

Investigating culture-related aspects of behavior for virtual characters

Birgit Endrass, Elisabeth André, Matthias Rehm, Yukiko Nakano

Angaben zur Veröffentlichung / Publication details:

Endrass, Birgit, Elisabeth André, Matthias Rehm, and Yukiko Nakano. 2013. "Investigating culture-related aspects of behavior for virtual characters." *Autonomous Agents and Multi-Agent Systems* 27 (2): 277–304. <https://doi.org/10.1007/s10458-012-9218-5>.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

Deutsches Urheberrecht

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/>



Investigating culture-related aspects of behavior for virtual characters

Birgit Endrass · Elisabeth André · Matthias Rehm · Yukiko Nakano

Abstract In this paper, culture-related behaviors are investigated on several channels of communication for virtual characters. Prototypical behaviors were formalized in computational models based on a literature review as well as a corpus analysis, exemplifying the German and Japanese cultures. Therefore, aspects of verbal behavior, communication management and nonverbal behavior were taken into account. In evaluation studies conducted in the targeted cultures, each aspect's impact on human observers was tested. With it, we investigated for which of the aspects, observers prefer agent behavior that was designed to resemble their own cultural background.

1 Introduction

Culture plays a crucial role in our lives. Mainly on a subconscious level it influences our behavior, decisions and judgments on everything we encounter.

Language is the most obvious barrier when people from different cultures communicate. But even if both communication partners speak the same language fluently, misunderstandings may occur due to different cultural backgrounds. Human communication includes a lot more than the content of speech, e.g. a vast part of our communication happens nonverbally. In addition, *how* we communicate can sometimes be more crucial than the semantics

B. Endrass (✉) · E. André
Human Centered Multimedia, Augsburg University, Augsburg, Germany
e-mail: endrass@hcm-lab.de

E. André
e-mail: andre@hcm-lab.de

M. Rehm
Department of Architecture, Design and Media Technology, Aalborg University, Aalborg, Denmark

Y. Nakano
Department of Computer and Information Science, Seikei University, Tokyo, Japan

of the communication. An example of such a behavioral difference includes the style in which conversations are managed. In Ref. [34] Ting-Toomey describes that interrupting the conversation partner is perceived as impolite in many cultures such as Germany, while in contrast, it is judged positively as an increased interest in the ongoing conversation in other cultures, such as Hungary. A conversation where two people of contrasting cultures meet, is bound to cause trouble. While one communication partner might feel offended because of being interrupted during the conversation, the other might have the impression that the interlocutor is not interested in the conversation. In either case, a negative impression of the conversation and, in particular, the communication partner is the result. However, as non-experts in inter-cultural communication, interlocutors might not be aware of the reason for their impression, but just feel a general dislike for that particular person. This way, people might be confronted with being refused in another culture without knowing the reason, which in turn can lead to frustration or even culture shock.

To avoid this problem, training sessions can be conducted before traveling to a foreign country by means of role plays with professional actors who behave in a culture-specific way, e.g. [22]. However, this approach is too expensive and time consuming for today's travelers. At this point, computer-based systems come into play. By meeting a different culture in a virtual world, a first contact can be experienced in a safe environment, without the risk of making a fool of oneself or embarrassing a member of the other culture.

Before implementing such a training scenario, it is crucial to build and test the behavioral models of the virtual characters to explore which aspects of behavior actually have an impact on human observers. Otherwise computer-based systems can be misunderstood due to the cultural background of the user as well as the programmer. If, for example, the programmer intends to achieve a certain impression on the user, the user might not necessarily understand this intention and perceive it otherwise.

In this paper, we describe an approach that investigates different aspects of behavior for the German and Japanese cultures. In evaluation studies, conducted in the two targeted cultures, we investigated whether participants prefer agent behavior that was designed to resemble their own cultural background. This is hypothesized due to the similarity principle [5], which states that interaction partners who perceive themselves as being similar are more likely to like each other. With the evaluation studies, we investigate which of the implemented aspects of behavior cause the desired effects and are thus suited to be addressed in a training system with virtual characters.

To measure the impact of each aspect separately, they have to be evaluated in isolation, since different aspects of behavior might correlate, e.g. the semantics of speech and speech accompanying gestures. To overcome these correlations, different studies were set up for the content of dialogs [12] and aspects of communication management as well as nonverbal behaviors [15]. Besides a complete integration of the whole set of behavioral aspects, this paper explores a general workflow that can serve as a guidance for other approaches that investigate culture for virtual characters. In addition, it provides details on the implementation and discusses the contribution of our findings to the research community.

