

Natural interaction with culturally adaptive virtual characters

Felix Kistler, Birgit Endrass, Ionut Damian, Chi Tai Dang, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Kistler, Felix, Birgit Endrass, Ionut Damian, Chi Tai Dang, and Elisabeth André. 2012.
"Natural interaction with culturally adaptive virtual characters." *Journal on Multimodal User Interfaces* 6 (1-2): 39–47. <https://doi.org/10.1007/s12193-011-0087-z>.



Natural interaction with culturally adaptive virtual characters

Felix Kistler · Birgit Endrass · Ionut Damian ·
Chi Tai Dang · Elisabeth André

Abstract Recently, the verbal and non-verbal behavior of virtual characters has become more and more sophisticated due to advances in behavior planning and rendering. Nevertheless, the appearance and behavior of these characters is in most cases based on the cultural background of their designers. Especially in combination with new natural interaction interfaces, there is the risk that characters developed for a particular culture might not find acceptance when being presented to another culture. A few attempts have been made to create characters that reflect a particular cultural background. However, interaction with these characters still remains an awkward experience in particular when it comes to non-verbal interaction. In many cases, human users either have to choose actions from a menu their avatar has to execute or they have to struggle with obtrusive interaction devices. In contrast, our paper combines an approach to the generation of culture-specific behaviors with full body avatar control based on the Kinect sensor. A first study revealed that users are able to easily control an avatar through their body movements and immediately adapt its behavior to the cultural background of the agents they interact with.

Keywords Full body tracking · Virtual agents · Natural interaction · Nonverbal behavior · Gesture recognition · Culture

1 Introduction

When people communicate with each other, they communicate on several channels, such as speech, gestures or postures. While verbal behavior is very explicit, there are other aspects of behavior that are more subtle. Interpersonal distance between interlocutors is an example of a behavioral aspect that is mostly managed subconsciously. Spatial behavior and interpersonal distances are culture-specific. Pease [20] describes several situations that lead to misunderstandings between members of different cultures due to different perceptions of spatial behavior. In one example, he reports on a conference in the US, where a Japanese attendee talked to an American participant:

(...) the two slowly began to move around the room, the American moving backward away from the Japanese and the Japanese gradually moving towards the American. This was an attempt by both the American and the Japanese to adjust to a culturally comfortable distance from each other. (...) Video recordings of this phenomenon replayed at high speed give the impression that both men are dancing around the conference room with the Japanese leading.

This example shows that members of different cultures have a different perception of appropriate interpersonal distance.

In a similar way, virtual characters might not be accepted when interacting in a culturally inappropriate manner. For example, Nass, Isbister and Lee [19] have shown that computer agents representing a user's ethnic group are perceived as socially more attractive and trustworthy.

F. Kistler (✉) · B. Endrass · I. Damian · C.T. Dang · E. André
Human Centered Multimedia, Augsburg University,
Universitätsstr. 6a, 86159 Augsburg, Germany
e-mail: kistler@hcm-lab.de

B. Endrass
e-mail: endrass@hcm-lab.de

I. Damian
e-mail: damian@hcm-lab.de

C.T. Dang
e-mail: dang@hcm-lab.de

E. André
e-mail: andre@hcm-lab.de

In reality, a vast majority of virtual agents are, however, based on a white ethnic background, see, for example, a study by Khan and de Angeli [14], who investigated the demographics of 147 virtual agents. Recently, a few attempts have been made to integrate culture-specific behaviors into virtual agents. However, the interaction with these agents is strongly limited due to a lack of robust recognition technology. Furthermore, this work has relied on little natural interaction devices, such as dancing pads for navigation or Wii-motes to support gesture interaction.

With the release of the Kinect sensor, Microsoft has paved the way to controller-free user interaction. By supporting the analysis of non-verbal behaviors in an unobtrusive manner, the Kinect sensor offers great potential for the simulation of culture-specific human-agent communication.

In the following, we present such an approach. The next section reviews work from the social sciences that has provided the theoretical background for our research. After that, we discuss previous approaches to implement virtual agents that reflect culture through their verbal and non-verbal behavior. We then introduce a recognition framework we developed for the Kinect to analyze how close users stand, how they orient their body and which gestures they perform as an indicator for their cultural background. The recognition framework has been combined with an animation framework to produce customizable agent behaviors on the basis of parameters that have been inspired by cultural theories. We also report on first findings of a study we conducted to investigate whether and how users adapt their body movements to an agent's cultural background.

2 Culture-related interactive behaviors

Culture-related differences manifest themselves on different channels. While most other work focuses on behaviors that are usually expressed in a conscious manner, such as speech, this paper concentrates on nonverbal communicative behavior, such as gestures and postures, but also social navigation behaviors, such as interpersonal distance, formations or body orientations.

