# Challenges for Virtual Humans in Human Computing

Dennis Reidsma, Zsófia Ruttkay, Anton Nijholt

Human Media Interaction, University of Twente,
PO Box 217, 7500AE Enschede, The Netherlands
{d.reidsma,z.m.ruttkay,a.nijholt}@ewi.utwente.nl
http://hmi.ewi.utwente.nl/

## 1  Introduction

The vision of Ambient Intelligence (AmI) presumes a plethora of embedded services and devices that all endeavor to support humans in their daily activities as unobtrusively as possible. Hardware gets distributed throughout the environment, occupying even the fabric of our clothing. The environment is equipped with a diversity of sensors, the information of which can be accessed from all over the AmI network. Individual services are distributed over hardware, share sensors with other services and are generally detached from the traditional single-access-point computer (see also the paper of Pantic *et al.* in this volume [51]).

'Unobtrusive support' means that where possible the user should be freed from the necessity of entering into an explicit dialog with all these services and devices. The environment shifts towards the use of *implicit interaction*, that is, "interactions that may occur without the behest or awareness of the user" [36]. However, not *all* interactions between user and environment will be implicit. It may not be possible, or it may not be desirable, e.g. because the user does not want to feel a loss of control over certain aspects of his environment. So how does the user achieve the necessary explicit interaction? Will (s)he address every query for information to the specific device or service that ultimately provides the information? Will (s)he give commands to the heating system, the blinds and the room lighting separately? Will each service and each device carry its own interaction interface? Clearly not. Interfaces will come to be developed that abstract from individual services and devices and offer the user access to certain combined functionalities of the system. The interfaces should support the mixture of explicit and implicit, and reactive and proactive, interaction required for a successful AmI environment. Finally, AmI environments are inherently multi-user, so the interface needs to be able to understand and engage in addressed multiparty interaction [48]. We argue that Virtual Humans (VHs) are eminently suited to fullfill the role of such interfaces.

An AmI environment can serve various purposes. It can be a home environment, an office environment, a public space or it can be used in an educational setting. Virtual humans can be available, among others, as friend, exercise adviser, health care specialist, butler, conversation partner or tutor. Sometimes

they know things better than you do, sometimes they have more control over parts of the AmI environment than you have and sometimes they persuade you to do things differently. You may not always like all aspects of the virtual humans that cohabit your house. Maybe the virtual tutor that is available to monitor your children's homework sometimes takes decisions that are not liked by your children at all. Your virtual friend is not very interesting if it always agrees with your opininions. A healthcare agent has to be strict. A virtual human that acts as a conversational partner for your grandmother may have some peculiar behavior sometimes (like a dog or cat has; remember the Tamagotchi). As in collaborative virtual environments we can have remote participation in activities in AmI environments. Virtual humans can then represent family members (with all their characteristics, including weaknesses) that are abroad and that nevertheless take part in family activities. Transformation of communicative behavior of virtual humans that represent real humans can be useful too [4]. Summarizing, in the AmI environments we foresee that virtual humans can play human-like roles and need human-like properties, including (semi-) autonomous behavior, personalities, individual characteristics and peculiarities.

However, the vast majority of existing, implemented applications of Virtual Humans are focused around one clear task, such as selling tickets, learning a skill, answering questions or booking flights and hotels, and are accessed in a clearly explicit manner. There, one can take it as a given that the attention of the user is on the system and the user is primarily engaged with the interaction. In an AmI environment this is no longer true. A dialog with a Virtual Human may be altogether secondary to several other activities of the user. A dialog with a Virtual Human may also be about many different issues, pertaining to different aspects of the environment, in parallel. This has a lot of impact on many aspects of a Virtual Human.

In the rest of this paper we will examine a (not neccesarily exhaustive) number of aspects to Virtual Humans that we feel as most relevant to their introduction in a complex AmI environment. Some of these aspects relate to the embedding of the Human / Virtual Human interaction in ongoing daily activities: issues of synchronization, turn taking and control. Other points touch upon the fictional/real polemic: how realistic should VHs be? Should a VH interface exhibit 'socially inspired' behaviour? Should a VH exhibit also the imperfections and shortcomings so characteristic of human communication? Some of the points will be illustrated with examples from our recent work on Virtual Humans, summarized in Section 4.

### 1.1   Structure of this paper

Different parts of this paper have appeared earlier in contributions to the workshops on Human Computing at the ICMI [64] and the IJCAI [65]. This paper supercedes and significantly extends the earlier publications. The paper is structured as follows. Section 2 sketches the background of Human Computing as we see it. Section 3 discusses the state of the art on Virtual Human research. Section 4 presents three example applications that have been developed at the

Human Media Interaction (HMI) group, that will be used to illustrate points throughout the paper. Subsequently, a number of aspects of VHs are discussed. Most of the aspects are centered around the general question of how human VHs really should be, which is raised in Section 5 and further elaborated in Sections 6 through 9. The paper ends with a short discussion in Section 10.

## 2   Human Computing: Background

Applications of computing technology are largely determined by their capabilities for *perception* of user and environment and for *presentation* of information to the user. An important characteristic of Human Computing is that the perception goes beyond the observation of superficial actions and events towards progressively more complex layers of *interpretation* of what is observed. Starting from observable events (form), such as gestures or speech, the system uses models of human interaction to interpret the events in terms of content, such as intentions, attitude, affect, goals and desires of the user. Elsewhere in this volume, the means of interpreting the users' behaviour are classified as the front end of Human Computing [51]. Conversely, in the other side of the system, which might then be called the back end, the intentions and goals of the system are realised in concrete actions, which have some direct effect on the environment and/or involve communication to the user.

   The models of how humans behave and communicate, mentioned above, are an integral part of the Human Computing paradigm. Not only for an adequate interpretation of the perceptions, but also to inform and inspire the patterns of interaction that should occur between system and user (see also [36]), and, especially when Virtual Humans inhabit the environment, as a basis for the realisation of the communicative intentions of the system. The concept is illustrated in Figure 1 and discussed in more detail in [58] and [60].

## 3   Virtual Humans: Related Work and State of the Art

Virtual Humans (VHs) [53] - also known under other names like humanoids [70], embodied conversational agents [11] or intelligent virtual agents - are computer models resembling humans in their bodily look and their communication capabilities. A VH can speak, display facial expressions, gaze and nod, use hand gestures and also perform motions of biological necessity, such as blinking or changing posture. Besides these 'back-end' related channels, a VH ideally should be equally capable in understanding speech and perceiving non-verbal signals of his human conversant partner. In generating and interpreting bodily signals, the VH should be mindful, to orchestrate his modality usage according goals, the state of the environment and last but not least, social communication protocols. From the point of view of HCI, there are two major motivation of 'putting a VH on the screen':

