in: J. Cassell, S. Prevost, J. Sullivan, and E. Churchill: Embodied Conversational Agents, The MIT Press, pp. 220-255, 2000.

The Automated Design of Believable Dialogues for Animated Presentation Teams

Elisabeth André, Thomas Rist,
Susanne van Mulken, Martin Klesen, and Stephan Baldes
German Research Center for Artificial Intelligence
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
email: {andre,rist,mulken,klesen,baldes}@dfki.de

1 Introduction

During the last few years, an increasing number of attempts have been made to develop systems that are able to generate presentations automatically. The aim to generate presentation variants on the fly to accommodate for different purposes and user needs is a driving force behind these developments. The availability of multiple media offers the choice between media as well as the possibility to compose media combinations that convey information in a more intelligible and appealing way. The output of such presentation systems comprises mere replications of illustrated texts laid out on a computer screen, but also multimedia documents with timed media, such as video and audio clips, and hypertext-style documents.

Trying to deploy the repertoire of skills of human presenters, some R&D projects have begun to use animated characters (or agents) in presentation tasks. Based on either cartoon drawings, recorded video images of persons, or 3-D body models, presentation agents enrich the repertoire of available presentation styles. For instance, consider the presentation of research results at a conference. Even though a slide show, a video clip or a poster may contain all the relevant information, the presence of a skilled speaker in addition to well prepared presentation material is usually much more appealing. A potential strength of animated characters is their ability to convey nonverbal conversational signals that are difficult to communicate in traditional media.

In this paper, we investigate a new style for presenting information. We introduce the notion of presentation teams which—rather than addressing the user directly—convey information in the style of performances to be observed by the user. The paper is organized as follows. First, we report on our experience with two single animated presentation agents and explain how to evaluate their success. After that, we move to presentation teams and discuss their potential benefits for presentation tasks. In section 2**,** we describe the basic steps of our approach to the automated generation of performances with multiple characters. This approach has been applied to two different

scenarios: sales dialogues and soccer commentary that are presented in section 3 and 4, respectively. Section 5 discusses early impressions gained from informal tests that have been conducted for the two applications. A comparison to related work is given in section 6. Section 7 provides a conclusion and an outlook on future research.

1.1 Animated Presentations with Embodied Conversational Characters

Our work has concentrated on the development of animated presenters that show, explain, and verbally comment on textual and graphical output on a window-based interface. The first project we conducted in this area was the PPP project (Personalized Plan-Based Presenter) that generated multimedia help instructions presented by an animated agent, the so-called PPP persona (André and Rist 1996). The overall behavior of the presentation agent is determined partly by a script and partly by the agent's self behavior. Presentation scripts specify the presentation acts to be carried out as well as their temporal coordination. For example, a script may instruct the character to point to an object in an illustration and explain its function. While a script is an external behavior determinant that is specified outside the character, our characters also have an internal behavior determinant resulting in what we call self behavior. A character's self behavior comprises not only gestures that are necessary to execute the script, but also navigation acts, idle time gestures, and immediate reactions to events occurring in the user interface. Note that the borderline between scripted behavior and self behavior is a matter of the degree of abstraction. The more detailed a script prescribes what a character should do, the less there is a need to equip a character with a rich repertoire of reasonable self behaviors.

In the AiA project (Adaptive Communication Assistant for Effective Infobahn Access), we developed a number of personalized information assistants that facilitate user access to the Web (André, Rist, and Müller 1999) by providing orientation assistance in a dynamically expanding navigation space. These assistants are characterized by their ability to retrieve relevant information, reorganize it, encode it in different media (such as text, graphics, and animation), and present it to the user as a multimedia presentation. The novelty of PPP and AiA are that the presentation scripts for the characters and the hyperlinks between the single presentation parts are not stored in advance but generated automatically from preauthored documents fragments and items stored in a knowledge base.

Reasons to embody the assistants were, among others, (1) that it might ease conveying particular types of information in an unobtrusive way (e.g., gestures; conversational and emotional signals), and (2) that it might have the conjectured *persona effect* (Lester et al. 1997)—that is, the presence of a *persona* might have a positive effect on the user's attitudes and experience of the interaction (for a critical review, see Dehn and van Mulken 1999).

To investigate whether this effect indeed holds if we compare persona conditions with no-persona conditions and to see whether it extends to objective measures rather than just subjective measures, we performed a psychological experiment.

In this experiment, we tested the effect of the presence of our PPP persona with respect to the user's understanding, recall, and attitudes. Twenty-eight subjects were shown Web-based presentations with two different types of content. In the experimental condition, a speaking and gesturing PPP persona made the presentations. In the control condition, the (audiovisual) information presented was exactly the same, except that there was no persona and pointing arrows replaced all gesturing. After the presentations, the subjects were asked comprehension and recall questions and subsequently provided with a questionnaire that measured their attitudes regarding the system and PPP persona. Statistical analyses of the results showed that there was no effect on comprehension or recall. However, analysis of the data on the subjects' attitudes indeed revealed a significant positive effect of persona. Subjects who had seen presentations guided by persona indicated on a questionnaire that they found the presentations themselves and the corresponding tests less difficult than subjects who had seen presentations without persona. In addition, subjects found these presentations significantly more entertaining (van Mulken, André, and Müller 1998).

In a follow-up study, we investigated whether the subjective persona-effect could be found to extend even toward an increased trustworthiness of the information presented by a lifelike character. In this study, subjects had to perform a navigation task. Subjects were in turn assisted in navigation by one of four agents: one was invisible and merely gave textual recommendations as to how to proceed with the task; the second presented these recommendations acoustically; the third was a speaking cartoon-style agent; and the fourth was a speaking agent based on video images of a real person. In the text and audio conditions, reference to a recommended path was accompanied by a highlighting of the corresponding parts of the navigation tree. In the conditions with an agent, such a reference was accompanied by pointing gestures. We hypothesized that the embodied agents would appear more convincing or believable and that the subjects would therefore follow the agents' recommendations more readily.

This hypothesis, however, was not supported by the data. We found numerical differences only in the expected direction: the proportion of recommendations actually followed by the subjects dropped off going from video-based to cartoon-style, audio, and text agents (for further details, see van Mulken, André, and Müller 1999). These findings suggest, among other things, that merely embodying an interface agent may not be enough: to come across as trustworthy, one may need to model the agent more deeply—for instance, by giving it personality. We return to this issue later.

in: J. Cassell, S. Prevost, J. Sullivan, and E. Churchill: Embodied Conversational Agents, The MIT Press, pp. 220-255, 2000.

1.2 From a Single Presenter to Presentation Teams

Often, systems that use presentation agents rely on settings in which the agent addresses the user directly, as if it were a face-to-face conversation between human beings. For example, an agent may serve as a personal guide or assistant in information spaces (as in AiA), it can be a user's personal consultant or tutor, or it may represent a virtual shop assistant who tries to convince an individual customer. Such a setting seems appropriate for a number of applications that draw on a distinguished agent-user relationship. However, other situations exist in which the emulation of direct agent-to-user communication is not necessarily the most effective way to present information. Empirical evidence suggests that, at least in some situations, indirect interaction can have a positive effect on the user's performance. For example, Craig and colleagues found that, in tutoring sessions, users who overheard dialogues between virtual tutors and tutees, subsequently asked significantly more questions and also memorized the information significantly better (Craig et al. 1999, experiment 1).

Along the lines of Alpert, Singley, and Caroll (1999), who use multiple agents to impose a visible and enacted structure on the instructional material presented, we hypothesize that placing such a structure on the presentation may help users to organize the information conveyed. Imposing an organizational structure on the material presented has been shown to facilitate the assimilation of new information with related prior knowledge (Bower et al. 1969). In addition, such organization deepens processing and makes the information easier to remember (Ornstein and Trabasso 1974). The individual personified members of a presentation team could serve as visual indices that might help the user in a sort of cued recall.

With regard to presentation design, a team of presenters enriches the repertoire of possible communication strategies. For example, they allow the conveyance of certain relationships among information units in a more canonical way. Among other things, this benefits decision support systems where the user has to be informed about different and incompatible points of view, pairs of arguments and counterarguments, or alternative conclusions and suggestions. For solving such presentation tasks, it seems natural to structure presentations according to argumentative and rhetorical strategies common in real dialogues with two or more conversational partners. For instance, a debate between two characters representing contrary opinions is an effective means of informing an audience of the pros and cons of an issue.