2.1 Gestures and postures

A significant amount of work has been spent on improving the expressivity of a character's gestures, see, for example, [22]. To enable a believable culture-specific simulation of interactions, the agents need to be able to display culturally appropriate gestures. Each culture has a specific repertoire of nonverbal communicative behaviors. In high-context cultures, messages and symbols are hard to interpret without consideration of the context, which includes, among other things, a person's social status and background, while in

low-context cultures, symbols and messages are direct and to the point [18]. As a consequence, gestures in high-context cultures are more implicit and more formal compared to gestures in low-context cultures.

Besides different gesture choice, the performance of a gesture can also vary across cultures. While gesturing expressively is considered as a sign of engagement in some cultures, such as behavior is regarded inappropriate in others. In earlier studies [7], we investigated the differences in gestural expressivity and body postures between the German and Japanese culture. Gestures were performed faster, with more power, with a wider space, more fluently and with higher rate of repeated strokes in Germany compared to Japan.

The interpretation of nonverbal behavior can vary with the cultural background of the observer as well. Folding one's arm's in front of the body, for example, might be interpreted as relaxation in one culture, while it could be judged as blocking or even aggression in another culture.

2.2 Formations and interpersonal distances

Culture related differences manifest themselves not only in apparent nonverbal behaviors, such as gestures. Subtle aspects of nonverbal behavior, such as interpersonal distance or body orientation, sometimes play a crucial role of people's perception, sometimes in a subconscious manner. Whenever two or more individuals interact with one another, a formation is developed. Kendon describes this phenomenon with his F-Formation theory [13] according to which, interactants will always orient themselves in a way that enables access to the center of the formation to each of them. Kendon distinguishes between closed and open formations. Closed orientations indicate that the interactants do not wish to be disturbed by other people whereas interactants in open orientations will allow others to enter the conversation. The shape of a formation is defined by the position of its members' bodies, thus strongly related to their interpersonal distance preferences. Hall [9] distinguishes between four different distance zones: intimate, personal, social and public distance. For Northern Americans these zones are: intimate zone up to 0.45 m, personal zone from 0.45 m to 1.2 m, social zone from 1.2 m to 3.6 m and the public zone starts at 3.6 m.

The influence of culture on interpersonal distance has been studied by various researchers. Sussman and Rosenfeld [24] for example studied the influence of culture, language and gender on conversational distance based on Hall's proxemics theory. Their results are particularly strong distinguishing high- and low-contact cultures which is exemplified for the Venezuelan and Japanese cultures. Ferraro [8] found that the average conversational distance for European Americans is approximately 0.5 m. For Latin Americans

this distance can drop down to 0.35 m. Arabian cultures perceive the conversational distance to be as low as 0.22 m. Watson [25] notes that in high-contact cultures interpersonal distances are small, interlocutors face each other directly, often touch one another and speak in a low voice, whereas in low-contact cultures interpersonal distances are greater, interactants face each other more indirectly, speak in louder voices and touching is less usual. Hofstede [10] postulates that members of individualistic cultures such as the US are likely to stand out visually during interaction. Thus, their interpersonal distance should be rather high. Collectivistic cultures, vice versa, are physically very close with in-groups, but reserved with out-groups. Assuming a conversation with friends or family members, the interpersonal distance should be smaller for collectivistic cultures compared to individualistic cultures.

3 Related computational approaches

An early study by Bailenson and colleagues [3] revealed that human users do not treat human-like agents as pure animations, but show different gaze and personal distance behaviors depending on whether they are confronted with human-like agents or nonhuman-like objects. They also found a number of interesting gender differences that are in line with studies of human-human interaction. In particular, they observed that female participants responded in a more sensitive manner to gaze and spatial behaviors of the agents than male participants. Unlike our work, they did not investigate the influence of the agents' culture on human behavior. Nevertheless, their work shows that virtual environments may provide an excellent test bed for studying proxemics behaviors.

The integration of social aspects, such as culture, into the behavioral models of virtual characters has been investigated only lately. In most cases, research involves the user as a passive observer. As a consequence, most researchers concentrate on perception studies in which users are requested to watch scenarios with virtual characters reflecting a particular cultural background [7, 11, 17].

Only few attempts have been made to enable users to interact with individualized virtual characters. One of the difficulties is the availability of appropriate sensor technology to recognize the user's non-verbal behavior. As a work around, a number of systems rely on conventional graphical user interfaces that enable users to select suitable behaviors. An example includes the Tactical Language Training System (TLTS) by Johnson and colleagues [12]. In order to complete the tasks provided by the system, learners have to learn a foreign language. So far, four versions of TLTS have been implemented: Iraqi, Dari, Pashto, and French. Through interaction with the people in the virtual world, the learner is supposed to develop cultural sensitivity and learn how to

communicate appropriately with members of the simulated culture. In addition to a speech-based interface, the system offers the learner GUI menus to select appropriate culture-specific gestures.

