of the approaches that have been taken in the first decade of ECA research, by presenting the social signaling skills of three agents we have developed in our group.

1. Introduction

In traditional spoken dialogue systems, the kinds of information services such as TRAINS (<http://www.cs.rochester.edu/research/trains/>) from the nineties [1], the focus was on getting a specific task performed by natural language dialogue. The power of a spoken dialogue system is made possible by constraining the domain; which helps semantic processing. Having a clear task, makes it possible to simplify pragmatic processing as well, as the scenario - getting information about a train journey, for instance - is quite well structured, following a simple script. The strategy of such a dialogue system consists in asking a series of questions with restrained options. When the system takes the initiative - starting the conversation with "You are talking to the X-system. You can book tickets to destinations from anywhere in Europe. From which city do you want to leave?" - this constrains the input sufficiently for speech recognition to perform reasonably well. The spoken dialogue system is thus able to fill in the slots that are needed to formulate a query on its database and provide the user with the

information wanted. Besides these information gathering and information providing actions, an important part of the dialogue actions consist in checking whether the system has correctly understood the user - a process referred to in some systems as grounding - and instantiating repair dialogues if this appears not to be the case. A spoken dialogue system is mainly concerned with content and control dimensions of interaction, less with what Goffman has termed the "ritual" dimension of interaction [10].

To give an idea of the dimensions involved in conversations which have also been found relevant for spoken dialogue systems, Figure 1, shows the major dimensions of conversational activity that are distinguished in the DIT++ taxonomy (<http://dit.uvt.nl>).

Although the DIT++ scheme provides a slot for "social obligation management functions", these are mainly related to formulaic elements of interaction. Spoken dialogue systems such as TRAINS have a similarly limited view on the social aspects. For instance, the choice of wordings may reflect a certain element of formality or politeness and the system will greet the user and present itself.

The virtual humans that have been developed since the start of this century (and perhaps slightly before that), on the other hand, have been endowed with more elaborate social skills. In the following sections we will discuss three embodied conversational agents that have been developed by the Human Media Interaction group, illustrating some of the approaches to equip embodied dialogue systems with social skills. In the final section we will discuss the approaches and sketch some future developments.

2. The virtual Receptionist

The virtual agent Karin (Figure 3) was one of the first embodied dialogue systems that we used to carry out some user experiments. Karin is a virtual receptionist who resides behind the information desk at the lobby of the Virtual Music Center (VMC) - a virtual replica of the theater in Enschede. The dialogue skills of Karin were based on an earlier non-embodied dialogue system that interfaced to a database containing the information on the performances in

Information transfer functions:
 information seeking
 information providing functions
Action discussion functions:
 commissives and directives
Dialogue Control functions:
 feedback
Interaction management:
 turn, time and contact management
 own/partner communication management
 discourse structure management
Social Obligations Management functions:
 salutation, self-introduction, apologizing,
 gratitude expression, valediction

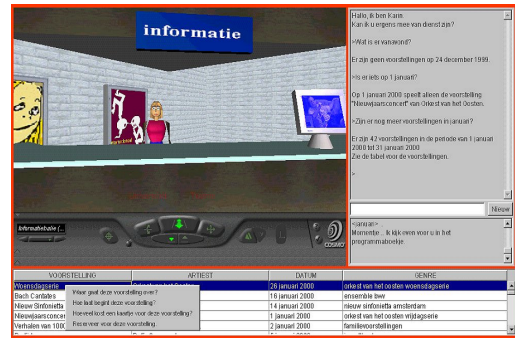


Figure 3. Karin: the virtual receptionist.

Figure 1. Dialogue involves several dimensions of activity. Basic information transfer one of these: the dimension of content. Besides information exchange this also involves getting people to do things or committing oneself to do things. On a meta-level, people take actions that ensure that the dialogue goes well: providing feedback on reception of messages or eliciting feedback, for instance. Taking care that the interaction goes well also means that one undertakes actions that make it clear whose turn it is to speak. Besides the level of content/task and control, there is a third level, which is always involved when two or more people interact, the “ritual”, or social dimension.