In addition, embodied presentation teams can serve as rhetorical devices that allow for a reinforcement of beliefs. For example, they enable us to repeat the same piece of information in a less monotonous and perhaps more convincing manner simply by employing different agents to convey it. Furthermore, in an empirical study, Nass and colleagues showed that subjects who watched news and entertainment segments on different TVs rated them higher in quality than news and entertainment segments shown on just one TV

(Nass, Reeves, and Leshner 1996). We suspect that such effects may even be reinforced if information is distributed onto several agents that represent different specialists.

Finally, programs on TV demonstrate how information can be conveyed in an appealing manner by multiple presenters with complementary characters and role castings. This presentation style is used heavily in advertisement clips and infotainment/edutainment that try to combine information presentation with entertainment. In contrast to TV presentations, however, the generation of performances on a computer system allows to take into account particular information needs and preferences of the individual user.

The observations above encouraged us to investigate scenarios in which the user observes (or overhears) a dialogue between several lifelike characters. A necessary requirement for the success of such presentations is that the agents come across as socially believable individuals with their own distinct personalities and emotions (cf. Bates 1994). The manual scripting of such dialogues is, however, not flexible enough to adapt presentations to the specific needs of an audience on the fly. Therefore, our work concentrates on the development of a generative mechanism that allows for the automated design of believable dialogues.

2 Designing Presentation Dialogues: Basic Steps

Given a certain discourse purpose and a set of information units to be presented, we determine an appropriate dialogue type, define roles for the characters involved, recruit concrete characters with personality profiles that go together with the assigned roles, and, finally, work out the details of the individual dialogue turns and have the characters perform them.

2.1 Dialogue Types and Character Roles

The structure of a performance is predetermined by the choice of a certain dialogue type. Various types of dialogues exist including debates, panel discussions, chats, interviews, consultation, sales, brokering, and tutoring dialogues. Which one to adopt depends on the overall presentation goal. In this chapter, we concentrate on scenarios common in TV transmissions: sales dialogues and chats about jointly watched sport events.

Once a certain dialogue type has been chosen, we need to define the roles to be occupied by the characters. Most dialogue types induce certain constraints on the required roles. For instance, in a debate on a certain subject matter, there is at least a proponent and an opponent role to be filled. In a sales scenario, we need at least a seller and a customer.

The next step is the casting of the designated roles. To generate effective performances with believable dialogues, we cannot simply copy an existing character. Rather, characters have to be realized as distinguishable

individuals with their own areas of expertise, interest profiles, personalities, and audiovisual appearance, taking into account their specific task in a given context.

When talking about a character's personality and affective state, we adopt the view of Moffat, who contends that "personality is consistent reactive bias within the fringe of functionality" (1997, 134). Psychologists have attempted to characterize personality by traits, relying on the statistical method of factor analysis to group words, commonly used to describe people, into chief organizing themes. The use of this technique has led to the consensus that five major factors or dimensions account for most of the variation in human personality. Although different researchers use slightly different terms for them, they can be summarized as *open, conscientious, extravert, agreeable,* and *neurotic* (McCrae and John 1992).

Closely related to personality is the concept of emotion. Emotions are often characterized as "valenced reactions to events, agents, or objects, with their particular nature being determined by the way in which the eliciting situation is construed "(Ortony, Clore, and Collins 1988, 13). Moffat differentiates between personality and emotion by using the two dimensions *duration* and *focus*. Whereas personality remains stable over a long period of time, emotions are short-lived. Moreover, while emotions are focused on particular events or objects, factors determining personality are more diffuse and indirect.

A further important component of a character's profile is its audiovisual appearance. Empirical evidence for this is, for instance, provided by Dryer who presented subjects with a set of animated characters to measure their perception of the characters' personality. He found that characters perceived as extravert and agreeable tend to be represented by rounder shapes, bigger faces, and happier expressions while characters perceived as extravert and disagreeable were typically represented through bold colors, big bodies, and erect postures (Dryer 1999).

2.2 Generation of Dialogue Contributions

After a team of presenters has been recruited, a performance is generated. As in our earlier work on presentation planning, we follow a communication-theoretic view and consider the generation of simulated dialogues a plan-based activity. However, to account for presentations given by a character team, a number of extensions have become necessary.

In André and Rist (1996), we argued that a presentation system should clearly distinguish between the creation of material and its presentation. Consequently, we refined the notion of a communicative act by differentiating between *acquisition* and *presentation* acts. While acquisition acts, such as designing a graphical illustration or retrieving it from a database, contribute to the contents of a conversation, presentation acts, such as showing the illustration to an audience, refer to its communicative function. In the

scenarios presented in this chapter, the user is not addressed directly. Instead, information is conveyed implicitly by a dialogue between several characters to be observed by the user. To account for the new communicative situation, we have to extend our previous communication model by introducing dialogue acts, such as *responding to a question* or *making a turn*, which refer to the interaction between the individual agents.

A further level of complexity arises from the fact that information is no longer simply allocated to just one agent, but instead distributed over the members of a presentation team whose activities have to be coordinated. To accomplish this task, we are investigating the following two approaches:

- *Agents with Scripted Behaviors:* In this approach, the system takes the role of a producer that generates a script for the agents that become the actors of a play. The script specifies the dialogue and presentation acts to be carried out as well as their temporal coordination. Since the script writer has almost complete control over all actors, this approach facilitates the generation of coherent dialogues. On the other hand, it requires that all the information to be communicated is a priori known by the script writer. Consequently, it is less suitable in situations where the actors have to immediately respond to events at presentation runtime, such as reactions from the audience. From a technical point of view, this approach may be realized by a central planning component that decomposes a complex presentation goal into elementary dialogue and presentation acts that are allocated to the individual agents. Knowledge concerning the decomposition process is then realized by operators of the planning component.
- *Autonomous Actors:* In this approach, the individual agents have their own communicative goals that they try to achieve. That is, there is no external script for the agents. Rather, both the determination and assignment of dialogue contributions is handled by the agents themselves. To accomplish this task, each agent has a repertoire of dialogue strategies at its disposal. However, since the agents have only limited knowledge concerning what other agents may do or say next, this approach puts much higher demands on the agents' reactive capabilities. Furthermore, it is much more difficult to ensure the coherence of the dialogue. Think of two people giving a talk together without clarifying in advance who is going to explain what. From a technical point of view, this approach may be realized by a distributed system with multiple reactive planners. The agents' dialogue strategies are then realized as operators of the individual planners.

Depending on their role and personality, characters may pursue completely different goals. For instance, a customer in a sales situation usually tries to get information on a certain product in order to make a decision, while the seller aims at presenting this product in a positive light. To generate believable dialogues, we have to ensure that the assigned dialogue contributions do not conflict with the character's goal. Furthermore, characters

may apply very different dialogue strategies to achieve their goals depending on their personality and emotions. For instance, in contrast to an extravert agent, an introvert agent will be less likely to take the initiative in a dialogue and exhibit a more passive behavior. Finally, what an agent is able to say depends on its area of expertise. Both the central and the distributed planning approach allow us to consider the character's profile by treating it as an additional filter during the selection and instantiation of dialogue strategies. For instance, we may define specific dialogue strategies for characters of a certain personality and formulate constraints that restrict their applicability.

Even if the agents have to strictly follow a script as in the script-based approach, there is still room for improvisation at performance time. In particular, a script leaves open how to render the dialogue contributions. Here, we have to consider both the contents and the communicative function of an utterance. For instance, utterances would be rendered differently depending on whether they are statements or warnings. To come across as believable, agents with a different personality should not only differ in their high-level dialogue behaviors but also perform elementary dialogue acts in a character-specific way. According to empirical studies, extravert characters use more direct and powerful phrases than do introvert characters (Furnham 1990), speak louder and faster (Scherer 1979), and use more expansive gestures (Gallaher 1992). Furthermore, the rendering of dialogue acts depends on an agent's emotional state. Effective means of conveying a character's emotions include body gestures, acoustic realization and facial expressions (see Collier 1985 for an overview of studies on emotive expressions).

To consider these factors, the planner(s) enhances the input of the animation module and the speech synthesizer with additional instructions, for instance, in an XML-based markup language.

3 Inhabited Marketplace

As a first example, we address the generation of animated sales dialogues that was inspired by Jameson et al. (1995) and Mehlmann et al. (1998). For the graphical realization of this scenario, we use the Microsoft Agent™ package (Microsoft 1999) that includes a programmable interface to four predefined characters: Genie, Robby, Peedy, and Merlin. Since the use of these characters might lead to wrong implications in car sales scenarios, we are currently designing our own characters whose visual appearance better fits the agents' role in such scenarios.