This paper is structured as follows. The next section introduces related work in the domain of enculturated virtual characters. Subsequently, the approach taken for our purpose is outlined that combines a theory-based approach with a empirical corpus study. Section 4 describes culture-related aspects of behavior on different channels for the German and Japanese cultures. Therefore, aspects of verbal behavior, communication management and nonverbal behavior are taken into account. For each behavioral dimension, first the theoretical background is introduced that explains the behavioral aspect and gives information on how it can be structured for integration into a computational model. Second, expectations

for the German and Japanese cultures are stated, based on theories of culture from the social sciences. Finally, building on these expectations, observations made in our video corpus are introduced for each aspect. Section 5 describes the integration of our findings into a multiagent system. Subsequently, we introduce the evaluation studies that were carried out in order to investigate which of the behavioral aspects have an impact on human observers of the studied cultural backgrounds. Finally, this paper is concluded in Sect. 7.

2 Related work

The integration of culture is rather new in the domain of virtual agents. The majority of work in the area focuses on education or training of inter-cultural skills.

Many educational approaches are designed for school environments and aim to create a cultural awareness on a general level using fantasy cultures. These approaches help create inter-cultural understanding by pointing out that behavior is sometimes just different from one's own but not necessarily wrong. Therefore, a user typically enters a virtual world in which he or she interacts with virtual characters of a different culture in a story-like fashion. An example includes the ORIENT system introduced by Aylett et al. [2] where alien characters interact according to their own rules of behavior. Following up on that approach, the eCute project [9] aims at teaching a general cultural awareness, understanding and sensitivity to late primary children and young adults using synthetic cultures and moral circles.

Another group of approaches focuses on training culture-specific competencies on a national level and therefore aim at teaching concrete knowledge and skills for the target culture. Most of the approaches were conducted in the US for military purposes. Examples include the Tactical Language Training System (TLTS) [25] that trains communication skills in foreign languages such as Iraqi Arabic, or the BiLAT system [26] that teaches culture-specific negotiation skills.

Other related work concentrates on research studies to provide a deeper insight in how enculturated virtual characters are perceived by human observers. Examples include work carried out by Koda et al. [27] who investigate the different perception of facial expressions on avatars across cultures, or a study by Yin et al. [36] who investigated a set of aspects such as outer appearance, language or way of argumentation for the Anglo-American and Latin-American cultures.

We consider the work described in this paper as belonging to the third group (research studies), since our aim is to formalize culture-related aspects of behavior and to investigate how human observers of the simulated cultural backgrounds perceive the behaviors. In comparison to studies introduced in the field so far, we evaluate each integrated aspect in isolation to find out which of the aspects improve the acceptance of a virtual character in a given cultural background.

3 Approach

In principle, there are two approaches that might be taken to integrate social aspects such as culture into virtual character systems: top-down and bottom-up. For the top-down approach, definitions of culture and in particular, descriptions of culture-related differences in behavior are extracted from literature and integrated into a multiagent system. Thus, it is a model-driven approach where schemes that are already established in the social sciences are transformed into a computational model. In contrast, the bottom-up approach is data-driven. Human

behavior is analyzed with regard to behavioral differences. Observed tendencies are then integrated into the simulated behavior of virtual characters.

The approach described in this paper combines the two approaches described above. Thus, the social sciences provide us with hypotheses about behavioral differences and a selection of behavioral aspects that are of interest when building a model that describes culture-specific behavior, while from the empirical data we learn how differences in these aspects manifest themselves in a concrete manner.

As a first step of the approach, literature from the social sciences is studied in order to extract culture-related differences in behavior for the targeted cultures. In the social sciences a huge amount of theories and studies are available. However, most descriptions are rather abstract and hard to formalize in computational models. In particular, dimensional models and dichotomies seem to be well suited for our purpose as they take a more descriptive approach to explain culture and thus, provide good guidelines for simulated virtual character behavior.

As a next step, the tendencies extracted from literature of social sciences are grounded in empirical data. Therefore, a video corpus recorded in the two target cultures is analyzed taking into account the expectations on behavioral differences.

Recording such a cross-cultural corpus is a time and resource consuming task [30]. First of all, a huge amount of data is required to explain culture-related differences in behavior. Therefore, at least two cultures are needed in which to record the corpus. Even when focusing on one culture only, observed behavior can be influenced by the context in which the recordings are taken. A cross-cultural study is even more challenging since the cultural aspects need to be singled out from amongst other factors such as context of the recordings or personalities of the participants. In addition, it needs to be assured that the participants in the different cultures are facing the same conditions. Therefore, the setup of the recordings needs to be well chosen to be reproducible elsewhere, while clear scripts need to be constructed and reasonable advice given to the participants. Also technical limitations, such as choice of camera or size of recording area, need to be specified in advance. In addition, the scenario recorded in the corpus needs to be able to grasp cultural differences in behavior, but at the same time needs to be kept general, such that it can be reproduced in all recorded cultures. When designing such a study, developers should also be careful to leave their own cultural background behind, as far as possible, and to avoid making culture-related design heuristics an issue.