[K] Hello, I am Karin. How can I help you?
 [U] Is there anything on this evening.
 [K] There are no performances on the 24th of December 1999.
 [U] What about the first of January?
 [K] On the first of January there is only the New Years Concert of Orkest van het Oosten.
 [U] Are there other performances in January?
 [K] There are 42 performances in the period from January 1, 2000 to January 31, 2000. Please take a look at the table.

Figure 2. Karin: example dialogue

the actual theater. The dialogue system allowed people to query the information about performances and also to order tickets. A short dialogue is presented in Figure 2.

To develop the dialogue system a Wizard of Oz study was conducted in order to get a sense of the kinds of questions people would ask and a sense of how they would ask them. On the basis of this corpus, a parser was developed that uses pattern matching to analyze the user’s input. The Karin agent will, as other spoken dialogue system, ask the user questions that will allow it to fill the slots it needs to query the database ([20]).

The introduction of an embodied version of the dialogue system raised questions about the proper way to have the agent behave with its body. What kinds of actions should

it perform? What kinds of nonverbal behaviour should it display and how should this be related to the verbal expressions? In our main study on Karin’s nonverbal behaviour we focussed on gaze. Where should the agent be looking at during the course of the interaction?

From the literature on gaze behaviours in interaction, we know that it is involved in several dialogue control function and in interaction management. In a basic sense, gaze is closely related to attention. As a listener, looking at the speaker signals some form of attention which clearly fulfills a contact management role. For a speaker, seeing that the listener is looking, fulfills a typical positive feedback-function. At the end of a turn speakers frequently look to the interlocutor, which can function as an indication that the turn is about to end (turn management). Besides these control functions, gaze can also function as a deictic, pointing device.

The gaze behaviour that we implemented in our agent was related to these conversation regulation aspects and deictic functions. While the user was typing, Karin would look towards the user, as a display of attention. When Karin spoke short sentences she would continue looking at the user, but at the beginning of somewhat longer utterances, we had Karin look away; turning her eyes and head upwards and sideways. At a certain point she would resume looking at the user. This is similar to the algorithm used in [8]. We also had her look at the table of performances that appears in the screen as a result of a query to direct the user’s attention to it.

In an experiment we looked at the effectiveness of this behaviour by comparing three versions of the system. Besides the version that implemented the behaviours mentioned above, we had a version in which Karin looked at the user most of the time and one in which she would change her gaze behaviour in a more or less random way. We had 48 people interact with one of the versions of Karin (16 per condition). They were instructed to make two reservations for a performance. It appeared that subjects who interacted with the system that implemented the gaze algo-

algorithm needed significantly less time to complete the task. This would indicate that the gaze behaviour had an important part in interaction management, making the conversation go smoother.

Besides keeping track of the time it took the participants to make the reservations, we also asked them to fill out a questionnaire that consisted of several judgements on a five point Likert scale related to the impression they got from the agent. The factors that we were interested in were *ease of use*, *satisfaction*, *involvement*, *efficiency*, *personality* and the *perceived naturalness of the behaviours*. It is well-known that gaze behaviours also play an important role on the social and affective dimensions of conversations, i.e. gaze plays an important role in social signalling (see [14] for an overview of functions of gaze). It is therefore not surprising that simple differences in the gaze pattern have an effect on the social perception of an agent.

Although we did not find any significant differences between the conditions with respect to judgement of naturalness of eye movements, there were significant differences between the conditions on several of the other factors. The version that implemented the algorithm performed the best on the factors *ease of use* (with judgements on statements such as “It is easy to get the right information”, “It took a lot of trouble to order tickets”,...), *personality* (“I trust Karin”, “Karin is a friendly person”,...), and *satisfaction* (“I liked talking to Karin”, “I like ordering tickets this way”,...).

What this indicates is that the nonverbal *behaviours* that may be taken as having primarily an interaction management function also have an effect on the social-affective dimensions. As Goffman already noted, the system (control) functions and the ritual functions cannot be separated, in the sense that whatever behaviour is performed, this may have effects on each of the dimensions¹.

Discussion One should note that the Karin agent, is basically a plain dialogue system with an embodiment added to it. The agent does not have a dedicated reasoning component that deals with the ritual functions of components. The nonverbal gaze behaviours are more or less hard-coded, so to speak, on top of the task-oriented dialogue system. The dialogue system does not provide special variables or modules for personality or friendliness. However, the experiment shows that varying the basic behaviours of an agent has clear effects on how it is being perceived as a social agent.