Figure 1 shows a dialogue between Merlin as a car seller and Genie and Robby as buyers. Genie has uttered some concerns about the high running costs, which Merlin tries to play down. From the point of view of the system, the presentation goal is to provide the observer with facts about a certain car. However, the presentation is not just a mere enumeration of the plain facts about the car. Rather, the facts are presented along with an evaluation under consideration of the observer's interest profile in value dimensions, such as

safety, sportiness, comfort, and economy, that are important to him or her (see section 3.2). Our system is neutral in so far that it presents positive as well as negative facts about the product.

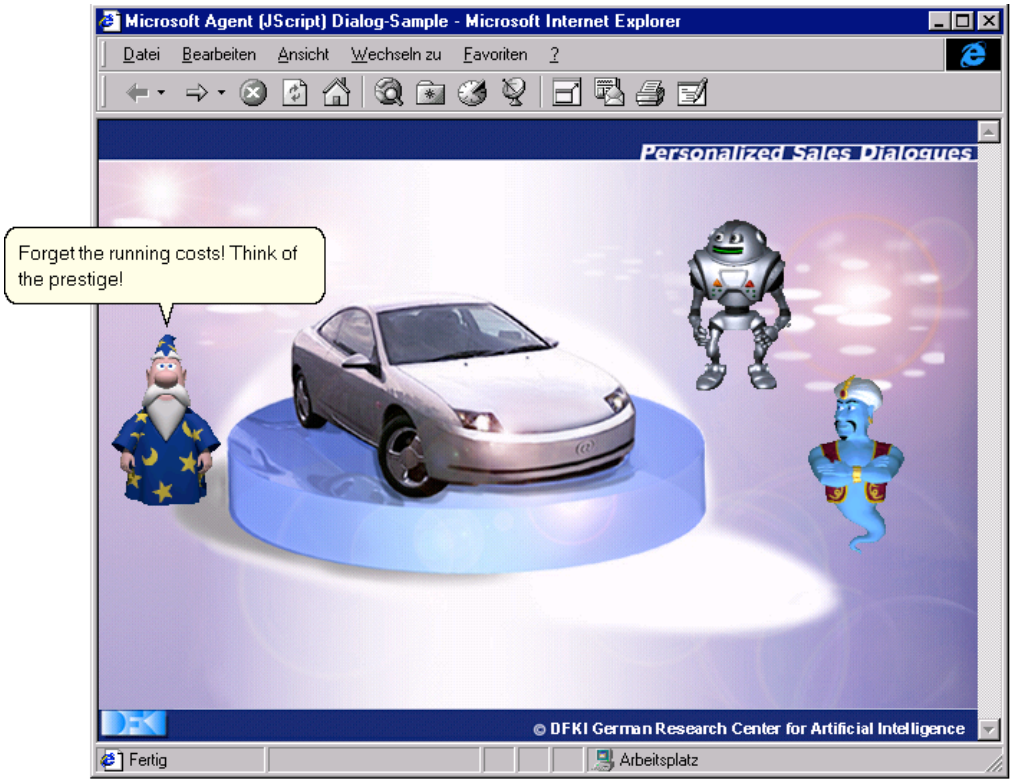

Figure 1 Screen shot of the inhabited market place.

3.1 Character Profiles

To enable experiments with different character settings, the user has the possibility of choosing three out of the four characters and assigning roles to them. For instance, he or she may have Merlin appear in the role of a seller or buyer. Furthermore, he or she may assign to each character certain interests in value dimensions which may reflect his or her own interests (see fig. 2). Our system allows for two operating modes. In the first mode, the system (or a human author) chooses the appropriate character settings for an audience. In the second mode, it allows the audience to test various character settings itself. The potential benefits of this mode will be further discussed in section 5.

SELLER Genie		BUYER1 Peedy		BUYER2 Merlin	
Agreeableness	**Extraversion**	**Agreeableness**	**Extraversion**	**Agreeableness**	**Extraversion**
◉ agreeable	○ extravert	○ agreeable	○ extravert	◉ agreeable	◉ extravert
○ neutral	○ neutral	○ neutral	○ neutral	○ neutral	○ neutral
○ disagreeable	◉ introvert	◉ disagreeable	◉ introvert	○ disagreeable	○ introvert

Figure 2 Role casting interface for the car sales scenario.

Personality dimensions may be set by the user as well. As mentioned in section 2.1, personality traits can be reduced to five big dimensions. Two of them seem to be most important for social interaction (cf. Nass, Isbister, and Lee, this volume). Therefore, we have decided to model these two personality factors

- *Extraversion* with these possible values: extravert, neutral or introvert;
- *Agreeableness* with these possible values: agreeable, neutral or disagreeable.

In this version of the sales scenario, we concentrated on one dimension of emotive response, namely valence (Lang 1995). It has the following values: positive, neutral, and negative. In our scenarios, emotions are essentially driven by the occurrence of events. The events in the sales scenario are the speech acts of the dialogue participants that are evaluated by the characters in terms of their role, personality traits, and individual goals (see sec. 2.1). The goals, in particular, determine the desirability of events; for example, a buyer will be displeased if he is told that a relevant attribute of a car (e.g., electric window lifters) is missing for a dimension that is important to him (e.g., comfort). In this scenario, we do not deal with emotion structures and emotion generating rules explicitly (e.g., see Elliott 1992) but rather connect the scenario-specific dialogue acts (e.g., DiscussValue, PositiveResponse, InformIf) to the relevant animation sequences and utterance templates by using the current internal state of the character as an additional contraint in the behavior selection mechanism, as described in the next section. This approach is similar to that of Lester and colleagues (Lester et al., this volume), where

pedagogical speech acts drive the selection and sequencing of emotive behaviors.

3.2 Source, Structure, and Representation of the Information to be Communicated

Part of the domain knowledge is an ordinary product database, organized in the form of an *n*-dimensional attribute vector per product. In this scenario, the products are cars with attributes, such as model type, maximum speed, horsepower, fuel consumption, price, air conditioning, electric window lifters, and airbag type. Thus, to a large extent, the contents of the database determine what an agent can say about a product. However, the mere enumeration of product attributes would not make an appealing sales presentation, especially if it gets long-winded, as in the case of complex products like cars. Furthermore, products and their attributes described in a technical language run the risk of sounding unfamiliar to the user. It seems much more appropriate to have a further description of the products that reflects the impact of the product attributes on the value dimensions of potential customers.

Such an approach can be modeled in the framework of multi-attribute utility theory (von Winterfeldt and Edwards 1986) and has already been used for the identification of customer profiles in an electronic marketplace for used cars (Mehlmann et al. 1998). In this project, a large German/American car producer and retailer provided the car database, whereas the value dimensions for the product "car" were adopted from a study of the German car market (Spiegel-Verlag 1993) that suggests that safety, economy, comfort, sportiness, prestige, and family and environmental friendliness are the most relevant. In addition, we represent how difficult it is to infer such implications. For instance, high gas consumption has a negative impact on economy, and this relationship is relatively easy to infer.

In this chapter, however, we do not address the question of how to build up and refine profiles based on the dialogue contributions of real customers interacting with the system. Rather, for the sake of simplicity, we assign to the characters a particular interest in one or several value dimensions before the planning of the agent's dialogue contributions starts or ask the user to do so. The user's assignments are interpreted by the system as an indication of his or her own interests. Furthermore, we use a more simplified model representing the implications of attribute values on the agent's value dimensions and for representing how difficult it is to infer them. Figure 3 shows an excerpt of the represented domain knowledge used for the generation of the car sales performances. It lists the value and the implications of the attribute "consumption." For instance, the attribute "consumption" of "car-1" has a negative impact on the dimensions "environment" and "running costs," an implication that is not difficult to infer. The impact on the dimension "prestige" is positive, but this relationship is less obvious.

```
FACT attribute "car-1" "ccar1";
FACT type "ccar1" "consumption";
FACT value "ccar1" 8;
FACT polarity "ccar1" "environment" "neg";
FACT difficulty "ccar1" "environment" "low";
FACT polarity "ccar1" "prestige" "pos";
FACT difficulty "ccar1" "prestige" "medium";
FACT polarity "ccar1" "running costs" "neg";
FACT difficulty "ccar1" "running costs" "low";
```

Figure 3 Excerpt of the domain knowledge: Value and implications of the attribute “consumption” (coded as “ccar1”).