  – By replacing the traditional computer-specific interaction modalities (keyboard, mouse) by the natural communicational capabilities of people, the
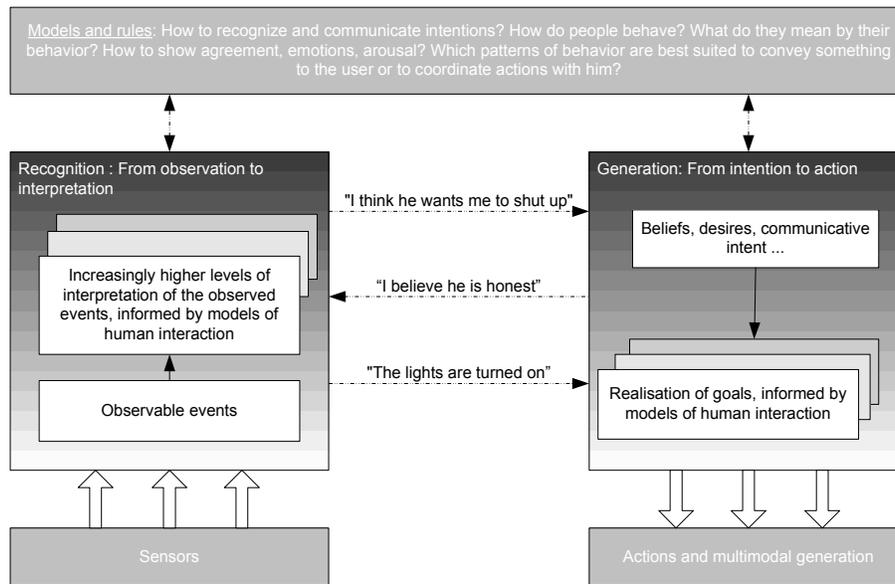
**Fig. 1.** Schematic overview of the various elements in the front-end and back-end of human computing, from observations via interpretations to appropriate reactions, mediated by various models expressing knowledge about the behavior of people.

services of computers will be accessible to a broad population, without respect to (computer) literacy, cultural and social background;
– VHs make new applications possible, where they fulfill traditional human roles, such as tutor, salesperson, and partner to play a game or chat with. Moreover, there are applications where the VH is in an entirely new role, without parallel in real life, such as human-like characters with fictional capabilities in (educational) games [28], interactive drama [43] or Virtual Reality applications [1], or a VH taking partial or split role known from real-life situations [6].

Often it is not easy to separate the two aspects in an application. For instance, a VH explaining how to operate a device [3], can be seen as a friendly interface replacing queries and search in an on-line textual help, but he also extends the services, by being able to demonstrate operations, possibly in a way tailored to the user's capabilities (e.g. handedness).

A key issue concerning VHs is the question whether they do indeed 'work'. In today's state of the art applications they cannot be mistaken for a real person. It is clear that they exist in a digital form, residing on a PC screen or projected to a wall as a set of pixels. Moreover, due to features in appearance and communication, they are not considered as a recording of a real person. The 'betraying features' may result from a conscious design decision to create a fictional char-

acter with limited human intelligence [38], or - more often - from technological limitations such as not having fast enough models to simulate hair, or Text-To-Speech engines capable of producing synthetic speech with intonation and other meta-speech characteristics. In spite of these observations, there are two basic principles which make VHs promising as effective and engaging communicating partners of real people.

1. In a broader context, it was shown that humans treat computers as social actors [56]. We tend to talk about (and to) computers as living creatures, ask them to perform some task, get emotional about their performance and provide feedback accordingly. One may expect that it is even more so if we see a human-like creature representing the services of the underlying system.
2. With the appearance of a human-like character on the screen, all the subtleties (or lack of them) reminiscent of human communication will be interpreted and reacted to. Though we know that we are not facing a real person, we "suspend our disbelieve" against the VH: we do not look at them as cool computer engineering achievements, but react to them as we do to real people. This does not require, per se, photorealism in appearance and behavior. The very term was coined in connection with successful Disney animation figures, who thank their success to consistent behavior, both concerning changes in their emotions and goals reflecting some personality and in the way how this is conveyed in their speech and body language. Expressivity is achieved by using subtle phenomena of human communication in an enhanced or exaggerated - thus non-realistic - way.

Human-like appearance and behavior of the interface, in particular when the computer is hidden behind a virtual human, indeed strengthen the acceptance of the computer as a social actor [23] or even a persuasive social actor [22] that changes our attitudes and behavior.

But what design rules are to be followed making a 'successful' VH? One whom the users will like, find easy to communicate with... and who helps the user to perform the task required? The quality of the signals of communications modalities used by the VH should be good: his speech intelligible, his motion not jerky, his face capable of expressing emotions. In the past decade, much effort has been spent on improving the human-likeness of individual modalities of VHs, such as improving the quality of synthesized speech [45], modeling expressive gesturing of humans [30; 31], deriving computational models to capture the kinematics [72], providing means to fine-tune the effort and shape characteristics of facial expressions and hand gestures [14], model gaze and head behavior and adding biological motions like blinking or idle body motion [20]. The fusion of multiple modalities has been dealt with, from the point of view of timing of the generated behavior, and the added value of using multiple modalities in a redundant way.

Besides these low-level technical requirements, several subtle human qualities have been proven to play an essential role in both the subjective judgment of a (virtual) human partner and the objective influence of him. It has turned out that by using smalltalk, common ground can be created and as a consequence,

the user will trust the VH more [8]. In addition to task-related feedback, showing interest and providing empathic reactions [19] contribute to the success of task performance e.g. in tutoring [46; 41] and coaching. Emotional reactions are reminiscent of humans. Modeling emotions and their benefits in judging the VHs have been extensively addressed in recent works [25; 42]. Initially, the 6 basic emotions were to be shown on the face [21], which has been followed by research on the display of other emotions and of cognitive states, taking into account display rules regulating when emotions are to be hidden [54] or overcast by fake expressions e.g. to hide lies [57], studying principles to show mixed emotions on the face and to reflect emotions in gesturing [50]. Besides temporary emotions, having long-term attitude towards a VH, like friendship, has been pointed out [69].

Moreover, people attribute some personality to VHs based on signals of non-verbal communication, the wording of speech and physical impressions of the embodiment. It has been shown that signals of modalities such as posture or the intonation of synthetic speech are sufficient to endow VHs with a personality [47]. From these experiments it was also clear that people tend to prefer a virtual communicative partner with similar (or in some cases, complementary) personality. So an effective VH should be endowed with a personality best matching the envisioned user. Thus no single universal, best VH may be designed, but characteristics of the user - such as personality, but also age, ethnicity, gender - need to be taken into account. The importance of cultural and social connotation of a VH has been pointed out [52; 55].

It has also been mentioned that virtual humans should be individuals, in their behavior, verbal and non-verbal style, body, face and outfit [62]. Some subtle aspects of interaction, like the style of the VH [63] and his/her attitude towards the user have proven to be important in judging them.

As of the less investigated bodily design, some case studies show that similar criteria of judgment are to be taken into account when deciding about gender, age, ethnicity, formal or casual dressing [44], or even (dis)similarity to the human user [5] or non-stereotypical embodiment like a fat health-advisor [71]. Other dimensions of VH design are to do with human-likeness and realism. Should the VH be humanlike, or are animals or even objects endowed with human-like talking/gesturing behavior more appropriate? The application context may be decisive: for children, it may be a factor of extra engagement when the tutor is a beloved animal known from a favorite tale (or TV program), or when an animal or object to be learned about is capable of telling informative facts about 'himself'. But non-human characters may be beneficial for adult users too, to suggest the limited mental capabilities of the VH (using e.g. a dog [35], or a paperclip as Microsoft did). The degree of realism of the VH is another dimension. Most of the current research has been driven by the motivation to be able to reproduce human look and behavior faithfully. Recently, the expressivity in non-photorealistic and artistic rendering [61] and in non-realistic communication as done in traditional animation films [74; 12] and on theater stage [68] are getting a place in VH design.