In a similar system presented by Wu and colleagues [27], the user interacts with a virtual character by selecting phrases and/or gestures using a PDA. A so-called etiquette engine calculates the appropriateness of the selected action, based on the social relationship between the interlocutors and categorizes the action. This interpretation has an impact on the character's behavior, which might be unexpected for the user. The system does not allow for natural non-verbal communication using body movements, but relies on a graphical user interface both for speech and gestures.

An approach by Aylett and colleagues [2] introduces an educational application called ORIENT. It differs significantly from the systems described above because it uses interaction modalities based on real, physical and tangible objects surrounding the user. Several input devices are provided for interaction with the characters, such as mobile phones and objects with RFID technology. In addition, a Nintendo Wii Remote controller is used to recognize symbolic gestures while a dance pad serves for navigation. The ORIENT system constitutes a major attempt to integrate users as active participants into a culturally sensitive application. Nevertheless, interaction is still rather cumbersome because users have to hold a device in their hands for performing gestures and the navigation using a dance pad is little intuitive because users have to hit specific areas with their feet. As a consequence, the devices enable active user participation, but require some training and thus might distract the user from the actual learning goal.

Instead of different gestures, Rehm and colleagues [23] aim to recognize the user's cultural background by analyzing patterns of gestural expressivity. In particular, they focused on the spatial extent and the speed of gestures as an indicator of the user's culture. The information on the user's gestural activity is used for adapting the behavior of virtual characters to reflect behavior patterns found in the user's culture. Apart from the gestural expressivity, the agents' spatial behavior and the volume of speech are influenced. As in ORIENT, users are, however, not able to freely interact with their full-body without any obtrusive devices. Furthermore, the scenario as such is rather artificial because the interaction is based on gestures alone.

A different approach by Pedica and colleagues [21] describes how positioning and spatial behavior of virtual avatars in a conversation can be automatically generated to extend the unnaturally looking conversations of current agent interactions. Nevertheless, they focus on the influence of the environment and the social context on the automatically generated agent behaviors, but do not take culture into account of their behavioral model or extend the user interaction itself.

Commercially available Kinect applications on the Xbox console mainly include sport and fitness games (e.g. Kinect Sports), racing games (e.g. Kinect Joy Ride), and party and puzzle games (e.g. Game Party in Motion). They all offer some sort of motion or gesture interaction and most of them include a user controlled avatar. Although, the user normally does not navigate the avatar directly through the virtual world. In most games, the avatar moves autonomously and the user can only control aspects of the movement. For example, the avatars in the cars in Kinect Joy Ride automatically accelerate and the user only has to steer left and right with an invisible steering wheel. In Kinect Sports, the user gets, among others, the task to take part in a 100-meters race. Users can therefore influence the speed of their avatars by running without moving in front of the Kinect sensor, but can not turn left or right.

Bleiweiss and colleagues [4] describe an avatar control by combining full body tracking with predefined animations. They only apply parts of the user's skeleton tracking directly on their avatars, while other parts are blended with predefined animations and some parts are even completely replaced by those animations. The predefined animations are triggered according to the current game situation and by recognized user gestures. For example, a sliding animation is applied when the avatar is sliding down a chute or a jumping animation is triggered when a user jump is recognized. In that way, specific animations better fit to the avatar and they can exaggerate the player's motions up to permitting supernatural movements of the avatars.

Our approach distinguishes from previous work by combining full body interaction with virtual characters simulating individualized interactive behaviors. Users are represented by avatars which imitate their body movements and postures. System controlled characters respond to the user by dynamically adapting their behavior depending on their own assumed cultural background.

4 Full body interaction

An important objective of our work is to enable human users to interact with virtual agents in a natural manner focusing on non-verbal behaviors. Technology to achieve this also offers great promise for cultural training applications: Instead of requiring the participants to select the correct behavior (e.g. the appropriate gesture) for their avatars from a menu or to control their avatars by pressing buttons or keys, they may now simply perform the intended actions by themselves.