3.3 Design of Product Information Dialogues

To automatically generate product information dialogues, we use a central planning component that decomposes a complex goal into more elementary goals. The result of this process is a dialogue script that represents the dialogue acts to be executed by the individual agents as well as their temporal order. Knowledge concerning the generation of scripts is represented by means of plan operators. An example is listed in figure 4.

```
NAME: "DiscussValue1"
GOAL: PERFORM DiscussValue $attribute;
PRECONDITION:
      FACT polarity $attribute $dimension "neg";
      FACT difficulty $attribute $dimension "low";
      FACT Buyer $buyer;
      FACT Disagreeable $buyer;
      FACT Seller $seller;
BODY:
      PERFORM NegativeResponse $buyer $dimension;
      PERFORM ResponseNegativeResp $seller $attribute
            $dimension;
```

Figure 4 Example of a plan operator for discussing an attribute value.

The operator represents a scenario where two agents discuss a feature of an object. It only applies if the feature has a negative impact on any dimension and if this relationship can be easily inferred. According to the operator, any disagreeable buyer produces a negative comment referring to this dimension (NegativeResponse). The negative comment is followed by a response from the seller (ResponseNegativeResp).

One possible response is listed in figure 5. It only applies if there is an attribute that has a positive impact on the dimension under discussion. In this case, the seller first tells the buyer(s) that it disagrees and then lists attributes with a positive impact on the dimension. Note that our plan operators include both the propositional contents of an utterance and its communicative function. This is in line with Cassell and colleagues, who regard conversational behaviors as fulfilling propositional and interactional conversational functions (Cassell et al., this volume). For instance, we explicitly represent that "Bad for the " $dimension "?" is a response to a negative comment.

```
NAME: "ResponseNegativeResponse2"
GOAL:
PERFORM ResponseNegativeResp $agent $attribute
          $dimension;
PRECONDITION:
     FACT Polarity $attribute $dimension "pos";
BODY:
     PERFORM Respond $agent
          (+ "Bad for the " $dimension "?");
     PERFORM EnumeratePos $agent $dimension;
```

Figure 5 Example of a plan operator for responding to a negative comment.

Our approach allows us to interleave character and dialogue act selection. For instance, in the operator in figure 4, characters are selected by instantiating the variable bindings with characters that satisfy the constraints.

The implementation of the planning approach is based on the Java™-based JAM agent architecture (Huber 1999). The outcome of the planning process is an HTML file that includes control sequences for the Microsoft characters**.** The performances can be played in the Microsoft Internet Explorer. Since the Microsoft Agent™ Programming tool does not allow for intonational markups, we do not convey emotions by acoustic realization in this scenario.

3.4 What About This Car? Two Generation Examples

In what follows, we present two examples that illustrate how the agents' personality and interest profiles influence the contents and the structure of the sales dialogue. For expository reasons, we use extreme parameter settings here, so that differences in behavior are readily distinguishable.

Genie:

Role: seller;
Personality factors: agreeable, extravert;
Interests: environmental issues

Peedy:

Role: customer;
Personality factors: disagreeable, extravert;
Interests: economy

Merlin:

Role: customer;
Personality factors: agreeable, introvert;
Interests: comfort

The conversations are based on just a few dialogue strategies. Essentially, each agent asks about the values of features that might have an impact—positive or negative—on a dimension it is interested in. After that, the value of this attribute is discussed.

Genie:

Hello, I'm Genie. What can I do for you?

;;; starts the conversation because it is
;;; extravert

Peedy:

We're interested in this car.

;;; responds to the question because it is
;;; extravert

Genie:

This is a very environment-friendly car.

;;; provides a summary by emphasizing the
;;; dimension "environment" that is most
;;; important to him. here, the default assumption
;;; is that the buyers have the same interests
;;; as the seller

It has a catalytic converter. It is made of recyclable material.

;;; provides evidence by listing all features
;;; which have a positive impact on the dimension
;;; "environment"

Peedy:

How much gas does it consume?

;;; starts asking a question because it is

;;; extravert, wants to know more about gas
;;; consumption because this feature has an impact
;;; on the dimension "economy," which is important
;;; to him

Genie:

It consumes 8 liters per 100 km.

;;; retrieves the value from the car database

Peedy:

I'm worrying about the running costs.

;;; the value of the attribute "gas consumption"
;;; has a negative impact on "economy"
;;; Peedy is disagreeable and therefore makes a
;;; negative comment

Genie:

Forget the running costs. Think of the prestige.

;;; tries to distract the buyers by mentioning any
;;; positive implication. "prestige" has been
;;; chosen since there is no other easier-to-
;;; infer positive implication

Peedy:

How much does it cost?

;;; requests more information on a second
;;; attribute that has an impact on "economy"

Genie:

It costs 50000 dollars.

;;; retrieves the value of the attribute "cost of
;;; purchase" from the database

Peedy:

That's rather expensive.

;;; negative impact on "economy"

Genie:

Consider what a nice car you get for it.

;;; tries to play down the negative comment

Merlin:

Does it have electric window lifters?

;;; Merlin is interested in the dimension
;;; "comfort" and the attribute "electric window
;;; lifters" has an impact on this dimension

Genie:

No.

;;; retrieves the value of the attribute "electric
;;; window lifters" from the database

Peedy:

Well, a bit more comfort wouldn't hurt.

;;; the value of the attribute "electric window
;;; lifters" has a negative impact on the dimension
;;; "comfort"

Merlin:

Could you please tell me more about its interior?

;;; asks after the interior because this attribute
;;; has an impact on comfort which is important to
;;; him

Genie:

It has a spacious interior.

;;; retrieves the value of the attribute "interior"
;;; from the database

Peedy:

<scratches its head>.

;;; would like to say something negative, but
;;; there is no dimension on which the attribute
;;; "interior" has a negative impact

Merlin:

Good!

;;; the value of the attribute "interior" has a
;;; positive impact on comfort, Merlin is
;;; introvert, but since Peedy doesn't say
;;; anything, Merlin makes a comment
...

The second dialogue is based on the same dialogue strategies and knowledge bases. However, we assign the roles for the seller and the buyers to different agents. Furthermore, we modify the agents' interest profile and personality.

Robby:

Role: seller;
Personality factors: agreeable, extravert;
Interests: sportiness

Peedy:

Role: customer;
Personality factors: disagreeable, introvert;
Interests: environmental issues

Merlin:

Role: customer;
Personality factors: agreeable, extravert;
Interests: safety

Robby:

Hello, I'm Robby. What can I do for you?

;;; starts the conversation because it is
;;; extravert

Merlin:

We're interested in this car.

;;; responds to the question because it is
;;; extravert

Robby:

This is a very sporty car.

;;; provides a summary by emphasizing
;;; "sportiness," which is most important to him

It can drive 100 miles per hour.

;;; provides evidence by listing features which
;;; have a positive impact on "sportiness"

Merlin:

Does it have air bags?

;;; starts asking questions because it is extravert
;;; requests more information on an attribute that
;;; has an impact on safety

Robby:

Sure!

;;; retrieves the value of the attribute "air
;;; bags" from the database

Merlin:

Excellent!

;;; positive impact on the dimension "safety"

Does it have anti-lock brakes?

;;; requests more information on a second
;;; attribute with impact on safety

Robby:

Of course!

;;; retrieves the value of the attribute "anti-
;;; lock brakes" from the database

Merlin:

Wonderful!

;;; positive impact on the dimension "safety"

Peedy:

How much gas does it consume?

;;; wants to know more about gas consumption
;;; because this feature has an impact on
;;; "environment," which is important to him

Robby:

It consumes 8 liters per 100 km.

;;; retrieves the value from the car database

Peedy:

Isn't that bad for the environment?

;;; the value of the attribute "gas consumption"
;;; has a negative impact on "environment,"
;;; less direct speech since it is introvert

Robby:

Bad for the environment?

;;; questions negative impact

It has a catalytic converter. It is made of recyclable material.

;;; provides counterarguments
...

The two dialogues partially discuss the same car attributes, but from different points of view. For instance, in both cases, one of the buyers criticizes the high gas consumption of the car. But in the first case, it is concerned about the high costs, while, in the second case, it is thinking of the environment. According to the applied strategies, the dialogues terminate after all relevant attributes of the car under consideration have been discussed.

4 Gerd and Matze Commenting on RoboCup Soccer Games

The second application for our work on multiple presentation agents is Rocco II, an automated live report system for the simulator league of RoboCup, the Robot World-Cup Soccer. Figure 6 shows a screen shot of the system that was taken during a typical session. In the upper-right window, a previously recorded game is played back while being commented on by two soccer fans: Gerd and Matze sitting on a sofa. Unlike the agents of our sales scenario, Gerd and Matze have been designed specificially for soccer commentary. Furthermore, this application is based on our own Java™-based Persona Engine (André, Rist, and Müller 1999).

