Turning Anonymous Members of a Multiagent System into Individuals

Andrea Bönsch∗§, Tom Vierjahn∗§, Ari Shapiro† and Torsten W. Kuhlen∗§
∗Visual Computing Institute, RWTH Aachen University, Germany
† USC Institute for Creative Technologies, USA
§ JARA-HPC, Aachen, Germany

Abstract—It is increasingly common to embed embodied, human-like, virtual agents into immersive virtual environments for either of the two use cases: (1) populating architectural scenes as anonymous members of a crowd and (2) meeting or supporting users as individual, intelligent and conversational agents. However, the new trend towards intelligent cyber physical systems inherently combines both use cases. Thus, we argue for the necessity of multiagent systems consisting of anonymous and autonomous agents, who temporarily turn into intelligent individuals. Besides purely enlivening the scene, each agent can thus be engaged into a situation-dependent interaction by the user, e.g., into a conversation or a joint task. To this end, we devise components for an agent’s behavioral design modeling the transition between an anonymous and an individual agent when a user approaches.

I. INTRODUCTION

Embedding computer-controlled, human-like, virtual agents (VAs) into immersive virtual environments (IVEs) for interaction is challenging: in order to make them believable and authentic characters, the VAs have to show various aspects of human behavior. One crucial, social aspect is their demonstration of visual attention by gazing combined with their respect of personal space, a flexible protective zone that individuals maintain around themselves [13] in real-life [12] as well as in virtual scenarios [1]. Following Equilibrium Theory, the combination of gazing and interpersonal distance, i.e., the distance individuals keep between each other, induces different levels of perceived intimacy by interaction partners in real-life situations [2]. According to Bailenson and colleagues, this theory also applies to human-agent-interactions in IVEs [3].

In order to interact with VAs, a user first needs to approach them. However, depending on the scenario, the approaching phase leads to different reactions by the approached VA. In general, two situations of user-agent-interaction can be distinguished with respect to the agent’s role.

First, VAs can represent the anonymous, however autonomous members of a virtual crowd (VC). These crowds are used for instance to optimize architectural designs (e.g., [9], [14]), to enliven architectural walkthrough scenarios (e.g., [4], [21]), or to conduct pedestrian interaction behavior studies (e.g., [8]). Each of these applications provide different levels of interactivity ranging from pure agent-agent-interaction to limited user-agent-interaction. With the former, the VAs adapt only their walking trajectories in order to avoid collisions, whereas with the latter they additionally make eye contact with an approaching user (e.g., [15], [17]). Here, small interpersonal distances are tolerated since the scenes are typically crowded.

Second, VAs can represent individuals, fulfilling various, situation-dependent roles in a direct and personal interaction with the user. For instance, the VAs are instructors (e.g., [10]), negotiation partners (e.g., [11]) or guides (e.g., [20]). In these settings, approaching an agent commonly triggers a mutual gazing as well as the start of an interaction, e.g., shown by Ólafsson and colleagues for an agent being a user’s dialogue counterpart [18]. Here, designing an agent’s behavior with respect to intimacy, comprising gazing and proxemics [13], is important during both the approaching phase and the following interaction. Indicated by a study that we have conducted, it is beneficial to model an awareness zone around a VA: when a user is entering this zone, the VA should make eye contact with the user as a visual reaction on the user’s presence while meeting the user’s personal space requirements by giving way [7].

To the best of our knowledge, both of the above situations – (1) approaching an anonymous member of a crowd to pass it and (2) approaching an individual to start an interaction – are typically considered separately. Thus, users can only interact with members of a crowd who are explicitly modeled as individuals. Consequently, users have to be specifically informed about these individuals. In games, this is often done by introducing them, or by visually highlighting the individuals, e.g., by markers floating around or above them.

However, in Virtual Reality such techniques have shortcomings: visual clues not being present in reality diminish the users’ feeling of being present in a real-life situation. Additionally, user-agent-interaction restricted to specific agents in crowded scenes results in artificial behavioral patterns of the user. By this, the powerfulness and usefulness of the applications is limited.