In the Karin study, users interacted with a real working version of the dialogue system. It showed how certain behaviours have effects on the conversation and the perception of the agent on the social/affective dimension. Agents

¹The interaction of interaction management and social dimensions is also explored in our current work on the perception of different turn-taking behaviours on the perception of the social skills of an agent [24].

have been used to learn more about the mapping between social signals and their meanings or effects in other types of studies as well. These may take the form of perception studies, in which subjects are asked to rate the behaviour of an agent on dimensions related to social skills by showing a short video clip. The goal of these studies is to establish some kind of dictionary (or gestionary) of social signals and their meanings. In the context of the SEMAINE project, we have carried out several of such studies ([4], [15], [16], for instance). Although, such studies solve part of the puzzle of associating social signals with their possible meanings, they have several shortcomings. The main problem is that they abstract away the context of the interaction. Showing a video of an agent making a particular gesture, head movement or gaze pattern, does not show the context in which this takes place. In a different context the same signal will often have a different effect as well.

3. The Virtual Tutor

The example of Karin shows that it is practically impossible to dissociate the various dimensions of conversation: content, control and social-emotional factors and that signals for interaction control will also work in part as social signals. In the case of a virtual receptionist, the task as such does not involve very complicated social skills, except perhaps for maintaining some level of politeness. In other kinds of interactions for which virtual agents have been employed, social skills are much more important for the task as such. Consider, for instance, the case of a tutor².

A tutor engages in interaction with a student to teach him or her certain knowledge or skills. Typical acts of the tutor include setting specific objectives for the student, motivating the student, giving instructions, setting a specific task, asking or answering questions, explaining, providing support, hinting, pumping for more information, giving examples, providing positive or negative feedback and evaluating the student. A tutor does not just need to provide information on an appropriate level in a way that the students can learn optimally, but also has to perform actions that motivate and challenge students. For this, tutors may need to praise or criticize students. A tutor should therefore not just pay attention to how well a student is understanding instructions but also to how the student is feeling.

Lepper ([19]) identified four main goals in motivating learners: challenge them, give them confidence, raise their curiosity and make them feel in control. The skills of a good tutor does incorporate social skills. The four motivating goals identified by Lepper can be achieved by varying the teaching tactic. Also for a given task, there may be dif-

²In the ECA literature tutors or coaches are popular tasks to study relational aspects of virtual humans ([5], and [12], and [18], are just three early examples), though one of the first important studies on relational aspects involved a Real-Estate Agent ([7]).

ferent strategies that a tutor can use to reach the learning objective. For instance, the tutor can choose the Socratic method which mainly involves asking questions to the student. This can raise the student's curiosity. This method should be chosen only if the student is quite confident and has some mastery over the subject. The kind of praise or negative feedback given can provide confidence. The tutor will choose its actions based on how the student feels.

INES is an intelligent tutoring system that was primarily designed to help students practice nursing tasks using a haptic device within a virtual environment ([17]). We paid special attention to affect control in the tutoring dialogues by selecting the appropriate feedback. Also the kind of teaching action, the affective language used, and the overall teaching tactics are adjusted to the presumed mental state of the student. For this, INES takes into account elements of the student's character, his or her confidence level, and an appraisal of the student's actions: did the student make many mistakes, how harmful are the errors that were made, how was the overall performance so far, how active is the student etc. Also taken into account when calculating these values are the difficulty of the task, for instance. This is used to estimate the affective and motivational state of the student (anxious-confident, dispirited-enthusiastic) as well as the performance on the task.

The tutoring situation is primarily a dialogue, and INES is a combination of an intelligent tutor system and a dialogue manager. The social-affective dimensions affect both the nature of the tutoring and the nature of the dialogue. Affective parameters will affect the style of the feedback. Compare, for instance, "It was quite a difficult task. Try again, but put the needle in more slowly." versus "You put the needle in too fast. Try again." This difference in formulations shows the kinds of verbal adaptations the agent is able to make.