4.1 Character Profiles

Apart from being smokers and beer drinkers, Gerd and Matze are characterized by their sympathy for a certain team, their level of extraversion (extravert, neutral, or introvert) and openness (open, neutral, not open). As in the previous application, these values may be interactively changed. Following Ball and Breese (this volume), we decided to focus on the following two emotional dispositions: arousal with the values calm, neutral, and excited and valence with the values positive, neutral, and negative. Both seem useful in characterizing important aspects of a soccer spectator's emotional state.

Emotions are influenced by the current state of the game. For instance, both agents get excited if the ball approaches one of the goals and calm down again in phases of little activity on the field. An agent is pleased if the team it supports performs a successful action and displeased if it fails.

4.2 Source, Structure and Representation of the Information to Be Communicated

Rocco II concentrates on the RoboCup simulator league, which involves software agents only (as opposed to the real robot leagues). Thus, the soccer games being commented on are not observed visually. Rather, the system obtains its basic input from the Soccer Server (Kitano et al. 1997), which delivers player location and orientation (for all players), ball location, game score, and play modes (such as throw-ins and goal kicks). Based on these data, Rocco's incremental event recognition component performs a higher-level analysis of the scene in order to recognize conceptual units at a higher level of abstraction, such as spatial relations or typical motion patterns. Recognized events are internally represented by instantiations of case frames which indicate the event type, the involved objects as well as time and location of the occurrence. The interpretation results of the time-varying scene, together with the original input data, provide the required basic material for Gerd's and Matze's commentary (André, Herzog, and Rist 1997).

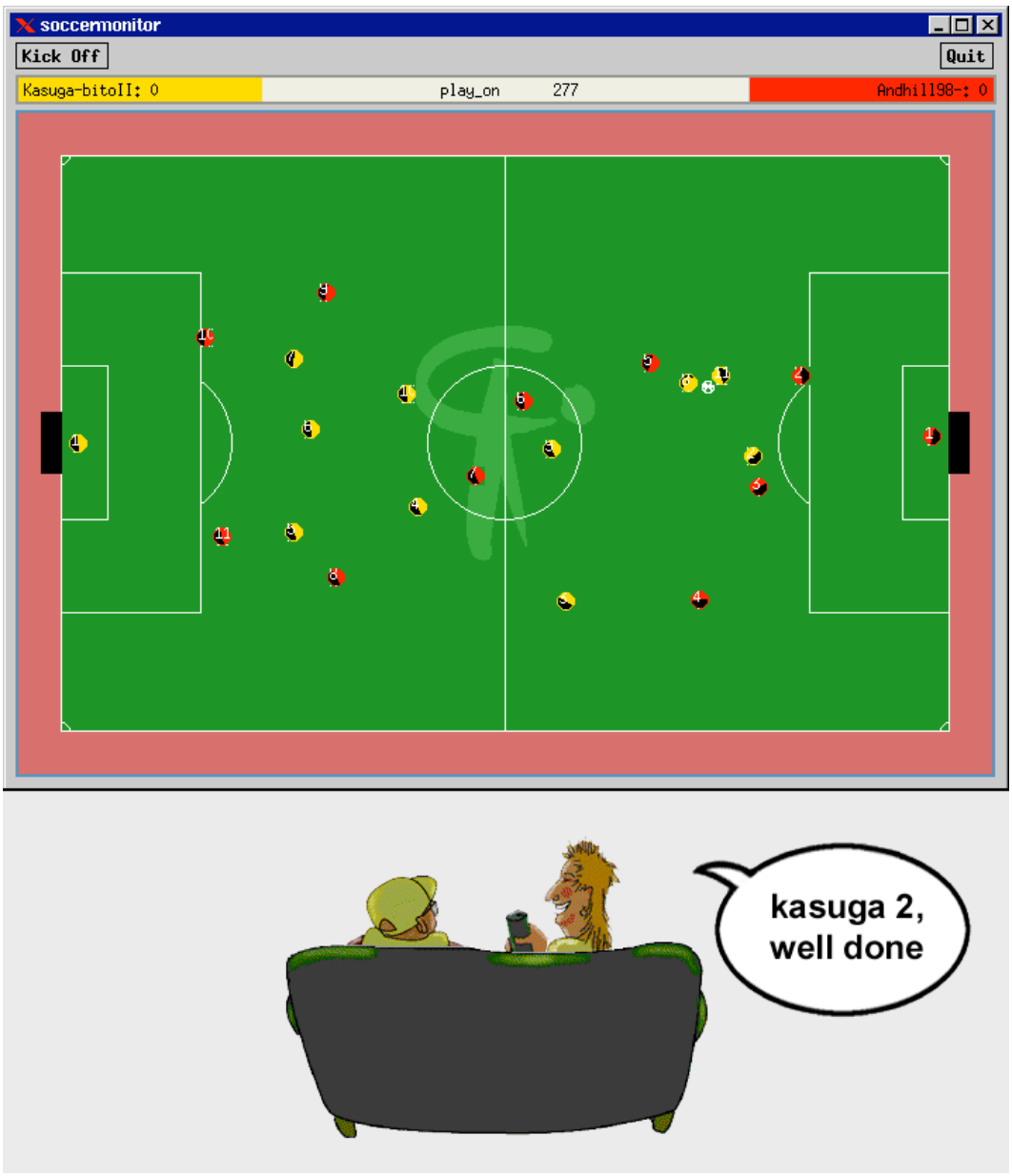

Figure 6 Commentator team Gerd and Matze.

4.3 Generation of Live Reports for Commentator Teams

Since Gerd and Matze comment on a rapidly changing environment, they have to produce speech utterances on the fly. In such situations, no global organization of the presentation is possible. Instead, the commentators have to respond immediately to the incoming data. Furthermore, they have to meet severe time constraints imposed by the flow of the game. They have to decide

whether to utter short telegram-style phrases or provide more detailed descriptions according to the situation. In some cases, it might even be necessary for the commentators to interrupt themselves. For example, if an important event (e.g., a goal kick) occurs, utterances should be interrupted to communicate the new event as soon as possible.

Unlike the agents in the car sales scenario, Gerd and Matze have been realized as (semi-)autonomous agents. That is, each agent is triggered by events occurring in the scene or by dialogue contributions of the other agent. Interactions in this scenario are based on less complex dialogue strategies than the sales scenario. In most cases, the goal of the commentators is to inform the viewer about ongoing events which can be usually realized by a sequence of event descriptions.

As in the previous application, we rely on a template-based natural language generator. That is, language is generated by selecting templates consisting of strings and variables that will be instantiated with natural language references to objects retrieved from the domain knowledge base or delivered by a nominal phrase generator. To obtain a rich repertoire of templates, extensive studies of the soccer language have been necessary. Since most studies concentrate on newspaper reports (Frank 1997), we decided to perform our own corpus analysis and transcribed and annotated 13.5 hours of TV soccer reports in English. Inspired by (Hovy 1987), we associate with each template the following linguistic features:

- *Verbosity:* The verbosity of a template depends on the number of words it contains.
- *Specifity:* The specifity of a template depends on the number of verbalized slots of the case frames and the specifity of the natural language expression chosen for the event type.
- *Force:* We distinguish between forceful, normal, and hesitant language. Forceful language is energetic and contains strong and confident phrases. Hesitant language is characterized by an increased number of linguistic hedges.
- *Formality:* We distinguish between formal, normal, and colloquial language. Templates marked as formal refer to grammatically correct sentences that are more common in newspaper reports. Colloquial templates, such as “ball played toward ?y,” are simple phrases characteristic of informal language. Especially, spoken soccer commentary is characterized by unfinished or ill-formed sentences and nonverbal utterances, such as “Ooh!” or “Aah!” (e.g., see Rosenbaum 1969). Such phrases will be considered in a future version of the system.
- *Floridity:* We distinguish between dry, normal, and flowery language. Flowery language is composed of unusual ad hoc coinages, such as "a lovely ball." Templates marked as flowery may contain metaphors, such as “squeezes the ball through**,”** while templates marked as dry, such as “plays the ball toward ?y” just convey the plain facts.

- *Bias:* Biased templates, such as "well done" contain an evaluation of an action or event. Bias may be positive, negative, or neutral.

Since the generation of live reports has to meet hard time constraints, we decided to use a four-phase filtering mechanism to prune the search space for template selection. Only the best templates of each filtering phase will be considered for the next evaluation step. The first filtering phase tries to accommodate the specific needs of a real-time live report. If time pressure is high, only short templates will pass this filtering phase where more specific templates will be given preference over less specific ones. In the second phase, templates that have been used only recently will be eliminated in order to avoid monotonous repetitions. . The third phase serves to communicate the speaker's attitude. If the speaker is strongly in favor of a certain team, templates with a positive bias will be preferred for describing the activities of this team. The fourth phase finally considers the speakers' personality. As in the sales scenario, forceful language is used primarily for extravert commentators, while hesitant language is used for introvert ones. We are not aware of any empirical studies concerning the use of flowery phrases. However, we assume that such phrases are used primarily by open individuals who are characterized as being creative and imaginative. If several templates remain for selection, verbose and specific templates are preferred for extravert characters, and terse and less specific ones for introvert characters. To increase the fluency of the commentary, the selected templates still allow for various modifications considering contextual constraints. For instance, "Now X" should not be uttered if X is already in focus.