The overall objectives of likability, ease of use and effectivity may pose conflicting design requirements. In an early study it was shown that even merely a slightly frowned eyebrow from a virtual human can result in measurable differences: a stern face was more effective, but liked less, than a neutral face [73]. There are some applications like entertaining games or crisis management where basically only one of the two objectives is to be met. However, in most of the applications both factors play a role, and the two objectives need to be balanced with respect to the progress made and the current emotional and cognitive state of the user.

The potentials of VHs are huge. There are still major challenges to specific disciplines to improve bodily and mental capabilities of VHs along the above dimensions, and to compile these components into full-fledged, consistent believable virtual characters [26]. Also, the necessity of cooperation of different disciplines, and particularly, dedicated studies providing basis for computational models of human-human interaction, have been underlined [34]. In such a 'big scenario' with yet many problems to be solved, many more subtle issues of human - (virtual) human communication are not (yet) looked at. In the rest of the paper, we focus on such issues. They are essential to have VHs in an AmI environment, where they are active and present in the daily life of the users. On the other hand, several of these subtle issues have major consequences on the design of VHs and raise further, principal questions concerning the VHs. This is the major focus of the forthcoming sections. In this paper we do not address the technical feasibility of the envisioned characteristics, as each of them would require in-depth expertise from specific fields. However, we note that several of the features we discuss would be feasible with present-day technology, while others would need research in new directions in some disciplines both to provide computational models and good enough implementations.

## 4   Virtual Humans: Engagement and Enjoyment

### 4.1   Introduction

In this section, we present three applications currently being developed at the HMI (Human Media Interaction) research group: the Virtual Dancer [59], the Virtual Conductor [9] and the Virtual Trainer [67]. These three novel applications are summarized in preparation to our general discussion on subtleties of VHs, where we will use illustrative examples from the applications. All three applications require virtual humans with capabilities beyond the ones in more restricted or traditional functions such as providing information or tutoring. These seemingly very different applications share some basic features, and have actually been developed relying on a similar framework. In all three applications, the VH:

- has visual and acoustic perception capabilities,
- has to monitor and react to the user continuously,
- has to use subtle variants of a motion repertoire generated on the fly, and

– uses both acoustic (music, speech) and nonverbal modalities in a balanced and strongly interwoven manner.

### 4.2   A Dancer

In a recent application built at HMI, a virtual human - the Virtual Dancer - invites a real partner to dance with her [59]. The Virtual Dancer dances together with a human 'user', aligning its motion to the beat in the music input and responding to whatever the human user is doing. The system observes the movements of the human partner by using a dance pad to register feet activity and the computer vision system to gain information about arm and body movements. Using several robust processors, the system extracts global characteristics about the movements of the human dancer like how much (s)he moves around or how much (s)he waves with the arms. Such characteristics can then be used to select moves from the database that are in some way 'appropriate' to the dancing style of the human dancer.

There is a (non-deterministic) mapping from the characteristics of the observed dance moves to desirable dance moves of the Virtual Dancer. The interaction model reflects the intelligence of the Virtual Dancer. By alternating patterns of following the user or taking the lead with new types of dance moves, the system attempts to achieve a mutual dancing interaction where both human and virtual dancer influence each other. Finding the appropriate nonverbal interaction patterns that allow us to have a system that establishes rapport with its visitors is one of the longer term issues being addressed in this research.

Clearly, the domain of dancing is interesting for animation technology. We however focus on the interaction between human and virtual dancer. The interaction needs to be engaging, that is, interesting and entertaining. Efficiency and correctness are not the right issues to focus on. In this interaction perfectness can become boring and demotivating. First experiences with demonstration setups at exhibitions indicate that people are certainly willing to react to the Virtual Dancer.

### 4.3   A Conductor

We have designed and implemented a virtual conductor [9] that is capable of leading, and reacting to, live musicians in real time. The conductor possesses knowledge of the music to be conducted, and it is able to translate this knowledge to gestures and to produce these gestures. The conductor extracts features from the music and reacts to them, based on information of the knowledge of the score. The reactions are tailored to elicit the desired response from the musicians.

Clearly, if an ensemble is playing too slow or too fast, a (human) conductor should lead them back to the correct tempo. She can choose to lead strictly or more leniently, but completely ignoring the musicians' tempo and conducting like a metronome set at the right tempo will not work. A conductor must incorporate some sense of the actual tempo at which the musicians play in her conducting, or else she will lose control. If the musicians play too slowly, the virtual conductor

will conduct a little bit faster than they are playing. When the musicians follow him, she will conduct faster yet, till the correct tempo is reached again.

The input of the virtual conductor consists of the audio from the human musicians. From this input volume and tempo are detected. These features are evaluated against the original score to determine the conducting style (lead, follow, dynamic indications, required corrective feedback to musicians, etc) and then the appropriate conducting movements of the virtual conductor are generated. A first informal evaluation showed that the Virtual Conductor is capable of leading musicians through tempo changes and of correcting tempo mistakes from the musicians. Computer vision has not been added to the system. That is, musicians can only interact with the conductor through their music. In ongoing work we are looking at issues such as the possibility to have the conducting behavior directed to (the location of) one or more particular instruments and their players, experimenting with different 'corrective conducting strategies' and extending the expression range of the Virtual Conductor.

### 4.4   A Trainer

The Virtual Trainer (VT) application framework is currently under development [67] and involves a virtual human on a PC, who presents physical exercises that are to be performed by a user, monitors the users performance, and provides feedback accordingly at different levels. Our VT should fulfill most of the functions of a real trainer: it not only demonstrates the exercises to be followed, it should also provide professionally and psychologically sound, human-like coaching. Depending on the motivation and the application context, the exercises may be general fitness exercises that improve the users physical condition, special exercises to be performed from time to time during work to prevent for example RSI (Repetitive Strain Injury), or physiotherapy exercises with medical indications. The focus is on the reactivity of the VT, manifested in natural language comments relating to readjusting the tempo, pointing out mistakes or rescheduling the exercises. When choosing how to react, the static and dynamic characteristics of the user and the objectives to be achieved are to be taken into account and evaluated with respect to biomechanical knowledge and psychological considerations of real experts. For example, if the user is just slowing down, the VT will urge him in a friendly way to keep up with the tempo, acknowledge with cheerful feedback good performance and engage in a small talk every now and then to keep the user motivated.

Related work on VTs can be found in, among others, [18], where a physiotherapist is described with similar functionality as ours, [13] with an interesting Tai Chi application, and [2], reporting about work on an aerobics trainer.

## 5   How Human Should Virtual Humans Really Be?

Traditionally, there are different qualities associated with machines (including computers) and humans. Machines, and machine made products are praised for

being reliably identical and precise, irrespective of time and conditions of production, exhaustive, fast, deterministic in handling huge amounts of information or products, slavishly programmable... as opposed to humans being nonrepetitive and less predictable, error prone and nondeterministic in their 'functioning'. On the other hand, when it comes to flexibility, adaptation, error recovery, engagement... humans are valued higher.

However, it has been suggested that implicit interaction, which is an important part of the interaction in an AmI environment, should be inspired in the patterns and behaviour found in Human-Human interaction [36]. And, depending on the application, humans already treat computers as social actors [56]. Humanlike appearance and behavior of the interface, in particular when the computer is 'hidden behind a virtual human', strengthen this acceptance of the computer as a social actor [23; 47] or even a persuasive social actor [22] that changes our attitudes and behavior. Virtual Humans, more than regular graphical user interfaces, invite natural human interaction behavior and therefore it is useful to be able to capture and understand this behavior in order to let it play its natural role in an interaction [51].