4.1 Full body tracking with Microsoft Kinect

For tracking the user's body, we make use of the Microsoft Kinect sensor. It basically consists of an infrared (IR) camera and an IR emitter that are used for providing a depth

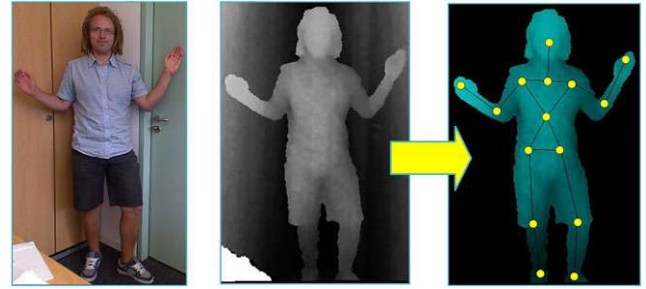


Fig. 1 Kinect: RGB-, depth-, and tracking-image

image in a 640×480 resolution at 30 frames per second. By using the framework “OpenNI” (Open Natural Interaction), the middleware “NITE” (Natural InTEraction)¹ and appropriate drivers,² we connect the Kinect to a PC, and users in front of the Kinect are tracked. The tracker output includes the joint positions of a simplified skeleton (currently 15 joints are supported) fitted on the three-dimensional shape of each user in the depth image. Additionally, this combination of software gives access to the Kinect's RGB camera.

Figure 1 shows a user in front of the Kinect (from left to right): at first in the Kinect's RGB image, secondly in the depth image and at last in the tracking output that includes the user shape (cyan) and the tracked joints (yellow circles) forming the simplified skeleton (black lines). By interpreting the tracked three-dimensional joint positions of the user, we support intuitive interaction without any device in the hand of the user.

4.2 Full body avatar control

The most obvious approach to integrate full body interaction in an application with a virtual world is to apply the positions (or rotations) of the joints tracked of a user to the joints of a virtual avatar. Most currently available Kinect games for the Xbox 360 include that mechanism.

In opposite to those games, we provide a different approach of how the users can navigate their avatar through the virtual scene. Instead of requiring poses, such as leaning forward or walking without moving, we are trying to achieve a more natural way of interaction by requiring the users to move in the real world themselves. However, the virtual navigation space for the avatar would then be limited to the real space in front of the screen. For this reason, we scale up the users' translation for the virtual characters (i.e. one step of the user results in multiple steps of the avatar) to cover a larger virtual navigation space.

To enable more natural movements of the virtual agent, the users' body orientation is applied to the character: If

¹<http://www.openni.org>.

²<https://github.com/avin2/SensorKinect>.

users turn left or right, their avatars will do the same. In this vein, users can orient their avatar towards other agents during a conversation.

In addition, the characters have predefined animations for walking forward, backward, left or right, that are applied according to the orientation and movement. This approach is similar to the one presented by Bleiweiss and colleagues [4]. However, they focus on exaggerating the users' motions and do not map the movement of the whole human body onto the characters.

As the legs of the virtual character are animated separately, we further only need to apply the tracked joint positions of the user's upper body (mainly the arms) to the bones of the virtual agent. We hope that this increases the users' sense of immersion and their identification with the character.

4.3 Full body gesture and posture recognition

In the paragraph above, we have presented an approach to control an avatar by tracking a user's physical movements. However, to allow for meaningful human-agent interactions, human movements have not only been captured, but also to be interpreted in a social context. To this end, we have implemented a toolkit for the recognition of four categories of posture and gestures:

1. *Static postures*: They describe specific relations between the tracked joints, and consequently the configuration of a part of the skeleton. They are directly read from the current joint coordinates in every frame. Examples are "lean forward", "hands above the head", and "arms crossed".
2. *Gestures with linear movement*: They describe a linear movement with a specific direction and speed of one or more joints. They require the calculation of the movement between the frames. Examples are "left hand moves right" and "hands move up".
3. *Combination of postures and linear movement*: They consist of a set of static postures and/or gestures with linear movement of 1 and 2 that are combined according to specific time constraints. This means that each specific gesture or posture has to endure at least for a specific amount of time, and there is also only a limited duration allowed until the next gesture or posture starts. Instead of only one posture or gesture, there can also be required a set of them in parallel which forms one state of a recognition automaton (see Fig. 2). The time constraints and the arrangement of the states result into the transitions of the automaton. In this vein, we can, for example, recognize waving as "hand over shoulder, alternating moving left and right" or walking without moving with "knees alternating up and down".
4. *Complex gestures*: Gestures of this category require a more detailed shape analysis by tracking one or more

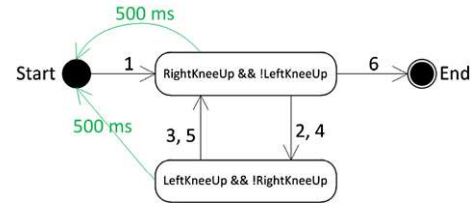


Fig. 2 Recognition automaton for walking without moving

joints (mostly the hands or the shoulder to hand vector) over a certain amount of time. A sequence of frames with the coordinates of the observed joints is processed by a gesture recognition algorithm, such as the Dollar\$1 algorithm [26] for 2D gestures (ignoring one coordinate) or the Protractor3D [16] for real 3D gestures.