Discussion Compared to the Karin agent, INES has modules built in that keep track of the user's mental state and modules that reason about the appropriate action to take, taking this mental state into account. This is reflected in the behaviours that also involve the execution of the task level. In this case different learning strategies may be chosen and actions that differ with respect to presumed confidence. The socio-affective dimension is not only expressed through the choice of learning strategy, but also in the verbal (and to a limited extent nonverbal) expressions that are chosen by the agent. The dialogue acts merge both affective and task dimensions. INES thus shows a different sort of agent compared to Karin, with the social skills intricately mixed in with the task and expressed through strategy and choice of words.

Another important difference relates to the user modeling. In the case of the virtual receptionist, the agent tries to

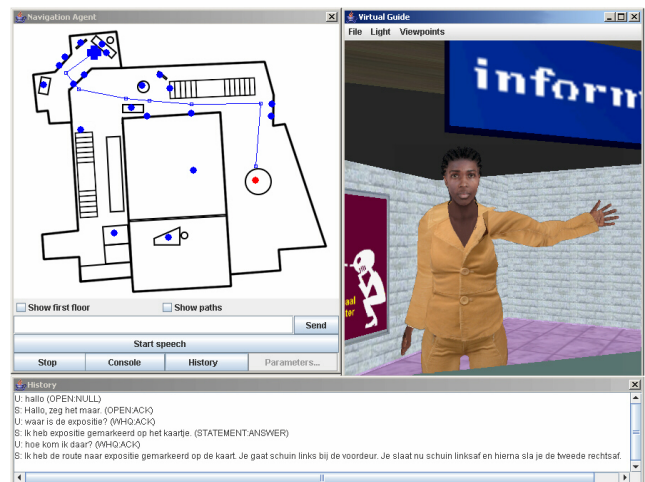


Figure 4. The Virtual Guide

guide the user in providing the information that is needed to make the reservation but is not further concerned with analysing the user's input. INES, on the other hand tries to get a sense of the affective state of the user by interpreting the actions taken and estimating the impact the performance in the exercise might have on the motivational state of the student. Moreover, the INES tutoring agent has an emotional model of its own in which emotional variables such as happy-for or sorry-for are kept track of (for more details see the paper cited).

In the next section we present a third virtual human in which social skills are manifested again in a different way. We return to the Virtual Music Center.

4. The Virtual Guide

The Virtual Guide³ is an embodied conversational agent that also resides in the Virtual Music Center, just as Karin. This agent is able to give directions. Visitors can ask the Guide for information using spoken or typed language as input, in combination with mouse clicks on a map of the environment (see Figure 4). The Virtual Guide responds using spoken language and gestures, and can also show things on the map. In this section we focus on the Guide's verbal behaviour, discussing how the Virtual Guide aligns her level of politeness to that of the user, so as to make her appear more socially intelligent.

Evidence from psycholinguistics has shown that the linguistic representations in social interactions automatically become aligned at many levels [21]. In other words, dialogue partners tend to copy aspects of each other's language. Following Bateman and Paris [3], our notion of alignment includes *affective style*, focusing on the verbal expression of politeness. We have equipped the Virtual

³Online demo at <http://wwwhome.ewi.utwente.nl/~hofs/dialogue/>

Table 1. Some sentence structures that can be handled by the Virtual Guide (translated from Dutch) and their politeness values (P).

| Form | Example sentences | P |
|------|--|----|
| IMP | Show me the hall. | -3 |
| DECL | You have to tell me where the hall is. | -2 |
| | I have to go to the hall. | -1 |
| | I am looking for the hall. | 0 |
| INT | Where is the hall? | 0 |
| | Where can I find the hall? | 1 |
| | Would you show me the hall? | 2 |
| | Do you know where the hall is? | 3 |

Guide with an adaptive politeness model that dynamically determines the user’s level of politeness during the dialogue and lets the Virtual Guide adapt the politeness of her utterances accordingly: a politely worded request for information will result in a polite answer, while a rudely phrased question will result in a less polite reaction.

Like most previous work, we build on Brown and Levinson’s politeness theory [6], which is based on the idea that speakers are polite in order to save the hearer’s face: a public self-image that every person wants to pursue. The concept of face is divided in *positive face*, the social need for a person to be approved of by others, and *negative face*, the need for autonomy from others. Whenever a speech act goes against either of these needs, this is called a Face Threatening Act (FTA). Brown and Levinson discuss various linguistic strategies to express an FTA at different levels of politeness. The *off-record strategy* is an indirect way of phrasing an FTA so that it allows for a non-face threatening interpretation. For instance, when someone says “This weather always makes me thirsty” this is probably a hint that he would like a drink. However, for the hearer it is easy to ignore the indirect request and treat the utterance only as an informing act instead.