Therefore, we argue in this paper that especially for applications representing real-life scenes, all embedded VAs have to support personal user-agent-interaction. Consequently, the VAs’ behavioral design needs to provide techniques for turning anonymous, autonomous members of VCs into intelligent individuals on users demand. To this end, we present two examples of representative scenarios (Section II) emerging from current research efforts. Afterwards, we devise suitable components of an agents behavioral design (Section III). Finally, we outline possible research directories (Section IV).
II. REPRESENTATIVE SCENARIOS

Current strategies by the German Federal Government aim at integrating modern information and communication technology into daily working routines. To this end, cross-linked intelligent cyber physical system (CPS), e.g., self-propelled or stationary robots, shall support human workers in order to improve and accelerate processes. Besides taking over unsafe, time-consuming, repetitive or physically exhausting tasks, the CPSs shall provide information to the workers in order to support them in making informed decisions and solving urgent problems on short notice. Two representative scenarios in these strategies are termed Industry 4.0 and Hospital 4.0.

In order to optimize the value creation process by means of largely self-organizing productions, CPSs shall be integrated into industrial production, referred to as Industry 4.0. This will affect the working routines of the industrial workers with different areas of expertise, the facilities management as well as the suppliers.

In the rehabilitation and health care sector, CPSs shall be integrated in order to optimize the patients’ treatment and care, referred to as Hospital 4.0. This will modify the operation principles of the medical and the nursing staff. Additionally, these developments affect the patients, their accompanying persons and visitors: points of contact for treatments, data acquisitions or questions may become technical interfaces.

Virtual Reality has the potential to support these efforts in the future in several ways, inter alia: (a) by means of virtual prototyping in order to pre-evaluate the planned IT-integration, e.g., regarding users reactions to the presence of and the requirement to team up with autonomous robots. (b) by means of real-time training simulations in order to initially evaluate and then train human workers in the interaction with teams of peers and assistance robots. (c) by means of behavioral economic experiments to investigate, e.g., changes of non-confounded peer effects or incentive effects on competition (e.g., [6]) in the new settings.

The resulting Virtual-Reality-based applications require the scenes to be populated with believable, autonomous, virtual peers and robots in order to adequately mirror real-life situations. Consequently, VCs based on anonymous, however autonomous agents are needed that represent self-reliantly working peers or other persons, e.g., patients and visitors. In particular, the applications must facilitate user-agent-interaction with any of these agents on users demand.

III. BEHAVIORAL DESIGN OF APPROACHED AGENTS

Simulations, as, e.g., required in the context of Industry 4.0 and Hospital 4.0, have to be realistic and convincing to positively effect the human’s performance in corresponding real-life situations. To this end, the virtual scenes have to be, i.e., populated by a crowd of anonymous, autonomous and authentic agents. When there is no direct user-interaction, these agents should walk around or work self-reliantly in the scene. However, the anonymous agents might be temporarily required to be engaged in a user-triggered interaction as individual, intelligent and conversational agents, e.g., as co-workers in an actual task at hand. Consequently, the VAs’ behavioral design needs to provide techniques for turning anonymous, autonomous members of crowds into intelligent individuals on users’ demand.

A simple and straight-forward technique to unambiguously declare which agent is required for an interaction consists of using a ray-casting-based point-and-click metaphor. However, selecting a larger group of agents to engage them in a user-agent interaction may get tedious. Another shortcoming is that this technique might be perceived as artificial in the context of human-like agents and thus might diminish the user’s feeling of being present in a realistic scene.

Commercial implementations commonly use natural speech commands like “Hey, Siri”, “Ok, Google” or “Alexa” in order to trigger an interaction with a conversational agent in form of a bodiless technical interface. However, implementing this direct speech for multiagent systems is challenging: users may not know the names of the agents they plan to involve in an interaction. One solution is attaching name badges to the agents. For instance, clinic staff frequently wears badges, name tags are commonly attached to patients’ beds, while accompanying persons usually stay anonymous. Thus, this method might be perceived as natural for a limited group of agents in a restricted set of scenes. Nevertheless, the badges’ restricted visibility is a drawback: depending on the user-agent-alignment they can be out of sight, while reading them requires a small interpersonal distance between user and agent. However, showing the names as labels, e.g., floating above the agents, as known from many games, might be perceived as being too artificial. Using phrases such as “Hey, you” instead of the agent’s names, is ambiguous if several agents are nearby. In order to clear the ambiguity, an agent in the user’s vicinity may ask in return whether she or he have been addressed. However, to pick this agent more information about the current situation supplementing the phrase is required. Another drawback of the direct speech as well as the aforementioned point-and-click metaphor is the requirement that the users have to explicitly state their requests for an interaction.

