Embodied Conversational Characters: Wandering between the Digital and the Physical World

Digitale Weltenwanderer: Synthetische Charaktere zwischen Realität und Virtualität

Elisabeth André, Klaus Dorfmüller-Ulhaas, Institut für Informatik, Augsburg,
Thomas Rist, Fachhochschule Augsburg

Summary The objective to develop more human-centered, personalized, and at the same time more entertaining interfaces immediately leads to the metaphor of an embodied conversational agent that employs gestures, mimics, and speech to communicate with the human user. Looking at past and current projects, the current paper discusses an ongoing and manifold evolution of embodied conversational agents from conversational settings with single presenters to interactive performances where the user may participate both as an observer and a presenter. We report on new trends, such as the integration of characters in mixed realities as well as endeavours to endow characters with social behaviors. ►►► **Zusammenfassung** Der Wunsch, die Mensch-Maschine-Kommunikation anthropozentrischer, persönlicher und zugleich unterhaltsamer zu gestalten, führt unmittelbar zur Metapher des personifizierten Interaktionspartners, der dem Systembenutzer in Form einer animierten Figur entgegentritt. Der vorliegende Beitrag skizziert aktuelle Entwicklungen im Bereich virtuelle kommunikative Agenten – angefangen von virtuellen Präsentatoren, die Information als Monolog präsentieren bis hin zu interaktiven Rollenspielen, an denen der Benutzer sowohl als Beobachter als auch als Präsentator teilnehmen kann. Wir berichten über neuere Trends, wie z. B. die Einbettung virtueller Charaktere in die reale Umgebung des Benutzers sowie die ambitionierte Bestrebung, sie mit sozialer Kompetenz auszustatten.

KEYWORDS I.2. [Artificial Intelligence] Lifelike Characters, Augmented Reality, Mixed Reality

1 Introduction

Despite various proposals for new interfaces and interaction paradigms, there is hardly one that matches both at the same time, richness and naturalness of human-human communication. During the last decade research groups as well as a number of commercial software developers have started to deploy embodied conversational characters in the user interface especially in those application areas where a close emulation of multimodal human-human communication is needed. This trend is motivated by a number of supporting arguments. First, virtual characters allow for communication styles common in human-human dialogue and thus can release users unaccustomed to technology from the burden to learn and familiarize with less native interaction techniques. Then, a personification of the interface can contribute to a feeling of trust in the system by removing anonymity from the interaction. Furthermore, well designed characters show great potential for making interfacing with a computer system more enjoyable.

Adopting the distinction between mind and body, incarnations of such characters can differ widely in type and amount of embodiment – starting from simplistic cartoon-style 2D representations of faces, fully embodied virtual humans in 3D virtual worlds to physically embodied androids co-habiting the user's real world. There is also a great variation in terms of its internal or "mental" abilities and skills – among other things, a character's ability to engage in a conversation with humans or possibly other synthetic creatures, and the extent to which it possesses further human-like qualities, such as personality and emotions as well as the ability to form social relationships with their human interaction partners.

By means of selected sample applications, Section 2 of this contribution recalls the yet ongoing development of animated presentation agents starting with TV-style information presenters to highly interactive multi-character scenarios in which information is conveyed to the user in the form of multi-party multi-threaded conversations.

While earlier applications concentrated on interactions between characters and humans in the virtual space, more and more attempts are being made to populate the user's physical environment with virtual characters. Obviously, the fact that the agents are now an integral part of the user's physical environment also affects the mode of interaction. Starting from Milgram's diagram of "Visual Continuity", Section 3 discusses several different ways of providing synthetic agents with a body and integrating them into the physical space. Providing a character with a body, however, is only one side of the medal. Rather, the success of an interface character in terms of user acceptance and interface efficiency very much depends on the character's ability to socially interact with the human user. To this end, Section 4 sketches first attempts to endow embodied conversational agents with social skills.