This raises the question of how human Virtual Humans should be. Should a VH interface exhibit 'socially inspired' behaviour? Should a VH exhibit also the imperfections and shortcomings so often present in human communication? Do we want a VH to be hesitant or even make errors, to be nondeterministic, or, just the opposite, are we eager to use superhuman qualities made possible by the machine? Do we need the fallible human qualities to increase the naturalness and believability of human looking software entities? Are there additional, practical values too of human imperfections? Is there a third choice, namely VHs which unify the functionally useful capabilities of machines and humans, and thus are not, in principle, mere replicas of real humans?

In this paper we focus on qualities of real humans that are characteristically present in everyday life, but are hardly covered by the efforts on attaining higher level cognitive model based behavior of VHs. We look at subtleties and 'imperfections' inherent in human-human communication, and investigate the function and merits of them. By subtleties, we mean the rich variations we can employ in verbal and nonverbal communication to convey our message and the many ways we can draw attention to our intentions to convey a message. By imperfections we mean phenomena which are considered to be incorrect, imperfect according to the normative rules of human-human communication. Both enrich the model of communication, the first one by taking into account more aspects and details (e.g. emotions and personality of the speaker), the second one by being more permissive about what is to be covered. We consider imperfections as those natural phenomena of everyday, improvised language usage which are not considered to be correct and thus, are not 'allowed' by some descriptive rules of proper language usage. For instance, restarting a sentence after an incomplete and maybe grammatically incorrect fraction is such an imperfection. We are aware though of the continuous change of norms for a language, pushed by actual usage. The space here does not allow dwelling on the relationship between intelligence and

the capability of error recovery and robust and reactive behavior, in general. Here we present the issue from the point of view of communication with virtual humans and from the point of view of perception of VHs.

## 6    Virtual Humans as individuals

"Who are you?" do people ask (usually as one of the first questions) from their VH interlocutor. The answer is a name, maybe extended with the services the VH can offer. In case of chat bots, a date of birth may be given, and the creator may be named as 'father', such as in the case of Cybelle[1]. Notably, the date of 'creation' makes sense in the fictional framework only. Moreover, any inquiry about further family members is not understood. The personal history is similarly shallow and inconsistent as of her hobbies: she has a favorite author, but cannot name any title by him. Deviations from this common solution can be found, when the VH is to stand for a real, though long-dead, person [7], and the very application is to introduce the reincarnated real person and his history to the user. The other extreme is feasible when the VH is in a role, such as that of a museum guide [39], where his refusal 'to talk about any personal matters' sounds to be a natural reaction. But in other applications, where it would be appropriate, we would never know about the family, schooling, living conditions, acquaintances and other experiences of the VH, neither about his favorite food or hobbies. One may argue that that is enough, or even preferred, to remain 'to the point' in well defined task oriented application like a weather reporter or trainer. However, even in such cases in real life some well placed reference to the expert's 'own life and identity' breaks the businesslike monotonicy of the service, and can contribute to create common ground and build up trust. B. Hayes-Roth endowed her Extempo characters with some own history as part of their 'anima' [32]. From the recent past, we recall a Dutch weather forecast TV reporter who added, when a certain never heard-of Polish town was mentioned as the coldest place in Europe, that this town is special for him as his father was born there. But he could have noted about some other aspects like special food or customs he experienced or knows of from that place. In case of a real fitness trainer's video, it is remarkable how the task related talk is interwoven with references to the presenter's personal experience on where she learnt the exercises, what she found difficult, etc. A VH could use his personal background to generate just some 'idle small talk' in addition to the task related conversation, or to relate it to the stage of task completion or difficulty and the reactions from the user, in order to increase the user's commitment. So for instance, a VT may include non task related small talk at the beginning or during resting times, or add task related background information to keep the user motivated during a long and/or difficult exercise.

In order to make a VH 'personal', it is not enough to endow him with a 'personal history'. Some mechanisms should be provided to be able to decide when and what piece of personal information to tell. E.g. to derive if there

---

[1] http://www.agentland.com/

is something in the personal knowledge of the VH which could be related to the factual, task oriented information to be told. This may span from simple tasks as discovering dates, names and locations, to the really complex AI task of associative and analogical reasoning.

Finally, the disclosure of the personal information and identity is a manifestation of personality: open, extrovert people (and VHs) may interweave more their story with personal references than introvert ones.

An interesting question is that a VH's 'personal history' may be also adapted to a situation, or a given user (group), not only its conversational style as suggested for robots [17] and VHs [63]. However, consistency within different interaction sessions with the same user (group) should be taken care of.

## 7   Here and today - situatedness

VHs hardly give the impression that they know about the time and situation they converse in with their user. Some VHs do reflect the time of the day by choosing an appropriate greeting. But much more could be done: keeping track of the day, including holidays, and commenting accordingly, providing 'geographical update' capability when placing a VH enabled service in a location, endowed with some social and political information about the place. Imagine a VT who knows that it is today a public holiday in Italy where the given VT is 'active'. Some special words to the user keeping up her exercise scheme on a holiday would be appropriate. But on a tropical Summer day, the heat may lead the VT to revise its strategy, remind the user the necessity of drinking, or even shorten the exercises, or suggest doing it the next time during the morning.

The identity of the user may be a source of further situatedness. As a minimum, a VH should 'remember' earlier encounters with the user. Asking the name or telling the same piece of small talk to the same person each time is disappointing. But how nice it sounds if a VT refers to yesterday's performance, knows of the user's religion does not allowing her to do exercises on Saturday, greets her specially on her birthday.

Finally, in order to perceive a VH as 'present', the VH must have means to gather information about the user and react to it. To begin with, the mere presence of the user and her identity should be detected, and her task related performance should be monitored. But think of a real trainer or tutor, who would very likely comment on changes like not wearing glasses, change in hair style, being sunburnt or showing signs of a cold. A Virtual Trainer could do similar comments.

## 8   Shortcomings as features

'Shortcoming' is a negative word. Shortcomings in the realisation of communicative acts (e.g. mumbling, sloppy gestures ...), the information transfer (e.g. ambiguous wording, redundant or incomplete content ...) or the knowledge or decisions: at first sight they are something to be avoided when you are designing

VHs. But... people without imperfections don't exist - luckily so, as the variety in 'imperfections' also makes people different and interesting. And besides, what is a shortcoming? Traits that are considered undesirable in one culture, may be unimportant or even considered good behaviour in the next. And behavior that may be considered imperfect because it deviates from a norm imposed by 'average human behavior' may actually have a clear communicative function.

In Section 5, we raised the question whether we should aim for 'perfect' VHs or whether there are practical values to human imperfections too. In this section we want to elaborate a bit on the question, examining some examples of 'imperfect' behavior on their potential merits.