For applying one of the recognition algorithms on the gestures of category 4 it is necessary to determine the start and end of a gesture. We currently have different workarounds to achieve the gesture separation, that are:

Interaction Accompanying Postures: For example, the left hand has to be held up while a gesture is performed with the right hand.

Distance Threshold: For example, a gesture lasts as long as the right hand is 30 cm in front of the shoulder.

The "Distance Threshold" does not need additional activities, but is less accurate as "Interaction Accompanying Postures".

The advantages of the gesture and posture types of the categories 1–3 are that they need no user dependent training like category 4, but can easily be defined by the interaction designers in advance. They do not have the problem of the separation, as 1 and 2 can be recognized per frame and 3 can wait until the requirements of the first state of its recognition automaton are fulfilled in one frame. Nevertheless, the completeness of recognizable gestures of category 3 is close to that of category 4. More complex gestures simply need more states in the automaton. Although, the modeling of such a gesture as an automaton can then be significantly more complex like the training of a gesture of category 4, which can also be seen as an automaton with one state per frame.

The recognition of gestures and postures of category 1–3 is implemented in our freely available FUBI framework [1]. An evaluation of this approach revealed that these gestures and postures can be recognized with a high level of accuracy. We recruited eighteen participants that had to perform specific gestures or postures within an interactive story telling environment. Each action was prompted by a symbol on the screen. Our participants successfully performed 97% of all possible actions (i.e. 65 out of 67). One of the unsuccessful actions was caused by a participant that did not remember how to perform it and thus acted out a completely different gesture. Only in one of the 67 actions, the participant knew

how to perform it, but did not succeed in time. More details on this study can be found in [15].

5 Creating responsive individualized virtual agents

In the previous section, we presented a recognition framework that maps body movements onto meaningful gestures and postures. In order to ensure engaging interaction, the virtual agents also need to show a believable behavior that reflects their cultural background in a consistent manner. To accomplish this task, we use the Virtual Beergarden scenario of the Advanced Agent Animation (AAA) application [5]. In the application, system-controlled characters as well as user-avatars can navigate through a virtual Beergarden environment and interact with each other. When two or more characters join an interaction, a formation entity is created. The characters are then members of this formation. This formation entity is defined according to Kendon's F-Formation theory [13]. The structure of a formation entity includes an o-space, a p-space and an r-space (Fig. 3). The o-space represents the formation center, the p-space is the space where the bodies of the members are located whereas the r-space is the space beyond a formation. Every member of a formation will continuously try to satisfy its preferences for interpersonal distance and orientation. This means that the shape of a formation is dynamic and strongly related to its members' preferences. When joining a formation, a character automatically chooses a position inside the formation's p-space and orients itself so that it faces the formation's o-space.

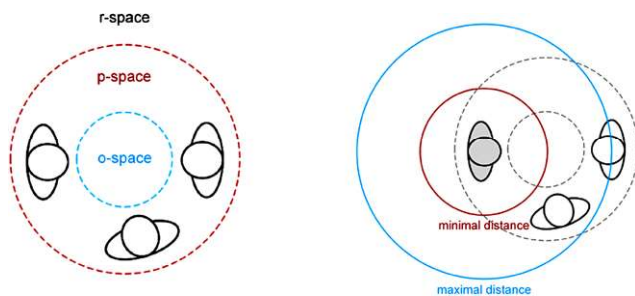
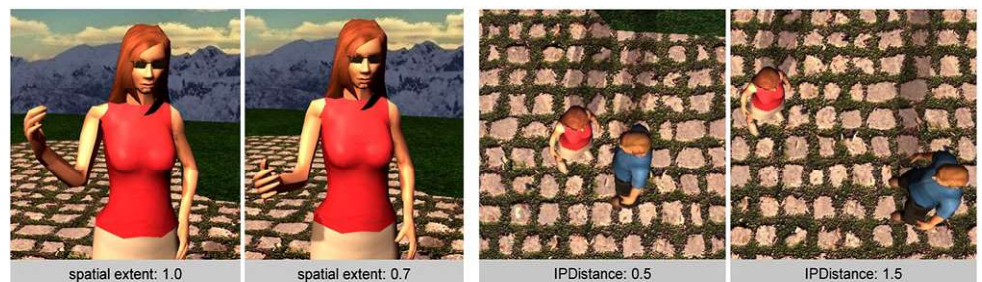


Fig. 3 Structure of a formation entity (*left*) and example of how the interpersonal distance is computed for a member of a formation (*right*)

Fig. 4 Individualized spatial extent (*left*) and interpersonal distance (*right*)



Nonverbal behaviors can be customized in the system along different parameters, taking into account culturally relevant aspects described in Sect. 2. In that manner, the expressivity of a gesture, for example, can be customized by varying the spatial extent of the gesture, playing it faster or slower, or repeating the stroke several times. In a similar manner, movement animations or gaze behavior can be individualized. The spatial behavior of a virtual agent can also be individualized by customizing its behavior within formation. This can be done by manipulating the interpersonal distance and orientation preferences. The interpersonal distance is controlled by each agent's constraints for minimal and maximal distance (Fig. 3) whereas the orientation is defined as a deviation from the normal, towards the o-space, orientation. All these parameters are defined for each character individually thus simulating, for example, the membership to a specific cultural group.