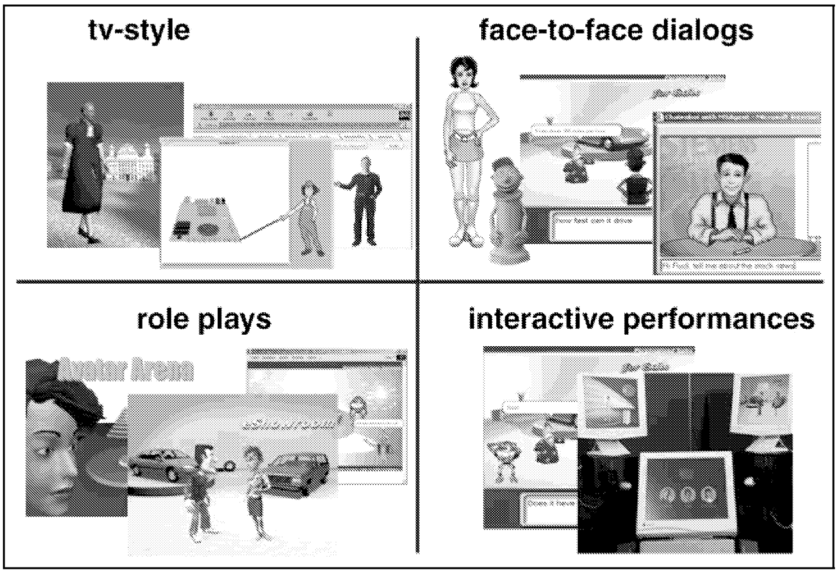

Figure 1 Conversational Settings for Embodied Conversational Characters.

2 Face-to-Face Communication in Virtual Environments

In the area of embodied conversational agents, we can observe an ongoing and manifold evolution as shown in Fig. 1.

2.1 TV Style Presenters

The first class of systems is used in applications in which a single character is deployed to present information. A popular example is the virtual news reader Ananova (www.ananova.com) that reads news scripts live from ITN, a British broadcaster producing news. Further examples include Noma's and Badler's virtual weather reporter [17], DFKI's PPP persona [24] as well as virtual product presenters found on commercial web pages. From the point of view of the user, viewing a presentation appears quite similar to watching a TV-news speaker or to watching a video clip because the flow of information is unidirectional from the agent to the user and no user-agent interaction is foreseen at display time. Therefore, the character's style of presentation is similar to a monologue, though multiple modalities may be deployed.

2.2 Face-to-Face Dialogues

A great number of contemporary systems aim to emulate aspects of a face-to-face dialogue with a character that is able to converse with the user. Differences among systems concern both available input modalities as well as output modalities of the virtual conversation partner.

Quite a number of commercial sites try to boost their web presence by means of virtual sales personnel that provide customers with a more personalized online-shopping experience 24 hours a day, 7 days a week. In most cases the user can "talk" to the character by typing NL expressions into a text-input widget while the character talks to the user either by voice output or likewise through speech bubbles. However, the conversational skills of these characters are often quite limited to that of a pattern-based chat robot that works similar to Weizenbaum's early Eliza system [31]. In the best case, such systems manage to map user input onto related contents available at the web site. The virtual chat agent Cybelle (www.agentland.com/) and the agent Aisa (www.smart.com) are examples of this kind. In the worst case, a conversation with such a character is neither informative nor entertaining. Needless to say that such characters are likely to be perceived by a user as useless if not annoying. In contrast, most research prototypes of embodied conversational characters are instances of complex multimodal dialogue systems, though the focus is usually on the generation of synchronised multimodal expression. Prominent examples include Peedy developed at Microsoft Research [3], the Internet Advisor Cosmo [14], the Steve

Agent (Rickel and Johnson 1999), the real estate agent REA [5], the GRETA Medical Advisor (Pelachaud et al. 2002a), the agent MAX [12] and the animated interface character Smartakus that is developed in the SmartKom project [30]. Most of these systems rely on sophisticated models for multimodal output generation. For instance, Smartakus incorporates a sophisticated spoken dialogue subsystem and has a "visual sense" that enables it to recognize and understand pointing gestures of the user.

2.3 Role Plays and Simulated Conversations

There are situations in which direct agent-user communication is not necessarily the most effective and most convenient way to present information. Inspired by the evolution of TV-commercials over the past 40 years, DFKI group has pioneered role-plays with virtual characters as a promising format for presenting information. A typical commercial of the early days of TV featured a sales person who presented a product by enumerating its positive features – quite similar to what synthetic characters do on web pages today. On TV, however, this format has been almost completely replaced by formats that draw on the concept of short entertaining scenes. Typically, such performances embed product information into a narrative context that involves two or more human actors. Episodic formats offer a much richer basis compared to the plain enumeration of product features, and thus meet the commercial industry's high demand for originality.

André and colleagues [1] suggest the use of presentation teams to convey information about products, such as cars, by performing role plays. Using this presentation style, the user receives information about cars by watching generated sales dialogues among virtual seller and buyer agents. The eShowroom allows the user to specify prior to a presentation (a) the agents' roles, (b) their attitude towards the product, (c) some personality traits (extravert vs. introvert, agreeable vs. not agreeable), and (d) their interests about certain aspects relevant for cars (e.g., the car's relation to prestige, comfort, sportiness, friendliness to the environment, costs etc.). Based on these settings, a variety of different sales dialogues can be generated for the same product.