### 8.1   User Adaptation

At first sight one might require that a VH is as perfect as can be, in its cognitive decisions, conversational and motion skills. One reason for lifting this requirement may be user adaptation. As a VH may come across users of different intellectual and communicational capabilities, the capability to recognize such a situation and scale down the VHs' own functioning is more beneficial than overwhelming the user with a perfect and too demanding performance. This is already happening in the case of some TTS services, where the generated English pronunciation is not the 'official nicest' one, but one 'tortured' according to the practice of users whose mother tongue, like many Asiatic languages, has a very different acoustic scheme. Adapting to the level of a user by hiding some of the cognitive capabilities of the VH is already a usual practice in gaming and in tutoring situations. Lastly, the VT scenario exemplifies a possible reason for adapting the motion skills to those of the user by showing up in the embodiment the most appropriate for the given user. His/her gender, age and motion characteristics may be similar to the user's, in order to avoid having a too huge gap between or example a 'fit and young' trainer and the 'fattish, somewhat rigidmoving' user [71]. On the other hand, deviations from usual trainers in the other (superman) direction may have positive effects too. Imagine the VT making every now and then extreme, beyond-realistic capabilities jumps to cheer up the user or grab attention.

### 8.2   Clarification and Commitment

Certain types of imperfections in communication need 'repair'. A (virtual) human who mumbles needs to be asked to repeat itself. A (virtual) human using ambiguous language, or serious disfluencies, may be asked for clarification, literally, or through nonverbal mechanisms such as a lifted eyebrow or a puzzled expression. The traditional judgment of such repair in Human-Computer interaction is as an undesirable necessary evil, hence the term 'repair'. However, as Klein *et al.* state, one of the main criteria for entering into and maintaining a successful joint activity is "Commitment to an intention to generate a multiparty product" [37]. This commitment needs not only to be present, but must

also be communicated to the conversational partner. Imperfections in the communications of a VH, and the subsequent so-called 'repair dialogues', could be a subtle means for both VH and human to signal their commitment to the interaction. One must be really committed to an interaction if one is going through the trouble of requesting and/or giving clarifications, repeating oneself, etc...

There are two sides to this commitment issue. The first relates to imperfections at the side of the human user. When we assume that entering into clarification and repair dialogues is a strong signal of commitment to the conversation, we see a clear reason to invest in the development of techniques for clarification dialogues beyond what is needed to let the conversation reach its intended goal. One may decide to make the VH initiate clarification dialogues when they are not absolutely necessary for reaching the goal, to signal commitment to understanding the user.

The second relates to imperfections at the side of the VH. If the human is forced to ask the VH for clarification all the time, this will be considered a drawback in the interaction capabilities of the system. However, there may be certain subtle advantages to a VH that uses ambiguous, disfluent or otherwise imperfect expressions. Judicious use of ambiguities at non-critical points in the conversation, at a point where it is likely that the user will ask for clarification (explicitly or nonverbally), gives the VH a chance to show off its flexibility and its willingness to adapt to the user. This gives the human user a feeling of 'being understood' and of commitment from the side of the VH. Again, such an approach would need a lot of investment in repair and clarification capabilities of the VH.

### 8.3  Signaling Mental State and Attitude

In other situations imperfections express an important part of the content of the conversation. Imperfections in the multimodal generation process, such as hesitation, stuttering, mumbling and disfluencies can signal indications of the VH's cognitive state (e.g. 'currently thinking', 'unhappy'), conversational state ('ready to talk'), attitude towards the conversation partner or the content (belief, certainty, relevance, being apologetic).

Cognitive or conversational state is often reflected by gaze and body postures of the VHs. However, the usage of (non-speech) vocal elements in these and other situations has not been addressed widely yet. For example, by analyzing a multi-party real-life conversation, we found that non-speech elements abundantly interwove with the 'meaningful, articulated' utterances.

While some of the erroneous and nonverbal utterances reflect the 'processing deficiencies' of the speaker (e.g. difficulty in formulating a statement in correct format), others have important function in regulating the dialogue (e.g. indicating request for turn taking by making some sound) or in qualifying the 'verbatim' content (e.g. a hesitant pause before making a statement indicates that the information to be conveyed may not be correct). In our analysis of multi-party real-life conversation we could identify speech situations as well as personality,

emotional and cognitive state of the speaker as indicators of the frequency and type of non-speech elements used.

### 8.4   Other Merits of Imperfections

The above subsections presented a number of considerations for turning VHs 'imperfect'. Before ending this section we will touch lightly upon a few of the countless other themes that in some way also can involve imperfections.

Disfluencies and other imperfections are a major characteristic of spontaneous speech. If a VH talks without disfluencies it may come across as stilted, not spontaneous enough. In the ongoing work mentioned earlier we do encounter a common phenomenon in the language usage itself: speakers do not express themselves in perfect sentences, especially if they are answering an unexpected, unusual question or contribute to a discussion. Often, they abandon an erroneous start of a sentence and correct, or repeat the start in another form.

Ambiguous wordings and underspecification will help to give the user a sense of 'freedom' in the dialogue, to feel less constrained [24]. (Note though that we then run the risk of the user indulging in the same kind of ambiguous language use to an extent that the VH cannot handle it). Hesitations, mumbling and disfluencies are also mechanisms that play an important role in reducing the amount of threat potentially perceived by the human user [10].

Making mistakes may have the positive side-effects of users perceiving the VH as more believable and more individual. Humans are not perfect, and usually there is not a single 'perfect' way of action. When researchers create a model of, for example, turn taking behavior as a general norm, and we then implement this model in a VH, the fact that such a model is an abstraction by nature makes certain that the VH will behave in a way that no real human would do. Making mistakes also gives a VH unrivalled opportunity for exhibiting an individual style. After all, as humans we can probably distinguish ourselves from others through the mistakes that we make as much as through the things we do perfect.

## 9   Some Communicational Challenges for VHs in an AmI Environment

Finally, there are some communicational challenges for Virtual Humans we would like to mention here. They have a lot to do with the fact that VHs in an AmI environment need to coordinate their actions carefully to events happening in the environment, as well as to the actions of the user (see also the paper by Jeffrey Cohn in this volume [16]). The last is especially tricky when the interaction is implicit, since then the user is not aware of the interaction and one cannot rely on the simple turntaking protocols that usually regulate interaction between a VH and a user.

### 9.1   Who is in Control?

VHs enter domains where they are not to be the dominant partner, and thus the control scenario is not well established. On one hand, the VH should not be completely dominant and thus needs to be able to follow the initiative of the human partner. On the other hand, VHs that show no initiative of their own do not seem to be human-like conversational partners but more like responsive machines that are to be controlled by the user using multimodal commands.

In the Virtual Dancer application, the control between human and VH is both mixed-initiative and implicit: the VH alternates phases of 'following dancing behavior' where it incorporates elements of the humans dance in its own dance with phases of 'leading dancing behavior' where it introduces new elements which the human will hopefully pick up.

For the Conductor, control is on the side of the VH, but with a special twist. Although the conductor is leading the musicians all the time, she cannot afford not to let herself be influenced by the actions of the musicians. If an ensemble is playing too slow or too fast, a conductor should lead them back to the correct tempo. She can choose to lead strictly or more leniently, but completely ignoring the musicians tempo and conducting like a metronome set at the right tempo will not work. A conductor must incorporate some sense of the actual tempo at which the musicians play in her conducting, or else she will lose control.

In the VT scenario, the issue of control is essential, and subtle. Basically, the user is to perform the instructions given by the VT. However, this is not what happens all the time. The reaction for the VT depends on the assessment of the situation, including past performance and knowledge of the user. When the VT concludes that the user has just lost tempo, or is getting a little lazy, than the VT reinforces his/her own tempo to the user. But if the user looks very exhausted, or (s)he has a 'bad day' of decreased performance, then the VT may 'give in' to the user and slow down his/her tempo to comfort the user. Hence the VT is attentive and reactive, instead of imposing a predefined scenario on the user. The decision concerning when and how 'to give in' to the user must be based on detailed domain-specific and psychological knowledge.