Taking these aspects into account different cultural profiles can be generated that resemble prototypical behaviors of different cultural backgrounds. Figure 4 exemplifies these differences by showing customized gestural spatial extent (left) and different interpersonal distances (right).

6 Evaluation of full body avatar control and different interpersonal distance behaviors

To test our Kinect-based recognition framework in combination with our approach to culture-specific agent parameterization, we conducted a first study with 26 volunteers to investigate (1) how intuitive full body interaction is for users and (2) how they respond to the agents' culture-specific behaviors, while we hypothesize that users prefer agent behavior designed to resemble their own cultural background. For the latter, we focused on culture-specific social navigation behavior and in particular interpersonal distance behavior. Our work extends studies by Bailenson and colleagues [3] who investigated the influence of gender on human proxemics behaviors in virtual environments by studying culture as a novel variable.

To this end, we created three kinds of characters, showing different culture-related interactive behaviors: (1) Neutral virtual agents showing mediate spatial behavior (65 cm

Fig. 5 Virtual characters showing prototypical individualistic (*left*) and collectivistic spatial behavior (*middle*), and the study setup (*right*)



to 1.2 m). (2) Prototypical collectivistic agents that keep a closer personal distance (35 cm to 85 cm). (3) Prototypical individualistic agents that have higher interpersonal distance constraints (95 cm to 1.55 m). Figure 5 exemplifies two groups of virtual characters that show prototypical individualistic (left) and prototypical collectivistic (middle) spatial behavior.

6.1 Evaluation setup and procedure

Users are represented by avatars following the approach described earlier. The avatars are shown from a third person perspective and replicate the users' behavior in real-time. Figure 5 (right) depicts our setup with the user in front of a screen and the Kinect placed right, below it. The screen displays the Virtual Beergarden where the user avatar (woman shown from the back) and two other virtual agents are talking to each other.

The evaluation started with a brief introduction, and after that the participants were allowed to test the full body avatar control. Then, we explained the social setting to the participants and told them to follow a simple script. The user was supposed to have an appointment with two friends in the Virtual Beergarden. When his or her avatar shows up in the Virtual Beergarden, the friends are already there.

As a first step in the script, the user had to perform a waving gesture for greeting the agents from distance and gaining their attention. Once the gesture was recognized by the system, the virtual characters waved back and verbally greeted the participant. As a next step, the user was requested to approach the group of virtual characters in order to join them for a conversation. After a few sentences spoken by the virtual characters, they stated their farewell and walked away, which finished the scenario.

Participants had to perform three test runs to ensure that they got used to the script and were not distracted by factors other than the virtual characters' behaviors. For these practice interactions, we used the neutral agents as described in Sect. 6.1. Afterwards, the actual evaluation study started. Participants had to run through the same script for two more times with the virtual characters that had increased or decreased distance zones, shown in random order. To avoid any bias caused by gender, all character groups consisted of one male and one female character. All characters had a Western

appearance with varying clothing, hair style and eye color randomly generated at each start of the application.

After each interaction in the evaluation study, we asked the participants to fill in a questionnaire addressing two major aspects: (1) the effectance of the interaction and (2) the impression the characters conveyed. All answers had to be given on a 7-point Likert scale with optional comments. To evaluate the effectance, we asked participants whether they felt that they had accomplished their task and whether they understood why the characters were behaving as they did. To evaluate participants' impression of the characters, they had to rate four adjectives: natural, intrusive, open and likeable. After completing the interaction with both groups of characters, participants were asked to fill in a general questionnaire. Besides some demographical questions, we asked how intuitive the interaction with the virtual characters via Kinect was perceived, and asked them to comment on the observed differences between the two virtual character groups. We additionally took video recordings of all interactions.

6.2 Evaluation results

We recruited 26 volunteers from our university campus to participate in the study. The mean age of the participants was 28.12 (SD 5.9), with age ranging from 23 to over 60.

First of all, we investigated whether participants found the interaction with the virtual characters using the Kinect intuitive. A t-test for one sample, testing whether the user rating was significantly above the neutral value of 4, revealed, that our participants thought the interaction was easy with a mean value of 6.231 ($t(25) = 9.959$, $p < 0.0005$). There was also evidence that our users enjoyed the interaction with the Kinect: Most participants started spontaneously playing with the system and tested the navigation and gesture imitation even before we introduced the system to them.