Using such a setting actually means a shift from a face-to-face character-user setting to a user-as-observer setting. The shift is motivated by a number of supporting arguments: First of all, they enrich the repertoire of modalities to convey information. For example, they allow a system to convey certain rhetorical relationships, such as pros and cons, in a more canonical manner. Furthermore, they can serve as a rhetorical device to reinforce beliefs. For instance, the same piece of information can be repeated in a less monotonous and perhaps more convincing manner simply by employing different agents to convey it. Furthermore, the single members of a presentation team can serve as indices, which help the user to organize the conveyed information. For instance, characters can convey meta-information, such as the origin of information, or they can present information from different points of view, e.g., from the point of view of a businessman or the point of view of a traveller. Last but not least, multiple characters allow us to convey social aspects, such as interpersonal relationships between emotional characters. Work in this direction has been done by Prendinger and Ishizuka [21] and by Rist and Schmitt [25] in their Avatar Arena system.

2.4 Multithreaded Multiparty Conversation

Casting role-plays with characters that can interact with both other characters and the user results in an open multi-party dialogue setting, which supports reactive as well as proactive user participation. One basic idea is to provide the user with the option of taking an active role in the dialogue if she or he wishes to do so. If not, however, the characters will continue the conversation on their own – maybe encouraging the user to give feedback from time to time.

Traum and Rickel [29] have addressed the issue of automatically generated multi-party dialogues in immersive virtual environments. In the context of a military mission rehearsal application, they address dialogue management comprising human-character and character-character dialogues. The characters are based on the Steve architecture [23] which has been enhanced by a multi-modal dialogue model to handle turn taking in such a challenging scenario.

Another multi-party application has been proposed by Isbister and colleagues [10] who concentrate on social interaction between several humans in a video chat environment which is supported by so-called Helper Agent. Helper Agent is an animated, dog-faced avatar that tracks audio from two-person conversations and intervenes if it detects longer silences.

A number of approaches to multiparty conversation have been inspired by research on interactive drama that aims at integrating a user in a scenario - either as an audience member or an active participant. An example includes the interactive installation CrossTalk [8].

CrossTalk has been designed for set-up in a public space, e.g., at a booth of a trade fair. CrossTalk characters live on two separated screens which are spatially arranged as to form a triangle with the user to stimulate a spatially extended interaction experience. CrossTalk takes Brenda Laurel's [13] paradigm of "computers as theatre" a step further and introduces a meta-theatre metaphor to emulate small talk between the characters which then becomes yet another performance, that is "meta-theatre". The purpose for introducing "off-duty" activities is twofold: Firstly it is a means to attract and bind the attention of passers-by, and secondly it gives the agents the authenticity of human-like actors, conveying the impression that they are permanently alive.

Technically speaking the realisation of interactive performances is quite challenging as one has to resolve on an operational level the conflict between predestination and freedom of interaction. Since the scenario is open-ended, neither the characters nor the users are able to tell what exactly may happen next. Instead the dialogue between the user and the characters evolves while time progresses.

3 Face-to-Face Communication in Shared Physical and Digital Environments

Traditional desktop-based applications are characterized by the fact that the user and the character share different realities: the character is tied to the flat screen while the user remains in the physical space. Recently, there has been a trend, however, to integrate synthetic agents in the user's physical space. Obviously, this development has a significant impact on the mode of interaction between humans and synthetic agents. In the following, we provide a number of criteria to characterize the resulting new interfaces.

3.1 The Position of the Character along the Virtuality Continuum

Following [15], we classify the contact between synthetic and human agents according to a "virtuality continuum". At one extreme, we find android agents that are completely integrated in the user's physical world and even allow for physical contact with the user. Mel, a robotic penguin developed by Sidner and colleagues [27] (see Image 1 in Fig. 2), is most likely the most sophisticated physical agent that engages in face-to-face communication with a human user. At the other extreme, there are purely virtual environments that are populated by human and synthetic agents. A prominent example is the pedagogical agent Steve [23] (see Image 4 in Fig. 2). Steve is aware of the user's presence in the virtual space, monitors her actions and responds to them, but has no access to the external world. That is it is only able to perceive user actions that are performed in the virtual space. In between, we find applications in which projections of virtual characters overlay the user's physical environment or projections of real persons are inserted into a virtual world. For instance, Cavazza and colleagues [6] propose a magic mirror paradigm which puts the user both in the role of an actor and a spectator by inserting the user's video image in a virtual world that is populated by synthetic agents (see Image 3 in Fig. 2). The group at Augsburg University focuses on the reciprocal problem, namely how to populate the user's physical environment with synthetic agents. In this application, a synthetic character called Ritchie jointly explores with the user a table-top application that combines virtual buildings of the city centre of Augsburg with a real city map being laid out on a real table. Image 2 in Fig. 2 shows the character Ritchie entering the Multimedia Lab. Mixed Reality applications are characterized by a tight spatial connection between physical and virtual objects. As shown in the image, the real door frame partially covers, for instance, the virtual character which in turn occludes the wall in the background.