### 9.2   Coordination of Modalities

There are several reasons why coordination of modalities in our applications needs a level of subtlety and sophistication beyond what is present in much current work.

For example, in present VH applications it is usually assumed that speech is the leading modality, and accompanying gestures and facial expressions should be timed accordingly. Dropping this assumption influences the design of a balanced multi-modal behavior that does not place natural language in a privileged role. It also influences the planning algorithms used to generate the behavior. We know of only one work where the instruction utterances' tempo is scaled to the duration of the hand movements in explaining an assembly task [40]. This clearly indicates how rarely this issue is attended by current research efforts.

Another challenging characteristic is the need for alignment of multimodal behavior to external channels or events. Such a feature is essential in any application where the VH is embedded in an external environment or does not converse by switching between pre-programmed active speaking and passive listening behavior but reacts according to a model of bidirectional communication behavior. 'Outside events' are usually not covered in the framework of modality coordination, but rather as an issue of planning specific gestures (e.g. pointing at or reaching for moving objects). However, the fact that behavior in any of the modalities may be constrained in its timing by sources outside the VHs' influence calls for subtle and strongly adaptive planning and re-planning.

In the Virtual Dancer application the issue of coordination focusses wholly on the alignment of the dance behavior to the music. There is also a relation between the performed dance 'moves' and the dance of the user, but as yet, there is no tight timing relation between those two.

Coordination in the Conductor application is really a mutual coordination between the intentions and actions of conductor on the one hand, and the musicians playing their parts on the other hand. One cannot simply say that the conductor coordinates her actions to the music as it is being played, nor is it realistic to assume that the conducting actions are planned first and the coordination of music played by the musicians is completely determined by that. There is a feedback loop of mutual influence going on: conducting and playing music is a joint action [15] in the truest sense.

In the VT scenario, the issue of coordinating speech, motion and a given piece of music has turned out to be central, relating to two aspects mentioned above. Namely, a virtual trainer explains postures and basic movements and conducts rhythmic exercises. The exercises may be performed in different tempi. The speech (e.g. counting and providing comments on posture), which is often of secondary importance, should be aligned to the motion. The alignment should be subtle, in one case making sure that the emphasized syllabe of the counting coincides with the end of the stroke of the movement, while in another case the counting starts together with a repetitive motion unit, or an expression is uttered in an elongated duration, during the finishing movement of a series. This requires real-time reactive planning of speech, occasionally resulting in what would usually be seen as 'unnaturally slow' utterances which are justified by the elongated duration of the corresponding motion. TTS systems may not even be prepared to generate such slowed down, unnatural speech. On the other hand, the timing of the exercise presented by the VT may be driven by external audiovisual cues of varying tempo. An exercise author may specify the tempo of an exercise by clapping or tapping, or as being aligned to music beats. However, this is often not enough in the case of VTs. For example, a VT may need to align its performance to the user doing the same exercise while counting along. This again requires a subtle real-time planning, affecting duration and alignment of sub-segments of a motion [66].

## 10    Discussion

We have argued that there are dimensions still to be exploited to turn VHs more lifelike, entertaining and in cases effective, especially in the context of Human Computing. We discussed the potentials of:

- making VHs look more individual and imperfect, may be configured to a given user's preferences;
- endowing VHs with identity and personal history;
- grounding VHs to the geographical and sociological place and time of the application being used;
- taking care of styled and natural conversation with phenomena of 'imperfections' reminiscent in real life.

The above features do not require, first of all, further improvement of single or multimodal communication on the signal level, but they do pose challenges on modeling mental capabilities like associative storytelling or require further socio-psychological studies of the nature and effect of social conversation in task related situations. What is needed, for several of the above enrichments, is multisignal perception of the conversant of a VH.

Dedicated evaluation studies are needed to put together a huge jigsaw image. It is clear already that the objective to engage the user in an activity and to perform a task well and efficiently may require a different VH design along several dimensions. Also, the application context (real/fictional) puts the user in different frame of mind to judge the VH, On the other hand, even less is known of the judgments of nonhuman capabilities of VHs. For example, it has turned out that a VH could 'read from the eye' of the user better than most of the people are capable of [49]. What to do with such a superhuman power of a VH? Or, another example is the reasoning speed and capability of a VH: do people take it as natural (from a VH) that he can recall multiple telephone books? Or should he 'fake' the human limitation of recalling data in a register? How to get away with shallow, or not deep/complete enough, models?

We have argued that carefully 'designed' imperfection in the communication may increase not only the believability but also the scope of applicability and effectiveness of the VH in question. People are are neither uniform nor perfect, but have different capabilities, and have means to recognize and cope with errors and limitations. Endowing VHs with the imperfections of humans can help making them more 'comfortable' to interact with. The natural communication of a VH should not be restricted to multimodal utterances that are always perfect, both in the sense of form and of content.

This scenario aims at hiding the 'machine' nature of the VH. This may be very much what we want on the communication level, but how about the cognitive level? For instance, if a VH is to find an item from a huge database, or perform a difficult calculation, should the imperfect  that is, slower, error prone  human behavior be mocked up, even if the computer is ready with the perfect answer immediately? In general, should a VH's amount and processing of knowledge

'resemble' the capabilities of humans, in order to make the VH believable and lifelike, as opposed to some omnipotent supercreature?

On the other hand, such a scenario may sound completely irrational, as it makes no use of the power of the computer. Given the fact that most humans are indoctrinated from birth with the adagio that 'computers are fast in calculating', hiding this capability behind artificial imperfection might even be perceived as unrealistic. Except maybe when the user is given to understand that (s)he interacts with the computer through the intermediation of the VH rather than with the computer embodied by the VH.

# Bibliography

[1] ABACI, T., DE BONDELI, R., CÍGER, J., CLAVIEN, M., EROL, F., GUTIÉRREZ, M., NOVERRAZ, S., RENAULT, O., VEXO, F., AND THALMANN, D. Magic wand and the enigma of the sphinx. *Computers & Graphics 28*, 4 (2004), 477–484.

[2] BABU, S., ZANBAKA, C., JACKSON, J., CHUNG, T., LOK, B., SHIN, M. C., AND HODGES, L. F. Virtual human physiotherapist framework for personalized training and rehabilitation. In *Proc. Graphics Interface 2005* (May 2005).

[3] BADLER, N. LiveActor: A virtual training environment with reactive embodied agents. In *Proc. of the Workshop on Intelligent Human Augmentation and Virtual Environments* (October 2002).

[4] BAILENSON, J. N., BEALL, A. C., LOOMIS, J., BLASCOVICH, J. J., AND TURK, M. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments 13*, 4 (2004), 428–441.

[5] BAILENSON, J. N., YEE, N., PATEL, K., AND BEALL, A. C. Detecting digital chameleons. *Computers in Human Behavior* (2007). in press.

[6] BAYLOR, A. L. S., AND EBBERS, S. The pedagogical agent split-persona effect: When two agents are better than one. In *Proc of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA)* (2003).

[7] BERNSEN, N. O., CHARFUELÀN, M., CORRADINI, A., DYBKJÆR, L., HANSEN, T., KIILERICH, S., KOLODNYTSKY, M., KUPKIN, D., AND MEHTA, M. First prototype of conversational h.c. andersen. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces* (New York, NY, USA, 2004), ACM Press, pp. 458–461.