A dialogue with the Virtual Guide is always initiated by the user, whose first utterance is then immediately analysed to determine its level of politeness. To this end, we associated the grammar used to parse user utterances with tags indicating their level of politeness on a scale from -5 (least polite) to 5 (most polite). The politeness level depends both on sentence structure, as illustrated in Table 1, and on the use of *modal particles* such as “perhaps” or “possibly”, as in “Could you perhaps show me the hall?”⁴ A detailed account of how user politeness is computed can be found in [9]. The system also determines whether the user chooses formal (*u*) or informal pronouns (*je*) to address the Virtual Guide. In its replies, the Guide will use the same choice of pronouns.

⁴Note that the language spoken by the Virtual Guide is Dutch, and the English translations provided in this paper may differ slightly in politeness from their Dutch counterparts.

After having analysed the user’s utterance, the Virtual Guide determines the affective style of its reaction. Its degree of alignment to the user can be changed, with the guide adapting its style immediately or only over a series of interchanges.

The first step in output generation is the selection of a sentence template with the desired level of politeness, computed from the politeness of the preceding user utterance and modified by the value of α . Currently the Guide has 21 different politeness tactics at its disposal, including those from Table 1; for a full overview see [9]. The tactics are grouped in clusters of sentence templates with an associated politeness range (e.g, from 4 to 5). During generation, the Virtual Guide randomly selects a template from the appropriate range. This way, a fitting template is guaranteed to be found, and some output variation is achieved even when politeness stays at the same level during the dialogue. Finally, gaps in the templates are filled in with formal or informal second person pronouns depending on the user’s pronoun choice.

We evaluated the politeness model using both interactive experiments and quantitative evaluations where human judges had to rate the politeness level of the verbal strategies of the Virtual Guide. The main quantitative results are that indirect tactics (e.g., “Someone should try again”) were generally rated as much less polite than predicted. Also, a frequent comment made by our judges was that subjects found more polite phrasings such as “If you don’t mind” out of place in the context of a request to look at the map. They said “Why would I mind?”, indicating the absence of any threat to autonomy. See [9] for more details.

In a first interactive experiment, we let 4 naive participants (students from our department, 2 male and 2 female) carry out three dialogues with the Virtual Guide. In dialogue 1, the Guide showed no alignment ($\alpha = 1$), and in dialogues 2 and 3 the Guide was set at full alignment ($\alpha = 0$). For dialogue 2 we asked the participants to be polite to the Guide, and for dialogue 3 we asked them to be impolite. They were free to determine the content of the dialogues (while staying within the direction giving domain).

The participants reported that they clearly noticed the effect of alignment in dialogues 2 and 3. Most of them said they liked the Guide’s linguistic style adaptation in the polite dialogue 2, but they found it less appropriate in the impolite dialogue 3, due to the nature of the application: it is the Guide’s ‘job’ to provide a service to the user, and the participants felt that in this role the Guide should always be polite, even to impolite ‘customers’. Though the users found an impolite guide somewhat inappropriate, they still thought it was ‘fun’ to see how the Guide adapted its language to theirs, resulting in exchanges such as:

U: How do I get from here to the exposition, pal?

S: I didn't understand what you said, mate.

The participants also commented on specific politeness tactics used by the Guide. For example, they thought that system utterances such as "It looks like I have been able to indicate the exposition on the map", intended to be polite, made the system sound insecure instead. The users also noted that when the Guide was overly polite this could be interpreted as sarcasm. On the other hand, the Guide also sometimes misinterpreted the user's level of politeness. The most striking example is when one user said "Help!" after the Guide had repeatedly failed to understand him. The system interpreted this utterance as impolite due to the imperative sentence structure, and promptly reacted by also using an imperative: "Say it differently."

Discussion Like the virtual tutor, the guide is able to show its social skills through adapting its verbal utterances. The behaviour is changed based on the behaviour of the user and can thus change dynamically. The examples in the user studies point out again, that it is not always easy to associate specific behaviours with specific functions. For instance, associating imperative sentences with directness or impoliteness. Content and context remain very important.