3.2 Available Communication Channels

Most research on embodied conversational agents focused on the realization of expressive behaviours in the virtual world. In computer graphics and animation, significant progress has been made on geomet-

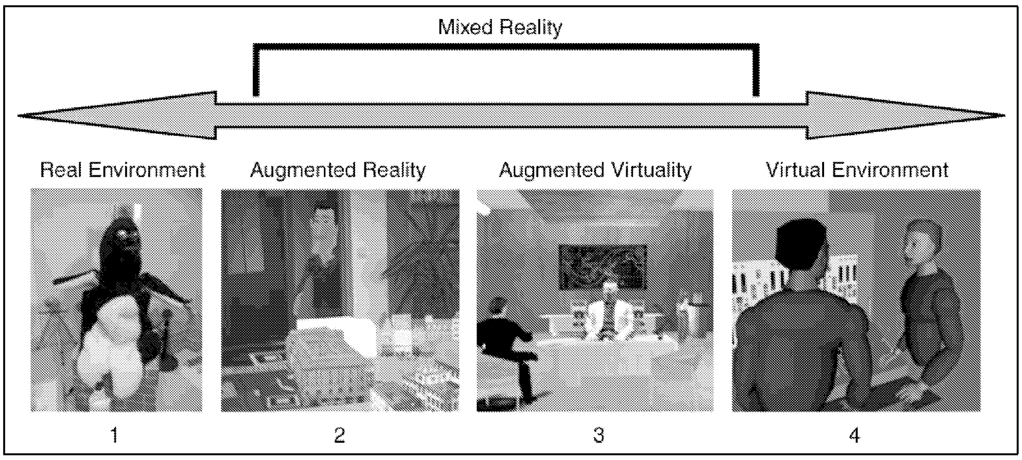

Figure 2 Milgram's Diagram of Virtuality Continuum Adapted to Embodied Conversational Characters: the robotic penguin developed by Sidner and colleagues [27] at MERL (1), the virtual character Ritchie entering the Multimedia Interfaces Lab at Augsburg University (2), Marc Cavazza acting as "Goldfinger" in an Augmented Virtuality application developed by his team at Teaside University [6] (3), the pedagogical agent Steve developed by Rickel and Johnson at ISI [23] (4).

ric face and body modelling and the development of animation and rendering techniques. There is now a broad spectrum of incarnations of characters differing widely in type and amount of embodiment – starting from simplistic cartoon-style 2D representations of faces, fully embodied virtual humans in 3D virtual as well as augmented realities.

The question arises of how to realize expressive behaviours when considering other character-based applications along the Milgram diagram of virtual continuum. The robotic penguin developed by Sidner and colleagues is able to track the face of the conversational partner and adjusts its gaze towards him or her. Even though the set of communicative gestures was strongly limited, an empirical study revealed that users indeed seem to be sensitive to a robot's conversational gestures and establish mutual gaze with it. The realization of communicative behaviours in mixed realities is limited by the fact that the correct projection of occlusions between physical and virtual objects still remains a great challenge. For instance, the Ritchie character is not able to shake hands with the user.

3.3 Amount of Shared Knowledge

Earlier agents, such as the PPP Persona, were usually tied to the desktop and had only limited perceptual abilities to access the user's physical surroundings [2]. The agent's knowledge about the users and their physical environment was primarily based on pre-stored information and simple text-based interactions. More recent projects equip animated agents with a set of sensors to detect and track people in front of the screen. An example is the kiosk agent Mack [28]. Similar to a human information guide, Mack can rely on the physical space shared with a visitor and refer to objects in the physical world, for example, by pointing back with its thumb.

In order to come across as believable, an agent should not just be equipped with a sensory mechanism. Instead it should reflect psychologically plausible attending behaviours that combine locomotion, visual search, and response to peripheral events. For instance, an agent that continuously stares at a target while moving towards it may appear rather unnatural. A first attempt to automate attending behaviours of a synthetic agent has been proposed by Chopra and Badler [7]. Examples of behaviours include spontaneous looking, visual search, and monitoring of events in a 3D environment.