Another important way of conveying personality is acoustic realization. We have not yet addressed this issue but simply designed two voices that may be distinguished easily by the user. Acoustic realization is, however, used for the expression of emotions. Drawing upon Cahn (1990), we have been examining how we can generate affective speech by parameterizing the TrueTalk™ speech synthesizer. Currently, we mainly vary pitch, accent, pitch range, and speed. For instance, arousal is expressed by a higher talking speed and pitch range. Unfortunately, the TrueTalk speech synthesizer only allows a few parameters to be set. Consequently, we can only simulate a small subset of the effects investigated by Cahn.

4.4 Kasuga against Andhill: Commented on by Gerd and Matze

In the car sales example, personality is essentially conveyed by the choice of dialogue acts. Gerd and Matze portray their personality and emotions by linguistic style, which is characterized by the choice of semantic content, syntactic form, and acoustical realization (Walker, Cahn, and Whittaker 1997). In addition, they make use of body gestures, such as jumping up and down and raising their arms. In the current version of Rocco II, each commentator concentrates on the activities of a certain team. That is, there is an implicit agreement between the characters concerning the distribution of dialogue

contributions. Responses to the dialogue contributions of the other commentator are possible, provided that the speed of the game allows for it. Furthermore, the commentators may provide background information on the game and the involved teams. This information is simply retrieved from a database. In what follows, we present a protocol of a system run. We start from these parameter settings:

Gerd:

Attitude: in favor of the Kasuga team;
Personality factors: extravert, open

Matze:

Attitude: neutral;
Personality factors: introvert, not open

Gerd:

Kasuga kicks off

;;; recognized event: kick off

Matze:

Andhill 5

;;; recognized event: ball possession, time
;;; pressure

Gerd:

we are live from an exciting game, team Andhill in red versus Kasuga in yellow

;;; time for background information

Matze:

now Andhill 9

;;; recognized event: ball possession

Gerd:

super interception by yellow 4

;;; recognized event: loss of ball, attitude: pro
;;; Kasuga, forceful language because it is
;;; extravert

still number 4

;;; recognized event: ball possession, number 4 is
;;; topicalized

Matze:

Andhill 9 is arriving

;;; recognized event: approach

Gerd:

ball hacked away by Kasuga 4

;;; recognized event: shot, flowery language

;;; because it is creative

Matze:

red 5, shot towards the goal

;;; recognized event: shot

Gerd:

brilliant save by Kasuga's goal keeper

;;; recognized event: ball possession, attitude:

;;; pro Andhill, forceful language because it

;;; it is extravert

...

5 Early Feedback from Users

So far, we have only conducted some informal tests with our two prototype system demonstrators. Based on these observations, we got a clearer idea about appropriate role castings and the way people may interact with the two systems.

Although we have so far implemented our system with only simple models of personality and emotion, an interesting issue is whether the audience of a performance is able to recognize the character's personality if it has been set by someone else.

We have tried this out in a small informal system test with ten subjects. The results suggested that, generally, if asked to describe the agents in terms of given personality dimensions, subjects are indeed able to recognize the personalities. However, if the assigned personalities mismatched the character's look and voice, they had much more trouble identifying the agent's personality. For example, Merlin's soft voice was judged to conflict with a disagreeable personality.

In addition, although perhaps not being necessary conditions for eliciting personality and interest ascriptions (cf. Nass et al. 1995), subjects' comments suggested that the look and voice of a character are indeed important cues concerning its personality and interest profile. In our test, subjects tended to believe that Merlin was more interested in comfort and safety, while they expected that Robby was more interested in the technical details of a car. This suggests that reusing the look and voice of characters for different roles is only possible to a certain extent.

Finally, we observed that our subjects were very eager to "play around" with the system and try setting the personalities for the different agents to see what effect this had on the way the presentation team conveyed the information. This may be seen as an indication for users wanting to understand the relationship between personality and behavior. This, in turn, encourages and provides support for applications that, for instance, address the development of person perception skills (Huard and Hayes-Roth 1997).

Comments revealing that the use of presentation teams has to follow careful consideration came from some subjects with respect to the soccer scenario.

They remarked that they felt that Gerd and Matze had distracted them from watching the soccer game. A reason for this may be that in this scenario the soccer game was presented in a rather abstract way (e.g., soccer players were represented as circles). Compared to this, the agents themselves and their idle time gestures (which were not only functional as in our previous empirical studies) may have been much more attractive to watch, even though they were only shown from behind. This suggests that we need to take care that the attractiveness of presentation teams per se and that of the information they comment on are appropriately set and also that idle time gestures require more careful selection on our part. For instance, if something unexpected or important happens, then the idle time movements should not be visually distracting.

In future, as soon as the system's implementation status allows, we would like to perform more formal evaluations. It might be interesting to investigate the effects of indirect interaction with presentation teams on variables such as recall and understanding. Moreover, in everyday work situations, it is often important to recall not only what was said but also who said it and when (source monitoring, cf. Schachter, Harbluk, and McLachlan 1984). It may be interesting to investigate to what extent presentation teams with perceptually easily distinguishable agents foster this kind of source monitoring (see also Craig et al. 1999, experiment 2).

6 Related Work

The generation of dialogues between multiple virtual presenters is a complex endeavor that requires research in a variety of disciplines including computer science, sociology, psychology, dramaturgy, and art and design. In this section, we will restrict ourselves to related work done in the intelligent user interfaces and natural language communities.

6.1 Animated Presenters

A number of research projects have explored lifelike agents as a new means of computer-based presentation. Applications similar to PPP and AiA were described by Noma and Badler who developed a virtual humanlike presenter

(Noma and Badler 1997) based on the Jack Software (Badler, Phillips, and Webber 1993), and by Thalmann and Kalra, who produced some animation sequences for a virtual character acting as a television presenter (Thalmann and Kalra 1995). While the production of animation sequences for the TV presenter requires a lot of manual effort, the Jack presenter receives input at a higher level of abstraction. Essentially, this input consists of text to be uttered by the presenter and commands such as *pointing* and *rejecting*, which refer to the presenter's body language. Nevertheless, here, the human author still has to specify the presentation script, while in the PPP and AiA systems, this process was automated. In addition, in contrast to the approach presented here, both systems employ just one agent for presenting information.

Byrne (Binsted and Luke 1999) and Mike (Matsubara et al. 1999) are two other systems that generate real-time natural language commentary on the RoboCup simulation league. A comparison between these systems and Rocco I, the predecessor of the commentator system described here, can be found in (André et al. 99). Mike and the previous Rocco system address the generation of expressive speech, but do not rely on animated characters. Most similar to our new commentary system is Byrne, since it makes use of an embodied commentator which is represented by a talking head. However, in order not to distract the audience too much, we decided not to show our agents from the front as in the case of Byrne.

The Agneta and Frida system (Höök et al. 1999) incorporates narratives into a Web environment by placing two characters on the user's desktop. These characters watch the user during the browsing process and make comments on the visited Web pages. Unlike our approach, the system relies on preauthored scripts, and no generative mechanism is employed. Consequently, the system operates on predefined Web pages only.

The system by Cassell and colleagues automatically generates and animates dialogues between a bank teller and a bank employee with appropriate synchronized speech, intonation, facial expressions, and hand gestures (Cassell et al. 1994). However, their focus is on the communicative function of an utterance and not on the personality and the emotions of the single speakers. Furthermore, they do not aim to convey information from different points of view but restrict themselves to a question-answering dialogue between the two animated agents.

Mr. Bengo (Nitta et al. 1997) is a system for the resolution of disputes with three agents: a judge, a prosecutor, and an attorney that is controlled by the user. The prosecutor and the attorney discuss the interpretation of legal rules. Finally, the judge decides who the winner is. The system is noteworthy because it includes a full multimodal interface consisting of components for the recognition and synthesis of speech and facial displays. The virtual agents are able to exhibit some basic emotions, such as anger, sadness, and surprise, by means of facial expressions. However, they do not rely on any other means, such as linguistic style, to convey personality or emotions.