[8] BICKMORE, T., AND CASSELL, J. Small talk and conversational storytelling in embodied interface agents. In *Proceedings of the AAAI Fall Symposium on Narrative Intelligence* (2000), pp. 87–92.

[9] BOS, P., REIDSMA, D., RUTTKAY, Z. M., AND NIJHOLT, A. Interacting with a virtual conductor. In Harper et al. [29], pp. 25–30.

[10] BROWN, P., AND LEVINSON, S. C. *Politeness : Some Universals in Language Usage (Studies in Interactional Sociolinguistics).* Studies in Interactional Sociolinguistics. Cambridge University Press, February 1987.

[11] CASSELL, J., SULLIVAN, J., PREVOST, S., AND CHURCHILL, E. F., Eds. *Embodied conversational agents.* MIT Press, Cambridge, MA, USA, 2000.

[12] CHAFAI, N. E., PELACHAUD, C., PELÉ, D., AND BRETON, G. Gesture expressivity modulations in an eca application. In Gratch et al. [27], pp. 181–192.

[13] CHAO, S. P., CHIU, C. Y., YANG, S. N., AND LIN, T. G. Tai chi synthesizer: a motion synthesis framework based on keypostures and motion instructions. *Computer Animation and Virtual Worlds 15*, 3-4 (2004), 259–268.

[14] CHI, D., COSTA, M., ZHAO, L., AND BADLER, N. The emote model for effort and shape. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000).

[15] CLARK, H. H. *Using Language.* Cambridge University Press, 1996.

[16] COHN, J. Foundations of human centered computing: Facial expression and emotion. In Huang et al. [33], pp. 5–12.

[17] DAUTENHAHN, K. *Socially Intelligent Agents in Human Primate Culture.* In Payr and Trappl [52], 2004, ch. 3, pp. 45–71.

[18] DAVIS, J. W., AND BOBICK, A. F.   Virtual PAT: A virtual personal aerobics trainer. In *Proceedings of Workshop on Perceptual User Interfaces (PUI'98)* (New York, 1998), IEEE, pp. 13–18.

[19] DE ROSIS, F., CAVALLUZZI, A., MAZZOTTA, I., AND NOVELLI, N. Can embodied conversational agents induce empathy in users? In *Proc. of AISB'05 Virtual Social Characters Symposium* (April 2005).

[20] EGGES, A., MOLET, T., AND MAGNENAT-THALMANN, N. Personalised real-time idle motion synthesis. In *PG '04: Proceedings of the Computer Graphics and Applications, 12th Pacific Conference on (PG'04)* (Los Alamitos, CA, USA, 2004), IEEE Computer Society, pp. 121–130.

[21] EKMAN, P. *The argument and evidence about universals in facial expressions of emotion.* John Wiley, Chicelster, 1989, pp. 143–146.

[22] FOGG, B. J. *Persuasive Technology: Using Computers to Change What We Think and Do.* The Morgan Kaufmann Series in Interactive Technologies. Morgan Kaufmann, January 2003.

[23] FRIEDMAN, B., Ed. *Human Values and the Design of Computer Technology.* No. 72 in CSLI Publication Lecture Notes. Cambridge University Press, 1997.

[24] GAVER, W. W., BEAVER, J., AND BENFORD, S.   Ambiguity as a resource for design. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2003), ACM Press, pp. 233–240.

[25] GRATCH, J., AND MARSELLA, S.  Tears and fears: modeling emotions and emotional behaviors in synthetic agents. In *AGENTS '01: Proceedings of the fifth international conference on Autonomous agents* (New York, NY, USA, 2001), ACM Press, pp. 278–285.

[26] GRATCH, J., RICKEL, J., ANDRÉ, E., CASSELL, J., PETAJAN, E., AND BADLER, N. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems 17*, 4 (2002), 54–63.

[27] GRATCH, J., YOUNG, M., AYLETT, R., BALLIN, D., AND OLIVIER, P., Eds. *Intelligent Virtual Agents, 6th International Conference, IVA 2006, Marina Del Rey, CA, USA, August 21-23, 2006, Proceedings* (2006), vol. 4133 of *Lecture Notes in Computer Science*, Springer.

[28] GUSTAFSON, J., BELL, L., BOYE, J., LINDSTRÖM, A., AND WIREN, M. The NICE Fairy-tale Game System. In *Proceedings of SIGdial 04* (April 2004).

[29] HARPER, R., RAUTERBERG, M., AND COMBETTO, M., Eds.   *Proc. of 5th International Conference on Entertainment Computing, Cambridge, UK* (September 2006), no. 4161 in Lecture Notes in Computer Science, Springer Verlag.

[30] HARTMANN, B., MANCINI, M., BUISINE, S., AND PELACHAUD, C.  Design and evaluation of expressive gesture synthesis for embodied conversational agents. In *AAMAS* (2005), F. Dignum, V. Dignum, S. Koenig, S. Kraus, M. P. Singh, and M. Wooldridge, Eds., ACM, pp. 1095–1096.

[31] HARTMANN, B., MANCINI, M., AND PELACHAUD, C.   Implementing expressive gesture synthesis for embodied conversational agents. In *Gesture Workshop* (2005), S. Gibet, N. Courty, and J.-F. Kamp, Eds., vol. 3881 of *Lecture Notes in Computer Science*, Springer, pp. 188–199.

[32] HAYES-ROTH, B., AND DOYLE, P. Animate characters. *Autonomous Agents and Multi-Agent Systems 1*, 2 (1998), 195–230.

[33] HUANG, T., NIJHOLT, A., PANTIC, M., AND PENTLAND, A., Eds.   *Proc. of the IJCAI Workshop on Human Computing, AI4HC07* (January 2007).

[34] ISBISTER, K., AND DOYLE, P.   *The Blind Man and the Elephant Revisited: A Multidisciplinary Approach to Evaluating Conversational Agents.* Vol. 7 of Ruttkay and Pelachaud [62], 2004, ch. 1, pp. 3–26.

[35] Isla, D. A., and Blumberg, B. M. Object persistence for synthetic creatures. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems* (New York, NY, USA, 2002), ACM Press, pp. 1356–1363.

[36] Ju, W., and Leifer, L. The design of implicit interactions - making interactive objects less obnoxious. *Design Issues* (Draft). Draft for Special Issue on Design Research in Interaction Design.

[37] Klein, G., Feltovich, P. J., Bradshaw, J. M., and Woods, D. D. *Common Ground and Coordination in Joint Activity.* Wiley Series in Systems Engineering and Management. John Wiley and sons, Hoboken, New Jersey, 2004, ch. 6, pp. 139–178.

[38] Koda, T., and Maes, P. Agents with faces: the effect of personification. In *5th IEEE International Workshop on Robot and Human Communication* (November 1996), pp. 189=–194.

[39] Kopp, S., Jung, B., Lessmann, N., and Wachsmuth, I. Max - a multimodal assistant in virtual reality construction. *KI 17*, 4 (2003), 11.

[40] Kopp, S., and Wachsmuth, I. Model-based animation of coverbal gesture. In *CA '02: Proceedings of the Computer Animation* (Washington, DC, USA, 2002), IEEE Computer Society, p. 252.

[41] Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. The persona effect: affective impact of animated pedagogical agents. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 1997), ACM Press, pp. 359–366.