Only a few characters are able to perceive a limited set of nonverbal signals from the user. For example, the SmartKom agent is able to recognize and understand the user's pointing gestures on a display and tries to read a small number of emotional expressions from his or her face [30]. Nakano and colleagues [16] developed a model of grounding for embodied conversational characters. The model has been tested within the Mack agent that analyzes the user's head movements and gaze to establish a common understanding between user and agent of what is being said and meant. Even though considerable progress has been made in the last decade in multimodal dialogue systems, the ability of a character to engage in an unconstrained multimodal conversation will remain a great challenge for years.

3.4 Mobility of the Character

Even though kiosk agents like Mack have an understanding of the user's physical world, they are tied to a fixed location. Mack shares with the user an informational reality, but cannot take on the role of a real companion since it is not able to follow the user if she moves away from the information kiosk.

Several options to allow humans to explore together with an agent a physical space have been explored in the Peach project [26]. In this project, a virtual presenter follows the visitors in a museum and provides advice on both stationary screens as well as the user's mobile device. Unlike stationary agents, mobile agents are able to provide continuous assistance to the user while she explores the physical space. To adapt their presentations to the user's physical context, the Peach agents derive knowledge about the user's location and orientation from different sensors which are connected to the mobile device. Nevertheless, they are still tied to a specific device and cannot move to an arbitrary location in the physical space.

At Augsburg University, we are currently experimenting with different possibilities to provide the user with a virtual companion in the real world. First of all, we support a setting in which the character is superimposed on a scene and faces the user (see Fig. 3). In a multi-user application, this option may create a higher illusion of privacy in the sense that the character is embedded in the user's personal viewing field and exclusively visible to him or her (and not to others). Furthermore, the character maintains continuous visual contact with the user no matter where he or she moves. Technically speaking, different coordinate systems are employed for specifying the user's and the agent's location.

Striving for a higher degree of immersion, we also explore a setting in which both character and user inhabit a common physical space (see Image 2 in Fig. 2). A characteris-

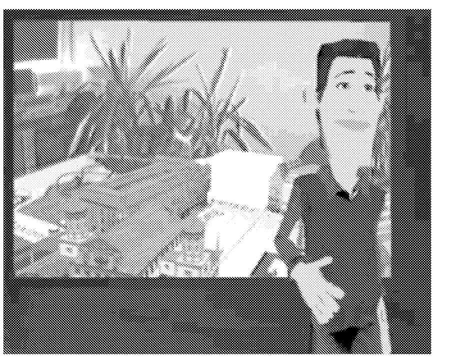

Figure 3 Embedding a Virtual Character as a Foreground Object in the User's Camera Space.

tics of this option is that the user rather has the feeling that the character perceives the scene in a similar way as he or she does and consequently shares experiences with him or her. The approach is, however, technically much more challenging since we have to compute occlusions between virtual and physical objects in real-time.

3.5 Mixed Modes of Interaction

Usually, interaction devices that allow the user to interact with digital objects are significantly different from tools and devices humans employ to manipulate the physical world. To allow for more continuous ways of interaction, the group at Augsburg is experimenting with tangible interfaces for mixed realities. In the Virtual Augsburg application, the user may position a cardboard box in the physical scene to signal the character where it should move (see Fig. 4). That is the user may elicit events in the digital world by the operation of physical devices. The synthetic character has, however, no mechanical effectors and thus is not able to manipulate the cardboard box itself.

Another example of an agent-based interface that allows for the manipulation of physical objects is described in [28]. The Mack agent monitors the user's interaction with a paper map and is even able to make indications on the map itself. Both applications are characterized by the fact that interaction with physical objects has to be restricted to a small set of domain-relevant objects in order to make sure that the characters register the user's manipulation of objects. Furthermore, there is just one representation – either physical or digital – for each object at the same time. In case of multiple representations, a system has to ensure that the user is able to recognize potential correlations between them.

4 Endowing Characters with Social Competencies

The concept of a virtual character promotes the idea that humans, rather than interacting with tools prefer to interact with an artefact that possesses some human-like qualities at least in a large number of application domains. If it is true – as Reeves' and Nass' Media Equation suggests – that people respond to computers as if they were humans [22], then there are good chances that people are also willing to form social relationships with virtual personalities. That is, a virtual character is not just another interface gadget. It may become a companion and even a friend to the user. A prerequisite for this vision to come true is that the virtual characters have a great deal of social competence that manifests itself in a number of different abilities.