Hayes-Roth and colleagues have implemented several scenarios following the metaphor of a virtual theater (Hayes-Roth, van Gent, and Huber 1997). Their characters are not directly associated with a specific personality. Instead, they are assigned a role and have to express a personality that is in agreement with this role. A key concept of their approach is improvisation. That is, characters spontaneously and cooperatively work out the details of a story at performance time, taking into account the constraints of directions either coming from the system or a human user. Even though the communication of information by means of performances was not the main focus of the work by Hayes-Roth and colleagues, the metaphor of a virtual theater can be employed in presentation scenarios as well.

The benefit of agent teams has also been recognized by developers of tutoring systems. For instance, Rickel and Johnson extended their one-on-one learning environment with additional virtual humans that may serve as instructors or substitute missing team members (Rickel and Johnson 1999). The main difference between their work and ours is that their agents directly address the user while in our case information is conveyed implicitly by means of a simulated dialogue between the agents.

6.2 Generation of Argumentative Discourse

Much work has been done on formal frameworks of argumentation and the generation of argumentative discourse. Most related to our work is the approach by Jameson and colleagues who developed a dialogue system that models noncooperative dialogues between a used car seller and a buyer (Jameson et al. 1995). The system is able to take on both the role of the seller and the buyer. In the role of the seller, the system tries to build up a usable model of the buyer's interests, in order to anticipate her reactions to the system's future dialogue contributions. In the role of the buyer, the system tries to arrive at a realistic estimation of the car's quality. However, while the objective of Jameson and colleagues is the generation of dialogue contributions that meet the goals of the single agents, our focus is on the development of animated agents that convey information by giving performances. Furthermore, Jameson and colleagues do not animate their agents and just produce written text. Consequently, they are not able to express human and social qualities, such as emotion and personality, through facial expressions and speech.

6.3 Conveying Emotions and Personality

Hovy describes one of the first natural language generators that not only is driven by the goal of information delivery but also considers pragmatic goals, such as conveying the social relationship between speaker and listener, during the generation process (Hovy 1987). His generation system PAULINE is able to produce a number of linguistic variants in dependency of parameters, such as the tone of interaction, the speakers opinion, and the available time.

While Hovy focuses on the generation of text, Walker and colleagues examine how social factors influence the semantic content, the syntactic form and the acoustic realization of conversations (Walker, Cahn, and Whittaker 1997). The generation of their dialogues is essentially influenced by the power the listener has on the speaker and the social distance between them. Such factors could be considered in our approach by treating them as additional filters during the generation process.

Recent work in the area of animated agents considers the full range of communicative behaviors including not only linguistic style but also body gestures and facial expressions. Ball and Breese present a bidirectional model of personality and emotion represented by Bayesian networks (Ball and Breese, this volume). The idea is to treat personality and emotion as unobservable variables in such networks and to define model dependencies between these unobservable variables and observable ones, such as linguistic style and facial expressions. The approach is noteworthy since it makes use of a uniform mechanism for both the diagnosis and the expression of emotions and personality that can be easily extended and modified. Furthermore, it accounts for the uncertainty that is characteristic of this domain.

Our own approach has been very much inspired by Cassell and colleagues, who follow a communication-theoretic approach and present an architecture based on discourse functions (Cassell et al., this volume). The goal of their approach is to interpret and generate conversational behaviors in terms of the conversational functions they have to fulfill in a dialogue. Pelachaud and Poggie have taken a similar approach, but they have concentrated mainly on the generation of facial displays (Pelachaud and Poggi, this volume).

7 Conclusion

In this chapter, we proposed performances given by a team of characters as a new form of presentation. The basic idea is to communicate information by means of simulated dialogues that are observed by an audience. This new generation task comprises content selection/organization, character allocation, and content rendering. Character allocation bears much resemblance to the media coordination problem in multimedia interfaces. Here, the basic idea is to decompose a complex goal into atomic information units that are then forwarded to several media-specific generators, for instance, for text and graphics in WIP (André and Rist 1995), or speech and gestures in REA (Cassell and Stone 1999). In a similar way, dialogue contributions may be allocated to the individual agents. However, while systems like WIP may start from a set of available media, in our case, new characters first have to be designed for each application, taking into account their specific task.

We have investigated these issues in two different application scenarios and implemented demonstrator systems for each of them. In the first application, a sales scenario, the dialogue contributions of the involved

characters are determined by a central planning component. In contrast, in the second application, we use a distributed planning approach to have two characters jointly watch and comment on soccer games. A main feature of the generated presentations is that the characters not only communicate plain facts about a certain subject matter but present them from a point of view that reflects their specific personality traits and interest profiles.

The purpose of our demonstration systems was not to implement a more or less complete model of personality for characters, such as a seller, a customer, or a soccer fan. Rather, the systems have been designed as test beds that allow for experiments with various personalities and roles. First informal system tests were quite encouraging but have also revealed some interesting aspects as to the importance of the matching of personality traits and surface characteristics, such as pitch of voice.

Currently, both systems are based on a small set of dialogue strategies and personality traits. It was, in fact, our intention to rely on rather simplified models to represent the characters' knowledge of the domain and interest profiles. One of the reasons was to show that recognizable effects can also be obtained by varying a small set of parameters. Furthermore, we wanted to demonstrate that the generation of appealing presentation dialogues requires only a minimal amount of additional knowledge modeling. For instance, in the sales scenario, we had to augment the car database by propositions that represent the impact of attributes on value dimensions. Of course, more fine-grained models may be used as well. Since our approach provides a declarative representation formalism for character and dialogue modeling, new dialogue strategies and personality traits may be added easily.

Our test beds provide a good point of departure for further research. In particular, we would like to systematically investigate further dialogue types to shed light on questions such as the following: What is the optimal number of roles, and what should an optimal casting look like? Which personalities do users prefer in which situations (see also Nass et al. 1995 and Nass, Isbister and Lee, this volume, for experiments devoted to this issue). Currently, these tasks are performed by a human user. From a technical point of view, it is also interesting to investigate to what extent the role casting and the assignment of personalities can be automated.

Acknowledgments

We are grateful to Marc Huber for his excellent support concerning the JAM framework. Many thanks also to Sabine Bein and Peter Rist for providing us with the Gerd and Matze cartoons and to Jochen Müller and Dirk Völz for technical support. We are also grateful to Justine Cassell for many useful and insightful comments. Part of this research was supported by the German Ministry for Education, Science, Research and Technology (BMBF) under contract 01 IW 806 and by the European Community under the contracts ERB 4061 PL 97-0808 and EP-29335.

References

Alpert, S. R., M. K.Singley, and J. M. Caroll. 1999 Intelligent virtual agents in an intelligent tutoring system. In *AIED-Workshop on Animated and Personified Pedagogical Agents*, 10–17. Le Mans, France.

André, E., and T. Rist. 1995. Generating coherent presentations employing textual and visual material. *Artificial Intelligence Review,* Special Issue on the Integration of Natural Language and Vision Processing 9(2–3):147–165.

———. 1996. Coping with temporal constraints in multimedia presentation planning. In *Proceedings of the AAAI '96,* vol. 1:142–147. Menlo Park, Cambridge, London: AAAI Press/The MIT Press.

André, E., G. Herzog, and T. Rist. 1997. Generating multimedia presentations for RoboCup soccer games. In H. Kitano, ed., *RoboCup '97: Robot Soccer World Cup I*, 200–215. New York: Springer.

André, E., T. Rist, and J. Müller. 1999. Employing AI methods to control the behavior of animated interface agents. *Applied Artificial Intelligence* 13:415–448.

André, E., K. Binsted, K. Tanaka-Ishii, S. Luke G. Herzog, and T. Rist. 1999. Three RoboCup Simulation League Commentator Systems. AI Magazine. to appear.

Badler, N. I., C. B. Phillips, and B. L. Webber 1993. *Simulating humans: Computer graphics, animation and control*. New York: Oxford University Press.

Ball, G. and J. Breese. 1999. Emotion and personality in a conversational character. This volume.

Bates, J. 1994. The role of emotion in believable agents. *Communications of the ACM* 37:122–125.

Binsted, K., and S. Luke. 1999. Character Design for Soccer Commentary. In M. Asada and H. Kitano, eds., *RoboCup-98: Robot Soccer World Cup II*, 22-33. New York: Springer.

Bower, G. H., M. C. Clark, A. M. Lesgold, and D. Winzenz. 1969. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior* 8:323–343.

Cahn, J. 1990. The generation of affect in synthesized speech. *Journal of the American Voice I/O Society* 8:1–19.