Politeness is a social skill that has been studied in several conversational agents. Presumably the first attempt at implementing politeness strategies was made by Walker et al. [25], with a recent follow-up in [13]. In their approach, the desired level of politeness of an utterance depends on the social distance between the dialogue participants, the power one has over the other, and the estimated face threat posed by the speech act. Other related work is that of [2, 18, 22] on the generation of tutoring responses, also based on Brown and Levinson's theory. All these systems perform politeness generation based on static input parameters, rather than a dynamic user model that is updated during interaction.⁵ Aspects that are taken into account in other work but not by our model include social distance and the face threat level of system dialogue acts.

5. General Discussion

In the previous sections we have presented three embodied conversational agents that we have been working on over the course of the last decade. They illustrate a range of ways in which agents can become social interactants. Our aim has not been to provide the full range of possibilities that have been explored in the field. By way of summary, we would like to point out some major aspects in the design of social agents.

⁵The politeness model proposed by André et al. [2] includes the user's emotional state, to be measured using physiological sensors. However, it seems this approach to user modelling has never been implemented.

We hope to have made the point clear that conversational agents are not one-dimensional, but are engaged in interactions on different dimensions which we referred to by such names as task and content, control and social-affective. A single behaviour may work on many dimensions in parallel. This is one aspect that makes the mapping between signal/behaviour and meaning/function less straightforward than is sometimes assumed. A better understanding of how signals work together in different conditions is needed but not so easy to achieve. Perception studies tend to decontextualise the signals and offer only limited insight. On the other hand, current video recordings of interactions that are available for analyse are often too particular, or too artificial. More and better methods and data collections will need to be developed and made available.

Behaviours displayed by conversational agents are unavoidably interpreted by the human interlocutor on multiple dimensions so that agents that are designed for simple dialogue will not escape judgements about their social skills, even though there are no components in the agent that are concerned with social interaction processing. Social skills are not only displayed through nonverbal signals, but also to what is being said and how it is said. Besides that, the way a task is performed may show interpersonal attitudes as well.

The examples we presented in this paper concerned social skills such as displaying friendliness, being able to motivate people and give confidence, and being polite. Other social skills that have been explored in the literature are showing rapport, empathy, or engagement, amongst others (see for instance, [11] and [23]).

The examples have shown that there can be considerable variation in the complexity of modeling social skills. In two of the agents that we presented, some sort of sensitivity to the social-affective state of the human interlocutor has been implemented. Social skills seem to require some understanding of the needs, desires, goals and emotional state of the other, by definition. Some of the agents that are around have more intricate user models⁶ than the agents we have presented. However, in general, the affect and social signal reading capabilities of most agents are rather limited. Not a lot of work on affective computing technology has been integrated in the ECA systems. This is one of the areas where next generations of social agents could improve upon. Undoubtedly, the next generations of social agents will become more versatile in their social skills with new projects dedicated to studying social signalling in human(-machine) interaction.

Acknowledgments This work has been supported in part by the European Community's Seventh Framework

⁶See some of the conversational agents developed at ICT (http://ict.usc.edu/projects/virtual_humans).

Programme (FP7/2007-2013) under grant agreement no. 231287 (SSPNet), and in part by the European Community's Seventh Framework programme under agreement no. 231868 (SERA).