For the effectance ratings, we did not get significant differences between the two conditions (collectivistic and individualistic characters), but we also applied t-tests for one sample that revealed that all ratings were significantly above the neutral value of 4. The participants thought, that they had accomplished their task while interacting with the prototypical collectivistic characters ($M: 6.65$; $t(25) = 18.158$,

$p < 0.0005$) and with the individualistic characters ($M: 6.73; t(25) = 26.100, p < 0.0005$). The participants further indicated, that they understood the characters' reactions with the prototypical collectivistic characters ($M: 6.19; t(25) = 10.953, p < 0.0005$) and with the individualistic characters ($M: 6.12; t(25) = 8.071, p < 0.0005$). These equally positive results in both conditions may be due to the fact that participants had the chance to get familiar with the system during the training phase.

To investigate whether participants rated the characters differently in the two conditions, we applied t-test for paired samples. For the obtrusiveness dimension, we found significant differences ($t(25) = 3.729, p < 0.001$). Thus, participants from Germany perceived virtual characters that had decreased interpersonal distance zones as significantly more obtrusive ($M: 4.08$) than virtual characters with increased zones ($M: 2.42$). We take this as evidence that users consider social distance behaviors that reflect their own culture as more appropriate than social distance behavior that do not. This result is in line with the literature according to which violations of interpersonal distance by stepping too close are described as being pushy or obtrusive, see Sect. 1. We did not achieve significance for the other dimensions. We assume that the participants did not connect these values to the spatial behavior of the characters.

As Bailenson and colleagues [3], we did not inform the participants that we were interested in proxemics behaviors and that these behaviors were varied in the two conditions. Also in the questionnaire, we did not explicitly refer to the agents' proxemics behavior, but only asked the participants whether they found the agents' behavior plausible. Nevertheless, the optional comments of the questionnaires revealed that most participants (16 out of 26) noticed that the agents behavior varied in their interpersonal distance behavior, while 10 participants did not observe any differences.

From the video recordings, we moreover noticed that some participants felt violated in their interpersonal distance by the group with decreased distance and e.g. continuously tried to retreat by stepping back. Therefore, we assume that interpersonal distance behavior between virtual characters and user avatars is perceived in a similar manner as human proximity behavior.

7 Conclusion

In this paper, we presented a human computer interface based on state-of-the-art sensor technology. A novel contribution of our approach is to bring together full body interaction in physical space that supports intuitive behavior (gestures and movements) with the social nature of the virtual agents' culture-specific behavior. For the simulation of different cultural backgrounds, the virtual characters' behavior can be customized taking into account different aspects

of such as gestural expressivity or interpersonal distance behavior.

This approach extends previous work by an interactive setting. In our previous work, culture-related differences were integrated into the behavioral models of virtual characters and evaluated by showing videos of the simulated behavior to human observers. Results of previous studies were promising in terms of user preferences for culture-related behavior that resembles one own's cultural background, in both verbal [6] and nonverbal behavior [7]. In this paper, we introduce our new approach of integrating the user into the virtual scenario and provide reactive virtual characters that give immediate culture-related feedback to the user's behavior.

First evaluation studies of our interactive system revealed that full body interaction can enrich applications with virtual characters. Users found the control of an avatar by their body movement intuitive. Additionally, we showed that gestures and postures can be recognized by our system with a high level of accuracy.

For our demonstrator, we focused on interpersonal distance behavior. Other approaches in that area consider interpersonal gaze behavior as well (see [3]). In our case, gaze behavior was intentionally left aside and the user was faced permanently by body and head orientation of the virtual characters during interaction. To this end, we were able to evaluate distance behavior separately and concentrate on reactive interactive behaviors providing immediate feedback.

Regarding culture-related interpersonal distance behavior, our evaluation study indicates that human users notice cultural differences reflected by agents and respond to it in a promising way. We thus see great potential of controller-free interfaces for cultural training scenarios. They enable users to interact in a natural manner and respond to virtual characters spontaneously without being disturbed by obtrusive interaction devices. For our future work, we aim at investigating additional interactive behavioral aspects recognizable by our system such as culture-related body postures.

Acknowledgements This work was funded by the European Commission within the 7th Framework Program under grant agreement eCute (education in cultural understanding, technologically enhanced).