Cassell, J., and H. Stone. 1999. Living hand to mouth: Psychological theories about speech and gesture in interactive dialogue systems. AAAI 1999 Fall Symposium on Narrative Intelligence, 34-42. Menlo Park: AAAI Press.

Cassell, J., T. Bickmore, L. Camphell, H. Vilhjalmsson, and H. Yan. 1999. The human conversation as a system framework: Designing embodied conversational agents. This volume.

Cassell, J., C. Pelachaud, N. I. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. 1994. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. Computer Graphics (SIGGRAPH '94 Proceedings), 28(4):413-420.

Collier, G. 1985. Emotional expression. Hillsdale, N.J.: Lawrence Erlbaum.

Craig, S. D, B. Gholson, M. H. Garzon, X. Hu, W. Marks, P. Wiemer-Hastings, and Z. Lu. 1999. Auto Tutor and Otto Tudor**.** In *AIED-Workshop on Animated and Personified Pedagogical Agents*, 25–30. Le Mans, France.

Dehn, D. M., and S. van Mulken. 1999. The impact of animated interface agents: A review of empirical research. *Journal of Human-Computer Studies*. to appear.

Dryer, D. C. 1999. Getting personal with computers: How to design personalities for agents. *Applied Artificial Intelligence* 13:273–296.

Elliott, C. 1992. The affective reasoner: A process model of emotions in a multi-agent system. Ph.D. diss. Northwestern University. The Institute for the Learning Sciences, Technical Report No. 32.

Frank, I. 1997. Football in recent times: What we can learn from the newspapers. In H. Kitano, ed., *RoboCup '97: Robot Soccer World Cup I,* 216–230. New York: Springer.

Furnham, A. 1990. Language and personality. In H. Giles and W. P. Robinson, eds., *Handbook of language and social psychology*, 73–95. Chichester, England: John Wiley & Sons.

Gallaher, P. E. 1992. Individual differences in nonverbal behavior: Dimensions of style. *Journal of Personality and Social Psychology* 63(1):133–145.

Hayes-Roth, B., R. van Gent, and D. Huber. 1997. Acting in character. In R. Trappl and P. Petta, eds., *Creating personalities for synthetic actors,* 92–112. New York: Springer.

Höök, K., M. Sjölinder, A.-L. Ereback, and P. Persson. 1999. Dealing with the lurking Lutheran view on interfaces: Evaluation of the Agneta and Frida System. In *Proceedings of the i3 Spring Days Workshop on Behavior Planning for Lifelike Characters and Avatars*. 125-136. Sitges, Spain**.**

Hovy, E. 1987. Some pragmatic decision criteria in generation. In G. Kempen, ed., *Natural language generation*, 3–17. Dordrecht: Martinus Nijhoff Publishers.

Huard, R., and B. Hayes-Roth. 1997. Character mastery with improvisational puppets. In *Proceedings of the IJCAI-97 Workshop on Animated Interface Agents: Making them Intelligent*, 85–89. Nagoya, Japan.

Huber, M. 1999. JAM: A BDI-theoretic mobile agent architecture. *In Proceedings of Autonomous Agents '99*:236–243. New York: ACM Press.

Nass, C., K. Isbister, and E.-J. Lee. 1999. Truth is beauty: Researching embodied conversational agents. This volume.

Jameson, A., R. Schäfer, J. Simons, and T. Weis. 1995. Adaptive provision of evaluation-oriented information: Tasks and techniques. In C. S. Mellish ed., *Proceedings of IJCAI '95*, 1886–1893. San Mateo, Calif.: Morgan Kaufmann.

Kitano, H., M. Asada, Y. Kuniyoshi, I. Noda, E. Osawa, and H. Matsubara. 1997. RoboCup: A challenging problem for AI. *AI Magazine* 18(1):73–85.

Lang, P. 1995. The emotion probe: Studies of motivation and attention. *American Psychologist* 50(5):372–385.

Lester, J. C., S. A. Converse, S. E. Kahler, S. T. Barlow, B. A. Stone, and R. S. Bhogal. 1997. The persona effect: Affective impact of animated pedagogical agents. In S. Pemberton, ed., *Human factors in computing systems, CHI'97 conference proceedings*, 359–366. New York: ACM Press.

Lester, J. C., S. G. Towns, C. Callaway, J. L. Voerman and P. J. FitzGerald. 1999. Deictic and emotive communication in animated pedagogical agents. This volume.

Matsubara, H., I. Frank, K. Tanaka-Ishii, I. Noda, H. Nakashima and K. Hasida. 1999. Character Design for Soccer Commentary. In M. Asada and H. Kitano, eds., *RoboCup-98: Robot Soccer World Cup II*, 34-49. New York: Springer.

McCrae, R. R., and O. P. John. 1992. An introduction to the five-factor model and its implications. *Journal of Personality* 60:175–215.

Mehlmann, O., L. Landvogt, A. Jameson, T. Rist, and R. Schäfer. 1998. Einsatz Bayes'scher Netze zur Identifikation von Kundenwünschen im Internet. *Künstliche Intelligenz* 3(98):43–48.

Microsoft. 1999. Microsoft Agent: Software Development Kit. Redmond, Wash.: Microsoft Press. <http://microsoft.public.msagent.>

Moffat, D. 1997. Personality parameters and programs. In R. Trappl and P. Petta, eds., *Creating personalities for synthetic actors*, 120–165. New York: Springer.

Nass, C., Y. Moon, B. J. Fogg, B. Reeves, and D. C. Dryer. 1995. Can computer personalities be human personalities? *International Journal of Human-Computer Studies* 43:223–239.

Nass, C., B. Reeves, and G. Leshner. 1996. Technology and roles: A tale of two TVs. *Journal of Communication* 46(2):121–128.

Nitta, K., O. Hasegawa, T. Akiba, T. Kamishima, T. Kurita, S. Hayamizu, K. Itoh, M. Ishizuka, H. Dohi, and M. Okamura. 1997. An experimental multimodal disputation system. In *Proceedings of the IJCAI '97 Workshop on Intelligent Multimodal Systems*.

Noma, T., and N. Badler. 1997. A virtual human presenter. In *Proceedings of the IJCAI '97 Workshop on Animated Interface Agents: Making them ntelligent*, 45-51.

Ornstein, P. A., and T. Trabasso. 1974 To organize is to remember: The effects of instructions to organize and to recall. *Journal of Experimental Psychology* 103(5):1014–1018.

Ortony, A., G. Clore, and A. Collins. 1988. *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Pelachaud, C., and I. Poggi. 1999. Performative facial expressions in animated faces. This volume.

Rickel, J., and W. L. Johnson. 1999. Virtual Humans for Team Training in Virtual Reality. In Proceedings of the *Ninth International Conference on Artificial Intelligence in Education*, 578-585. Amsterdam: IOS Press.

Rosenbaum, D. 1969. Die Sprache der Fußballreportage im Hörfunk. Dissertation. Universität des Saarlandes. Philosophische Fakultät.

Schachter, D. L., J. L. Harbluk, and D. R. McLachlan. 1984. Retrieval without recollection: An experimental analysis of source amnesia. *Journal of Verbal Learning and Verbal Behavior* 23:593–611.

Scherer, K. R. Personality markers in speech. 1979. In K. R. Scherer and H. Giles, eds., *Social markers in speech*, 147–209. Cambridge: Cambridge University Press.

Spiegel-Verlag. 1993. SPIEGEL-Dokumentation: Auto, Verkehr und Umwelt. Hamburg: Augstein.

Thalmann, N. M., and P. Kalra. 1995. The simulation of a virtual TV presenter. In *Computer graphics and applications*, 9–21. Singapore: World Scientific Press.

van Mulken, S., E. André, and J. Müller. 1999. An empirical study on the trustworthiness of lifelike interface agents. In H.-J. Bullinger and J. Ziegler, eds., *Human-Computer Interaction* (Proceedings of HCI-International 1999), 152–156. Mahwah, New Jersey: Lawrence Erlbaum Associates.

in: J. Cassell, S. Prevost, J. Sullivan, and E. Churchill: Embodied Conversational Agents, The MIT Press, pp. 220-255, 2000.

———.. 1998. The persona effect: How substantial is it? In H. Johnson, L. Nigay and C. Roast, eds., *People and Computers XIII* (Proceedings of HCI-98), 53-66. Berlin: Springer.

Walker, M., J. Cahn, and S. J. Whittaker. 1997. Improving linguistic style: Social and affective bases for agent p[ersonality. In *Proceedings of Autonomous Agents'97*, 96–105. Marina del Ray, Calif.: ACM Press.

von Winterfeldt, D., and W. Edwards. 1986. Decision analysis and behavioral research. Cambridge: Cambridge University Press.