References

- [1] J. F. Allen, B. W. Miller, E. K. Ringger, and T. Sikorski. A robust system for natural spoken dialogue. In *Proceedings of the 1996 Annual Meeting of the Association for Computational Linguistics (ACL'96)*, pages 62–70. ACM, 1996.
- [2] E. André, M. Rehm, W. Minker, and D. Buhler. Endowing spoken language dialogue systems with emotional intelligence. In *Affective Dialogue Systems*, LNCS 3068, pages 178–187, 2004.
- [3] J. Bateman and C. Paris. Adaptation to affective factors: architectural impacts for natural language generation and dialogue. In *Proceedings of the Workshop on Adapting the Interaction Style to Affective Factors at the 10th International Conference on User Modeling (UM-05)*, 2005.
- [4] E. Bevacqua, D. Heylen, C. Pelachaud, and M. Tellier. Facial feedback signals for ecas. In *In Proceedings of AISB'07: Artificial and Ambient Intelligence*, Newcastle University, Newcastle upon Tyne, UK, April 2007.
- [5] T. W. Bickmore and R. W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.*, 12(2):293–327, 2005.
- [6] P. Brown and S. C. Levinson. *Politeness - Some universals in language usage*. Cambridge University Press, 1987.
- [7] J. Cassell and T. W. Bickmore. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Model. User-Adapt. Interact.*, 13(1-2):89–132, 2003.
- [8] J. Cassell, O. Torres, and S. Prevost. Turn taking vs. discourse structure: How best to model multimodal conversation. In Y. Wilks, editor, *Machine Conversations*, pages 143–154. Kluwer, The Hague, 1999.
- [9] M. de Jong, M. Theune, and D. Hofs. Politeness and alignment in dialogues with a virtual guide. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 207–214, 2008.
- [10] E. Goffman. Replies and responses. *Language in Society*, 5(3):2257–313, 1976.
- [11] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy. Creating rapport with virtual agents. In *IVA*, pages 125–138, 2007.
- [12] J. Grolleman, E. van Dijk, A. Nijholt, and A. van Emst. Break the habit! designing an e-therapy intervention using a virtual coach in aid of smoking cessation. In W. IJsselstein, Y. de Kort, C. Midden, B. Eggen, and E. van den Hoven, editors, *Proceedings Persuasive 2006. First International Conference on Persuasive Technology for Human Well-being*, volume 3962 of *Lecture Notes in Computer Science*, pages 133–141, Berlin Heidelberg, 2006. Springer Verlag. ISBN=3-540-34291-5, ISSN=0302-9743.
- [13] S. Gupta, M. A. Walker, and D. M. Romano. Generating politeness in task based interaction: An evaluation of the effect of linguistic form and culture. In *Proceedings of the Eleventh European Workshop on Natural Language Generation (ENLG-07)*, pages 57–64, 2007.
- [14] D. Heylen. Head gestures, gaze and the principles of conversational structure. *International journal of Humanoid Robotics*, 3(3):241–267, 2006. ISSN=0219-8436.
- [15] D. Heylen. Multimodal backchannel generation for conversational agents. In *Proceedings of the workshop on Multimodal Output Generation (MOG 2007)*, page 8192, University of Twente, 2007. CTIT Series.
- [16] D. Heylen, E. Bevacqua, M. Tellier, and C. Pelachaud. Searching for prototypical facial feedback signals. In *IVA*, pages 147–153, 2007.
- [17] D. Heylen, A. Nijholt, and R. op den Akker. Affect in tutoring dialogues. *Applied Artificial Intelligence*, 1-2(19), 2005.
- [18] L. Johnson, P. Rizzo, W. Bosma, M. Ghijsen, and H. van Welbergen. Generating socially appropriate tutorial dialog. In *Affective Dialogue Systems*, LNCS 3068, pages 254–264, 2004.
- [19] M. Lepper. Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In *In Computers as Cognitive Tools*, page 75105. Lawrence Erlbaum Associates, 1993.
- [20] A. Nijholt and J. Hulstijn. Multimodal interactions with agents in virtual worlds. In N. Kasabov, editor, *Future Directions for Intelligent Information Systems and Information Science*, volume 45 of *Studies in Fuzziness and Soft Computing*, pages 148–173. Physica-Verlag, Heidelberg, Germany, 2000. ISBN=3-7908-1276-5.
- [21] M. J. Pickering and S. Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226, 2004.
- [22] K. Porayska-Pomsta and C. Mellish. Modelling politeness in natural language generation. In *Proceedings of the Third International Conference on Natural Language Generation (INLG-04)*, LNAI 3123, pages 141–150, 2004.
- [23] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artif. Intell.*, 166(1-2):140–164, 2005.
- [24] M. ter Maat and D. Heylen. Turn management or impression management? In *Proceedings of 9th International Conference on Intelligent Virtual Agents (IVA)*, Amsterdam, The Netherlands, 2009.
- [25] M. Walker, J. Cahn, and S. Whittaker. Linguistic style improvisation for lifelike computer characters. In *Entertainment and AI/A-Life, Papers from the 1996 AAAI Workshop.*, 1996. AAAI Technical Report WS-96-03.