Analyzing the large amount of data is a time consuming task. In particular, the annotation of the different behavioral channels takes a lot of time and effort, since e.g. annotators with different language skills are required. In addition, annotation schemes have to be carefully designed to ensure that on the one hand the large amount of data can be summarized in a descriptive manner, and on the other hand the desired aspects of behavior are covered in enough detail to be formalized in a computational model.

For the approach described in this paper, we analyzed the video corpus recorded for the Cube-G project (CULTure-adaptive BEhavior Generation for interactions with embodied conversational agents) [31] in the German and Japanese cultures.

For the video corpus, more than 20 participants were recorded in each culture, each running through three scenarios. Each scenario was recorded with one student interaction partner and one professional actor. The participants did not know that they were interacting with actors. In this vein, we were able to ensure high control over the recordings. On the one hand, we could assure that participants did not know each other in advance. On the other hand, we were able to influence that all scenarios lasted for around the same time. Actors were told to be as passive as possible and to allow the participant to lead the conversation. Only if communication seemed to stagnate, actors should get more active. In that manner, as many topics or ideas as possible were brought up by the participants and not by the actors. To allow



Fig. 1 Screenshots of the video corpus in three prototypical scenarios recorded in Germany

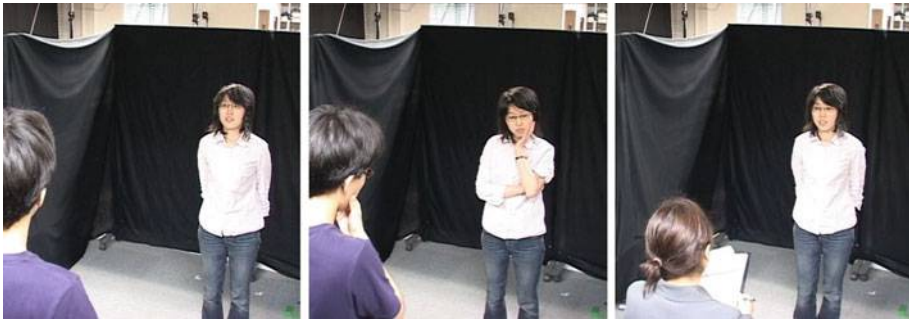


Fig. 2 Screenshots of the videos corpus in three prototypical scenarios recorded in Japan

all gender combinations, we hired four actors: one female and one male actor from each culture. It should be noted that dyads were held in each person's mother tongue and thus Japanese students interacted with Japanese actors and German students with German actors.

For the acquisition of the corpus, three prototypical social interaction scenarios were videotaped: a first-time meeting, a negotiation, and a conversation with someone of a higher social status.

Figures 1 and 2 show example conversations from our recordings in all three scenarios. Each of the participants was told that they would have to solve a task with another student and therefore the two of them had to get acquainted with one another. While having a small talk conversation to get to know each other, the video taping had already started to record the first scenario. For the analysis of culture-related behaviors, described in the subsequent sections, this first-time meeting scenario was investigated due to its prototypical nature.

4 Culture-related behaviors across Germany and Japan

As pointed out in the previous section, in our approach first research literature was reviewed to extract hypotheses about culture-related behavior for the German and Japanese cultures, while a corpus analysis was conducted later based on the hypotheses. Culture can manifest itself on different levels of behavior, while different expressive channels are influenced by culture. Using virtual characters for the simulation of cultural background, different channels

of behavior can be addressed. Verbal behavior can be simulated using text-to-speech systems, while nonverbal behaviors can be naturally demonstrated through the characters embodiment.

In the following subsections aspects of verbal behavior, communication management and nonverbal behavior are taken into account. For each group the following steps are executed: (1) theoretical background from the social sciences is introduced, (2) expectations for the German and Japanese cultures are stated and (3) findings from our corpus study are described.

4.1 Verbal behavior

Regarding verbal behavior, culture can determine the content of a conversation on different levels. On the one hand, there are differences in what people say and on the other hand, there are variations in how people say things. For the choice of “what to say” different topics could be appropriate in certain contexts in some cultures while they are not appropriate in other cultures. Variations in “how to say it” can be found in e.g. different politeness strategies.