One of those is the ability to interpret a user's affective state and react in an empathetic way. Work by Picard [20] and others has created considerable awareness for the role of affect in human computer interaction. Nevertheless, there is a yet ongoing debate whether it is necessary for a computer system to feel emotions as well. When striving for believable virtual personalities, however, the representation of affective state together with a simulation of emotion triggering seems indispensable. In fact, the modelling of affective characters has become a major research trend. Most of the current approaches rely on the so-called OCC model, named after its authors Ortony, Clore, and Collin [18]. The attractiveness of this model to computer scientists probably results from the fact that it defines different types of emotions in terms of emotion triggering conditions which can be easily formalised as rules in a computer-based reasoning system. Most research on the expression of affective behaviors focuses on virtual agents. An exception is the work by Breazeal [4] who developed a robotic agent that integrates a motor system to mimic certain emotive signals.

Furthermore, most current approaches have been designed for the classical face-to-face setting in which one user interacts with one virtual character. When dealing with socially more complex settings, such as multi-party conversations, scalability becomes a big challenge since one has to consider the current social context, too. Surprisingly little attention has been paid so far to the modelling of social context factors in which a conversation takes place. The work by Prendinger and Ishizuka [21] deserves mentioning here. They allow a human script writer to specify social distance and social power relationships among the characters involved in an application, such as a multi-player game scenario.

Another approach has been taken by Rist and Schmitt [25] who

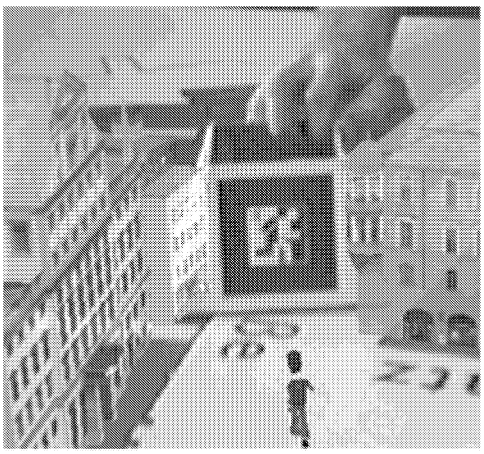

Figure 4 A Tangible Cube Interface to Manipulate Virtual Objects in the Real Space.

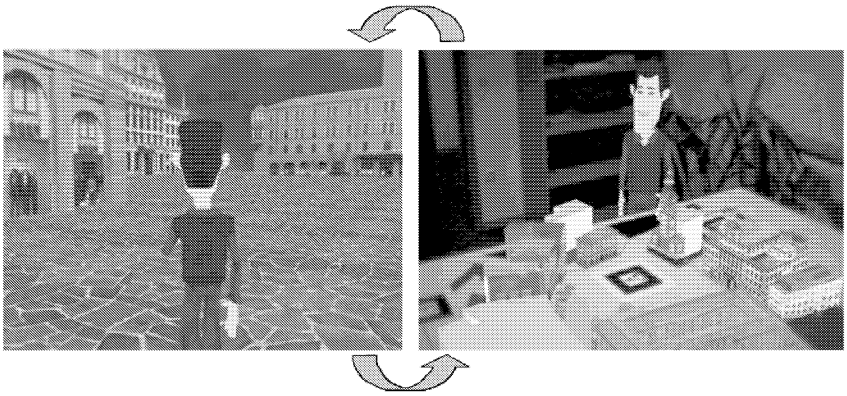

Figure 5 Moving from the Real to the Virtual World and vice versa.

aim at emulating dynamic group phenomena in human-human negotiation dialogues based on socio-psychological theories of cognitive consistency dynamics [19]. To this end, they consider a character's attitudes towards other characters and model a character's social embedding in terms of liking relationships between the character and all other interaction partners. While social behaviors are mainly reflected by the agent's communicative behaviors in the systems described in [21] and [25], Thalmann and colleagues [9] concentrate on the simulation of social navigation behaviours in virtual 3D environments including the social avoidance of collisions, intelligent approach behaviours, and the calculation of suitable interaction distances and angles. Completely new questions arise when dealing with social group dynamics between human and agents in mixed realities. For instance, how should a virtual character react if the user approaches it? Should it move towards the user or rather give way?