[42] Lim, M. Y., Aylett, R., and Jones, C. M. Emergent affective and personality model. In *IVA* (2005), T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist, Eds., no. 3661 in Lecture Notes in Computer Science, Springer, pp. 371–380.

[43] Mateas, M., and Stern, A. Façade: An experiment in building a fully-realized interactive drama, 2003.

[44] McBreen, H., Shade, P., Jack, M. A., and Wyard, P. J. Experimental assessment of the effectiveness of synthetic personae for multi-modal e-retail applications. In *Proceedings of the fourth international conference on Autonomous agents* (2000), pp. 39–45.

[45] Moppes, V. V. Improving the quality of synthesized speech through mark-up of input text with emotions. Master's thesis, VU, Amsterdam, 2002.

[46] Moundridou, M., and Virvou, M. Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning 18*, 3 (2002), 253–261.

[47] Nass, C., Isbister, K., and Lee, E.-J. *Truth is beauty: researching embodied conversational agents.* In Cassell et al. [11], 2000, ch. 13, pp. 374–402.

[48] Nijholt, A. Where computers disappear, virtual humans appear. *Computers & Graphics 28*, 4 (2004), 467–476.

[49] Nischt, M., Prendinger, H., André, E., and Ishizuka, M. Mpml3d: A reactive framework for the multimodal presentation markup language. In Gratch et al. [27], pp. 218–229.

[50] Noot, H., and Ruttkay, Z. M. Gesture in style. In *Gesture-Based Communication in Human-Computer Interaction* (2004), A. Camurri and G. Volpe, Eds., vol. 2915 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 324–337.

[51] Pantic, M., Pentland, A., Nijholt, A., and Huang, T. *Human Computing and Machine Understanding of Human Behaviour: A Survey*, vol. 4451 of *Lecture Notes on Artificial Intelligence, Spec. Vol. AI for Human Computing.* 2007.

[52] PAYR, S., AND TRAPPL, R., Eds. *Agent Culture. Human-Agent Interaction in a Multicultural World.* Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2004.

[53] PLANTEC, P. M., AND KURZWEIL, R. *Virtual Humans.* AMACOM/American Management Association, November 2003.

[54] POGGI, I., PELACHAUD, C., AND DE CAROLIS, B. To display or not to display? towards the architecture of a reflexive agent. In *Proceedings of the 2nd Workshop on Attitude, Personality and Emotions in User-adapted Interaction. User Modeling 2001* (July 2001), pp. 13–17.

[55] PRENDINGER, H., AND ISHIZUKA, M. Social role awareness in animated agents. In *AGENTS '01: Proceedings of the fifth international conference on Autonomous agents* (New York, NY, USA, 2001), ACM Press, pp. 270–277.

[56] REEVES, B., AND NASS, C. *The media equation: how people treat computers, television, and new media like real people and places.* Cambridge University Press, New York, NY, USA, 1996.

[57] REHM, M., AND ANDRÉ, E. Catch me if you can: exploring lying agents in social settings. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems* (New York, NY, USA, 2005), ACM Press, pp. 937–944.

[58] REIDSMA, D., OP DEN AKKER, H. J. A., RIENKS, R., POPPE, R., NIJHOLT, A., HEYLEN, D., AND ZWIERS, J. Virtual meeting rooms: From observation to simulation. *AI & Society, The Journal of Human-Centred Systems to appear* (June 2007).

[59] REIDSMA, D., VAN WELBERGEN, H., POPPE, R., BOS, P., AND NIJHOLT, A. Towards bi-directional dancing interaction. In Harper et al. [29], pp. 1–12.

[60] RIENKS, R., NIJHOLT, A., AND REIDSMA, D. Meetings and meeting support in ambient intelligence. In *Ambient Intelligence, Wireless Networking and Ubiquitous Computing*, T. A. Vasilakos and W. Pedrycz, Eds., Mobile communication series. Artech House, Norwood, MA, USA, 2006, ch. 17, pp. 359–378.

[61] RUTTKAY, Z. M., AND NOOT, H. Animated chartoon faces. In *NPAR '00: Proceedings of the 1st international symposium on Non-photorealistic animation and rendering* (New York, NY, USA, 2000), ACM Press, pp. 91–100.

[62] RUTTKAY, Z. M., AND PELACHAUD, C., Eds. *From Brows to Trust: - Evaluating Embodied Conversational Agents,*, vol. 7 of *Kluwers Human-Computer Interaction Series.* Kluwer Academic Publishers, Dordrecht, 2004.

[63] RUTTKAY, Z. M., PELACHAUD, C., POGGI, I., AND NOOT, H. *Excercises of Syle for Virtual Humans.* Advances in Consciousness Research Series. John Benjamins Publishing Company, to appear.

[64] RUTTKAY, Z. M., REIDSMA, D., AND NIJHOLT, A. Human computing, virtual humans and artificial imperfection. In *ACM SIGCHI Proc. of the ICMI Workshop on Human Computing* (New York, USA, November 2006), F. Quek and Y. Yang, Eds., ACM, pp. 179–184.

[65] RUTTKAY, Z. M., REIDSMA, D., AND NIJHOLT, A. Unexploited dimensions of virtual humans. In Huang et al. [33], pp. 62–69.

[66] RUTTKAY, Z. M., AND VAN WELBERGEN, H. On the timing of gestures of a virtual physiotherapist. In *Proc. of the 3rd Central European Multimedia and Virtual Reality Conference* (November 2006), Pannonian University Press, pp. 219–224.

[67] RUTTKAY, Z. M., ZWIERS, J., VAN WELBERGEN, H., AND REIDSMA, D. Towards a reactive virtual trainer. In *Proc. of the 6th International Conference on Intelligent Virtual Agents, IVA 2006* (Marina del Rey, CA, USA, 2006), vol. 4133 of *LNAI*, Springer, pp. 292–303.

[68] SI, M., MARSELLA, S., AND PYNADATH, D. V. Thespian: Modeling socially normative behavior in a decision-theoretic framework. In Gratch et al. [27], pp. 369–382.

[69] STRONKS, B., NIJHOLT, A., VAN DER VET, P. E., AND HEYLEN, D. Designing for friendship: Becoming friends with your eca. In *Proc. Embodied conversational agents - let's specify and evaluate them!* (Bologna, Italy, 2002), A. Marriott, C. Pelachaud, T. Rist, Z. M. Ruttkay, and H. H. Vilhjálmsson, Eds., pp. 91–96.

[70] THÓRISSON, K. R. *Communicative humanoids: a computational model of psychosocial dialogue skills.* PhD thesis, MIT Media Laboratory, 1996.

[71] VAN VUGT, H. C., KONIJN, E. A., HOORN, J. F., AND VELDHUIS, J. Why fat interface characters are better e-health advisors. In Gratch et al. [27], pp. 1–13.

[72] WACHSMUTH, I., AND KOPP, S. Lifelike gesture synthesis and timing for conversational agents. In *Gesture Workshop* (2001), I. Wachsmuth and T. Sowa, Eds., vol. 2298 of *Lecture Notes in Computer Science*, Springer, pp. 120–133.

[73] WALKER, J. H., SPROULL, L., AND SUBRAMANI, R. Using a human face in an interface. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 1994), ACM Press, pp. 85–91.

[74] WANG, J., DRUCKER, S. M., AGRAWALA, M., AND COHEN, M. F. The cartoon animation filter. *ACM Transactions on Graphics 25*, 3 (2006), 1169–1173.