References

1. Augsburg University (2011) Full Body Interaction Framework: <http://hcm-lab.de/fubi.html>
2. Aylett R, Paiva A, Vannini N, Enz S, André E, Hall L (2009) But that was in another country: agents and intercultural empathy. In: Proceedings of AAMAS 2009
3. Bailenson JN, Blascovich J, Beall AC, Loomis JM (2001) Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence* 10:538–598

4. Bleiweiss A, Eshar D, Kutliroff G, Lerner A, Oshrat Y, Yanai Y (2010) Enhanced interactive gaming by blending full-body tracking and gesture animation. In: ACM SIGGRAPH ASIA Sketches, SA '10. ACM Press, New York
5. Damian I, Endrass B, Huber P, Bee N, André E (2011) Individualizing agent interactions. In: Proc of 4th int conf on motion in games (MIG 2011)
6. Endrass B, Nakano Y, Lipi A, Rehm M, André E (2011) Culture-related topic selection in SmallTalk conversations across Germany and Japan. In: Intelligent virtual agents 2011. Springer, Berlin
7. Endrass B, Rehm M, Lipi AA, Nakano Y, André E (2011) Culture-related differences in aspects of behavior for virtual characters across Germany and Japan. In: Yolum T, Stone S (eds) Proceedings of AAMAS 2011, pp 441–448
8. Ferraro GP (1998) The cultural dimension of international business. Prentice Hall, New York
9. Hall ET (1966) The hidden dimension. Doubleday, New York
10. Hofstede GJ, Pedersen PB, Hofstede G (2002) Exploring culture—exercises, stories and synthetic cultures. Intercultural Press, Yarmouth
11. Jan D, Herrera D, Martinovski B, Novick D, Traum D (2007) A computational model of culture-specific conversational behavior. In: Pelachaud C, et al (eds) Intelligent virtual agents 2007. Springer, Berlin, pp 45–56
12. Johnson WL, Valente A (2008) Tactical language and culture training systems: using artificial intelligence to teach foreign languages and cultures. In: Innovative applications of artificial intelligence (IAAI 2008). AAAI Press, Menlo Park, pp 1632–1639
13. Kendon A (1991) Conducting interaction: patterns of behavior in focused encounters. Cambridge University Press, Cambridge
14. Khan R, Angeli AD (2007) Mapping the demographics of virtual humans. In: Proceedings of the 21st British HCI group annual conference on people and computers: HCI...but not as we know it—Volume 2, BCS-HCI '07. British Computer Society, London
15. Kistler F, Sollfrank D, Bee N, André E (2011) Full body gestures enhancing a game book for interactive story telling. In: Si M, Thue D, André E, Lester J, Tanenbaum J, Zammitto V (eds) Interactive storytelling. Lecture notes in computer science, vol 7069. Springer, Berlin/Heidelberg, pp 207–218
16. Kratz S, Rohs M (2011) Protractor3d: a closed-form solution to rotation-invariant 3d gestures. In: Proceedings of the 16th international conference on Intelligent user interfaces, IUI '11
17. Mascarenhas S, Dias J, Afonso N, Enz S, Paiva A (2009) Using rituals to express cultural differences in synthetic characters. In: Proceedings of AAMAS
18. Mashell's Journal (2012) <http://mashellsjournal.blogspot.com/2009/06/high-and-low-context-culture.html>
19. Nass C, Isbister K, Lee EJ (2000) Embodied conversational agents, chap. Truth is beauty: researching embodied conversational agents. MIT Press, Cambridge, pp 374–402
20. Pease A (1993) Body language: how to read other's thoughts by their gestures. Sheldon Press, London
21. Pedica C, Vilhjálmsón H (2009) Spontaneous avatar behavior for human territoriality. In: Ruttkay Z, Kipp M, Nijholt A, Vilhjálmsón H (eds), Intelligent virtual agents. Lecture notes in computer science. Springer, Berlin/Heidelberg, pp 344–357
22. Pelachaud C (2009) Studies on gesture expressivity for a virtual agent. Speech Commun 51:630–639. doi:10.1016/j.specom.2008.04.009. <http://dl.acm.org/citation.cfm?id=1550956.1551018>
23. Rehm M, Bee N, André E (2008) Wave like an Egyptian—accelerometer based gesture recognition for culture specific interactions. In: HCI 2008 culture, creativity, interaction
24. Sussman NM, Rosenfeld HM (1982) Influence of culture language, and sex on conversational distance. J Pers Soc Psychol 42:66–74
25. Watson O (1970) Proxemic behavior: a cross-cultural study. Mouton De Gruyter, Berlin
26. Wobbrock JO, Wilson AD, Li Y (2007) Gestures without libraries, toolkits or training: a \$1 recognizer for user interface prototypes. In: Proceedings of the 20th annual ACM symposium on User interface software and technology, UIST '07, pp 159–168
27. Wu P, Miller C (2010) Interactive phrasebook conveying culture through etiquette. In: CATS2010 held on ITS2010