5 Conclusions

Research in the area of embodied conversational characters has brought about a great variety of characters that differ widely with respect to embodiment, observable behaviour and communication skills, and internal "cognitive complexity". In this paper, we identified three trends. First of all, there is an evolution from face-to-face communication between a single agent and a single user to multi-party multi-threaded interactions between several human and synthetic interlocutors and observers. Secondly, increasing attempts are being made to integrate embodied conversational agents into the user's natural environment. Users and characters no longer inhabit separated spaces, but share an informational and physical reality. Obviously, shared realities bring new perspectives for humans and synthetic individuals to get in touch with each other. A lot of empirical work is required to decide on the nature of such encounters: Should individuals meet in the user's world, should they get together in the digital world or remain in their own worlds and just exchange information about virtual and physical objects? Another intriguing challenge is the realization of so-called traversable interfaces [11] that allow human and synthetic agents to cross the border from the digital world to the real world and vice versa. An example of a traversable interface is shown in Fig. 5. When moving from one space to another, the user has to correctly correlate different visual representations of one and the same object. A hard problem from the perspective of user interface design is therefore the question of how to make the transfer between different worlds as smooth as possible.

Finally, progress being made in the embodiment of agents also poses high demands on the design of appropriate interactions because the user will expect – at least in part – humanlike verbal and non-verbal conversational behaviours of such an agent. In the long run, it is therefore inevitable to enrich embodied conversational agents with social competencies to be accepted by the human user.

Acknowledgements

Image 1 in Fig. 2 is courtesy of Candy Sidner, MERL. Image 3 in Fig. 2 is courtesy of Fred Charles, University of Teeside. Image 4 in Fig. 2 is courtesy of Lewis Johnson, University of California.

References

[1] E. André, R. Rist, S. van Mulken, M. Klesen, and S. Baldes. The automated design of believable dialogues for animated presentation teams. In J. Cassell, S. Prevost, J. Sullivan, and E. Churchill, editors, *Embodied Conversational Agents*, pp.220–255. The MIT Press, 2000.

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

[3] G. Ball, D. Ling, D. Kurlander, J. Miller, D. Pugh, T. Skelly, A. Stankosky, D. Thiel, M. van Dantzich, and T. Wax. Lifelike computer characters: The persona project at microsoft. In M.J. Bradshaw, editor, *Software Agents*. AAAI/MIT Press, Menlo Park, California, 2000.

[4] C. Breazeal. Affective interaction between humans and robots. In *Proc. of the 2001 European Conf. on Artificial Life*, pp.582–591, 2001.

[5] J. Cassell, T. Bickmore, L. Camphell, H. Vilhjalmsson, and H. Yan. The human conversation as a system framework: Designing embodied conversational agents. In J. Cassell, S. Prevost, J. Sullivan, and E. Churchill, editors, *Embodied Conversational Agents*, pp.29–63. The MIT Press, 2000.

[6] M. Cavazza, O. Martin, F. Charles, S.J. Mead, X. Marichal, and A. Nandi. Multi-modal acting in mixed reality interactive storytelling. *IEEE Multimedia*, 11, 2004.