4.1.1 Theoretical background

Since speech is a very wide research field, we concentrate on the domain of small talk conversations as a prototypical conversation during a first-time meeting. Small talk is particularly interesting from a culture-specific perspective, since conversations in this domain can differ vastly across cultures but are, at the same time, available in most cultures. According to Isbister et al. [24], the categorization into safe and unsafe topics varies with cultural background. Consequently, a topic such as talking about family members can be considered as safe in one culture and as unsafe in another.

Small talk is often thought of as a neutral, non-task-oriented conversation about safe topics, where no specific goals need to be achieved. But besides being a simple chat, small talk can serve different purposes, such as establishing social relations, getting acquainted with a conversation partner or avoiding undesirable silence.

Although rules for small talk seem to be loose, there are certain structures that explain the flow of an average small talk conversation. In Ref. [33], Schneider describes a sequence that exemplifies the prototypical structure of small talk. This sequence of utterances is characterized as follows: (1) Question, (2) answer, (3) reverse question, understanding, acknowledgment or evaluation, (4) zero or more idle-moves. According to Schneider [33], this sequence is prototypically used to discuss a topic within a small talk conversation, and can be restarted for every other topic.

In addition to defining a prototypical sequence within a small talk conversation, Schneider [33] categorizes topics that, in principle, might occur in small talk conversations:

- (1) The **immediate situation** holds topics that are elements of the so-called *frame* of the situation. In order to explain the idea of a frame, the author of [33] uses a small talk situation that takes place at a party. Possible topics within a party frame could be the atmosphere, drinks, music, participants or food.
- (2) The **external situation** or “supersituation” describes all topics that hold the larger context of the immediate situation. Topics within this category could be the latest news, politics, sports, movies or celebrities.
- (3) For the **communication situation** interlocutors are seen as a subset of the immediate situation. Thus, topics focus the conversation partners e.g. their hobbies, family or career.

According to Schneider [33], a typical small talk conversation begins with the immediate situation and shifts to either the external situation or to the communication situation afterwards, which is dependent on the social surrounding. While shifting to social topics is more

common in a social context, such as a party situation, shifting to private topics is typical for a conversation between strangers that want to avoid silence.

However, Schneider [33] only considered Western cultures in his studies and did not have a look at different topic selection in different cultures. Thus, observations do not necessarily hold true for other cultural groups as well.

4.1.2 Expectations for Germany and Japan

Regarding verbal behavior, Hofstede's dimensional model of culture [21] gives some interesting insights. The individualism dimension of the model describes the degree to which individuals are integrated into a group. On the individualist side, ties between individuals are loose, and everybody is expected to take care for him- or herself. On the collectivist side, people are integrated into strong, cohesive in-groups. Ting-Toomey [34] describes people's self-conception in three layers: the private self, the collective self, and the public self. For individualistic cultures (such as Germany), she states that members are willing to reveal information about the private self in the public layer, while collectivistic cultures (such as Japan) tend to be more guarded regarding information about the private self. Bringing together small talk as a typical public conversation and Schneider's categorization of topics [33], we expect topics covering the communication situation to be more common in the German culture than in the Japanese culture.

Culture-related dichotomies provide two distinctive versions of cultural background in a converse manner. Hall [19] introduces such a dichotomy focusing on different perceptions of temporal regulation and introduces monochronic and polychronic cultures. This distinction describes different perceptions of temporal regulation. Members of different cultures differ regarding how they structure, interpret or understand time.

While Germany counts as a monochronic culture, Japan is a polychronic culture. One behavioral pattern described for monochronic cultures is that members tend to do one thing at a time, while members of the polychronic group prototypically tend to do several things at a time. Generalizing these behavioral patterns, monochronic cultures tend to finish one thing before starting another, while it is more common in polychronic cultures to switch back and forth between tasks. Regarding verbal behavior in the domain of small talk, we expect that topics are more likely to be discussed after one another in monochronic cultures and, thus, in a more sequential manner than in polychronic cultures, where we anticipate switching back and forth between topics as more natural.

In sum, the choice of topics as well as their sequence within a dialog should vary across the two cultures of Germany and Japan. Inspired by definitions of culture and corresponding stereotypical behavior found in literature, we extracted the following expectations about culture-specific differences in small talk behavior distinguishing prototypical German and Japanese conversational behavior:

- (1) Less topics covering the communication situation should occur in Japanese small talk conversations than in German ones.
- (2) Topics are discussed in a more sequential manner in German small talk conversations than in Japanese ones.