To this end, it should be avoided that users are solely responsible for explicitly triggering a user-agent-interaction. Thus, we suggest embedding a selection system that automatically determines candidates for a user-agent-interaction by evaluating a predefined model: based on the user’s behavior and actions in the scene, the model should determine those agents who are likely to be demanded as interaction partners. These candidates are then turned from anonymous agents to individuals. All other agents stay anonymous. The overall concept of this model is illustrated in Figure 1.

Setting up an adequate model is not trivial. However, four aspects of a user’s behavior should be taken into account in order to determine a suitable set of interaction candidates:

First, evaluating the interpersonal distance between user and agents by means of the user’s relative positioning, speed and walking trajectory is reasonable. This is already common in crowd simulations to avoid collisions or to add an authentic gazing between passing agents. Examples are the time-to-collision metric by Karamouzas and colleagues [16] or the minimum predicted distance by Pettré and colleagues [19]. However, a model
One example is the subjective interpretation of the interpersonal distance between user and agent: people tolerate smaller interpersonal distances in narrow passages. In such situations, this measure is thus not well suited to indicate whether the user wants to pass by or to make contact. In contrast, if the interpersonal distance is small in a larger empty area, it is a strong indicator for the desire of making contact.

Another example for environment-dependent weighting is gazing: if only little space is available, focusing on an agent for a longer duration during approaching might be due to observing the agent’s motions to be able to react to sudden movements in order to prevent collisions while passing. Else, it might be a sign of making contact.

Based on environment-dependent weighting factors, one aggregated score \( s \) has to be computed per agent (see Figure 1). By means of this score, the multiagent system then determines whether a VA is an interaction candidate or not. If so, the respective agent is turned from an anonymous member of the crowd into an individual, intelligent and conversational agent. After the interaction ended, indicated by speech or by the user’s departure, or when the agent is not engaged in an interaction, the VA can be turned back into the anonymous member of the crowd.

If the score indicates that the user’s aim is very likely to pass the approached agents, they stay anonymous. Nevertheless, following our recommendation of an awareness zone [7], it seems beneficial if the agents give visible feedback on the user’s presence, e.g., by gazing and giving way. Depending on the desired level of realism, the agents’ actions for the awareness zone may be more or less explicit: a study investigating a VA’s approaching strategies indicated that realistic and human-like behavior do not necessarily have the highest priority [5]. Thus, we expect a clearly visible reaction on the user’s approach to increase the user’s comfort in the IVE, even if that does not resemble a real human’s reaction.

**IV. Discussion**

We have devised a behavioral design for a multiagent system comprising VAs that can temporarily turn from anonymous members of a crowd into intelligent individuals. As triggers for the transitions we suggested implementing a set of weighted metrics evaluating the user’s approach based on, e.g., distance, gazing, gestures and speech.

To implement the outlined behavioral design, more insight into the situation-dependent approaching behavior between humans as well as humans and VAs in IVEs has to be gained. Additionally, detailed investigations on the required metrics used as transition triggers have to be done, taking different environments and numbers of agents into account. We are confident that Virtual-Reality-based multiagent systems will benefit from a behavioral design that combines approaches based on crowds as well as on individual agents.

Having a model yielding a per-agent-score indicating whether a VA is an interaction candidate or not allows for creating VAs with different personal traits: characteristics like shyness or obtrusiveness can be achieved by varying the weighting factors per agent. This will eventually lead to truly individual VAs.
ACKNOWLEDGMENT

This work was funded by the project house ICT Foundations of a Digitized Industry, Economy, and Society at RWTH Aachen University.

REFERENCES