[7] S. Chopra-Khullar and N. Badler. Where to look? automating attending behaviors of virtual human characters. *Autonomous Agents and Multi-Agent Systems*, 4(1–2):9–23, 2001.
[8] P. Gebhard, M. Kipp, M. Klesen, and Th. Rist. Authoring scenes for adaptive, interactive performances. In *Proc. of the 2nd Int'l Joint Conf. on Autonomous Agents and Multiagent Systems*, pp.725–732. ACM Press, 2003.
[9] A. Guye-Vuillième and D. Thalmann. A high level architecture for believable social agents. *Virtual Reality J.*, 5:95–106, 2001.
[10] K. Isbister, H. Nakanishi, T. Ishida, and C. Nass. Helper agent: Designing an assistant for human-human interaction in a virtual meeting space. In *Proc. of CHI*, pp.57–64. ACM Press, 2000.
[11] B. Koleva, H. Schnadelbach, S. Benford, and C. Greenhalgh. Traversable interfaces. In *Proc. of CHI*, pp.233–240. ACM Press, 2000.
[12] S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth. Max – a multimodal assistant in virtual reality construction. *Künstliche Intelligenz – Special Issue on Embodied Conversational Agents*, 4:11–17, 2003.
[13] B. Laurel. *Computers as Theatre*. Addison-Wesley, 1993.
[14] J. Lester, J.L. Voerman, S.G. Towns, and C.B. Callaway. Deictic believability: Coordinated gesture, locomotion, and speech in lifelike pedagogical agents. *Applied Artificial Intelligence*, 13:383–414, 1999.
[15] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Trans. on Information Systems*, E77-D(12):1321–1329, 1994.
[16] Y.I. Nakano, G. Reinstein, T. Stocky, and J. Cassell. Towards a model of face-to-face grounding. In *Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL 2003)*, pp.553–561, 2003.
[17] T. Noma and N. Badler. A Virtual Human Presenter. In *Proc. of the IJCAI'97 Workshop on Animated Interface Agents: Making them Intelligent*, 1997.
[18] A. Ortony, G.L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.
[19] C. Osgood and P. Tannenbaum. The Principle of Congruity in the Prediction of Attitude Change. *Psychology Review*, 62:42–55, 1955.
[20] R. Picard. *Affective Computing*. The MIT Press, 1997.
[21] H. Prendinger and M. Ishizuka. Social role awareness in animated agents. In *Proc. of the 5th Int'l Conf. on Autonomous Agents*, pp.270–277. ACM Press, 2001.
[22] B. Reeves and C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, 1996.
[23] J. Rickel and W.L. Johnson. Animated agents for procedural training in virtual reality: Perception, cognition and motor control. *Applied Artificial Intelligence*, 13:415–448, 1999.
[24] T. Rist, E. André and J. Müller. Adding animated presentation agents to the interface. In *Proc. of the 2nd Int'l Conf. on Intelligent User Interfaces (IUI)*, pp.79–86. ACM Press, 1997.
[25] T. Rist and M. Schmitt. Applying socio-psychological concepts of cognitive consistency to negotiation dialog scenarios with embodied conversational characters. In *AISB'02 Symp. on Animated Expressive Characters for Social Interactions*, 2002.
[26] C. Rocchi, O. Stock, M. Zancanaro, M. Kruppa, and A. Krüger. The museum visit: Generating seamless personalized presentations on multiple devices. In *Proc. of the 9th Int'l Conf. on Intelligent User Interfaces (IUI)*, pp.316–318. ACM Press, 2004.
[27] C.L. Sidner, C.D. Kidd, C.H. Lee, and N.B. Lesh. Where to look: A study of human-robot engagement. In *Proc. of the 9th Int'l Conf. on Intelligent User Interfaces (IUI)*, pp.78–84. ACM Press, 2004.
[28] T. Stocky and J. Cassell. Shared reality: Spatial intelligence in intuitive user interfaces. In *Proc. of the 7th Int'l Conf. on Intelligent User Interfaces (IUI)*, pp.224–225. ACM Press, 2002.
[29] D. Traum and J. Rickel. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proc. of the 1st Int'l Joint Conf. on Autonomous agents and multiagent systems*, pp.766–773. ACM Press, 2002.
[30] W. Wahlster, N. Reithinger, and A. Blocher. Smartkom: Multimodal communication with a life-like character. In *Proc. of Eurospeech'01*, 2001.
[31] J. Weizenbaum. Eliza – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.


1

2

3

1 Prof. Dr. Elisabeth André is a full professor of Computer Science at the University of Augsburg, Germany, and head of the laboratory for Multimedia Interfaces at Augsburg University. She is the Chair of the ACL Special Interest Group on Multimedia Language Processing (SIGMEDIA). Furthermore, she is on the editorial board of Artificial Intelligence Communications (AICOM), Cognitive Processing (International Quarterly of Cognitive Science), Universal Access to the Information Society (UAIS). She is also the Area Editor for Intelligent User Interfaces of the Electronic Transactions of Artificial Intelligence (ETAI), and a member of the editorial board of Computational Linguistics for the period 2002–2004.
Address: Institut für Informatik, Eichleitnerstraße 30, 86159 Augsburg, Germany,
Tel.: +49-821-5982341, Fax: +49-821-5982349,
E-Mail: andre@informatik.uni-augsburg.de

2 Dr. Klaus Dorfmüller-Ulhaas is a senior researcher at the Lab for Multimedia Concepts and Applications at Augsburg University. He received his PhD in Computer Science from the Vienna University of Technology for his work on: "Optical Tracking – From User Motion To 3D Interaction". His current research interests include: Aug-

mented Reality, Virtual Reality, Computer Vision, Tracking, Projective Geometry, and Camera Calibration.
Address: Institut für Informatik, Eichleitnerstraße 30, 86159 Augsburg, Germany, Tel.: +49-821-5982343, Fax: +49-821-5982341,
E-Mail: klaus@informatik.uni-augsburg.de

3 Prof. Dr. Thomas Rist
is Professor for Interactive Media at the University of Applied Sciences Augsburg and Senior Scientist at the Intelligent User Interfaces Lab of DFKI Saarbrücken. He has served as a program committee member for various international conferences (including among other conferences AVI, AAMAS, IUI, IVA, HCI, UAHCI, CASA, CHI). He was chair of the i3-Spring Days conference series 1999–2001 and has chaired the international conference of Intelligence Virtual Agents 2003.
Address: Fachhochschule Augsburg, Friedberger Straße 2a, 86161 Augsburg, Germany, Tel.: +49-821-5586249, Fax: +49-821-5586160, E-Mail: tr@rz.fh-augsburg.de