4.1.3 Empirical findings

To analyze our video data, topics occurring in the conversations were annotated. In total, 21 German and 11 Japanese videos were investigated, including all gender combinations.

Table 1 Average distribution of topic categories observed in our video data

	Germany (%)	Japan (%)
Immediate situation	18	29
External situation	41	52
Communication situation	41	19

The lower amount of Japanese conversations can be explained by the availability of translated dialogs.

Following Schneider [33], we categorized topics occurring in the conversations into immediate, external or communication situation. For analysis, we build lists of frequency data holding the occurrences of topic categories during the conversations, and compared the two cultures.

Analyzing the distribution of topics across the cultures, we observed topics covering the immediate and external situation more often in the Japanese conversations than in the German ones, while topics covering the communication situation occurred more often in the German conversations. Comparing the choice of topic categories across Germany and Japan, we used the independent two-tailed t test. We found significant differences for all three topic categories [$t(30) = -2.61$, $p = .014$, $r = .43$ for the immediate situation; $t(30) = -2.20$, $p = .036$, $r = .37$ for the external situation, and $t(30) = 2.201$, $p = .035$, $r = .37$ for the communication situation].

This is in line with literature, where it is outlined that in the Japanese culture fewer personal information is provided during first-time meetings. In addition, we found that topics covering the immediate situation and external situation are more common in Japanese conversations.

Using the average percentage of occurrence of each topic category, we obtained a prototypical distribution of topics in prototypical German and Japanese conversations. This average distribution is presented in Table 1.

From our second tendency extracted from literature, we expect topics to be discussed in a more sequential manner in prototypical German small talk conversations compared to Japanese ones.

For our analysis, we counted the amount of topics that arise in each small talk dialog as well as the shifts between topics. Please note, that the same amount of topics does not necessarily mean the same amount of topic shifts and vice versa. In that manner, we were able to compare conversations in terms of how many topics occurred in a conversation in relation to the sequence of their occurrence or reappearance respectively.

Due to the time consuming process of annotation of verbal behavior (Japanese and German speech had to be transcribed and translated to English), we conducted a preliminary analysis first [14] including the first eight available annotated videos from our corpus (four in each culture). In this preliminary analysis results were promising. It showed that in both cultures the exact same number of topics was discussed (26 topics in each video set). Regarding shifts between topics, we found 38 topic shifts in German conversations and 46 topic shifts in Japanese conversations (all videos were approximately 5 min long).

Analyzing the complete data set however, results are less promising and we did not achieve significant results. The tendency is still in line with our expectation, assuming that there should be more frequent topic shifts in the Japanese data set. Looking at the data, we conclude that the preliminary analysis contained the most extreme examples of sequential versus parallel topic occurrences. Still the data give some useful insight into how temporal verbal behavior can be integrated into our computational model. However, since results are not statistically

significant, we did not include culture-related differences in topic shifts in our evaluation study.

4.2 Communication management

Another interesting aspect of human conversational behavior that is considered culture-dependent, are so-called communication management behaviors. As exemplified in the introduction section of this paper, interrupting the conversation partner, for example, is perceived differently across cultures.

4.2.1 Theoretical background

Communication management behaviors are used to control the flow of a conversation and can be defined using so-called regulators. According to Ting-Toomey [34], regulators are consisting of *vocalics*, *kinesics* and *oculesics*:

- Vocalics: This category includes verbal feedback signals, e.g. expressions such as “uh-huh” in the English language. Other features of vocalics are the usage of silence in speech or interruptions of the communication partner’s speech. Depending on how, and how often these behaviors are used, a different rhythm of speech evolves.
- Kinesics: Kinesics include bodily nonverbal behaviors that are used to manage communication such as gestures or body postures.
- Oculesics: This category includes nonverbal behaviors that are conducted with the head or the eyes. Communication can, for example, be managed by gazing at the conversation partner or turning the head away.

According to Ting-Toomey [34], the aforementioned regulators are considered culture-specific behaviors. In addition, regulators are learned at a very young age and are, thus, used at a very low level of awareness. Ting-Toomey [34] also states that the discriminative use of regulators often causes inter-cultural distress and misunderstandings.

4.2.2 Expectations for Germany and Japan

In this paper, verbal regulators are considered as an aspect within culture-related small talk dialogs. In particular, the usage of pauses in speech and overlapping speech are taken into account. According to Ting-Toomey [34], the beliefs expressed in talk and silence are highly dependent on culture. Besides the function of pauses and overlapping speech, their frequency and positioning within the conversation can vary across cultures.

In the previous subsection, we introduced cultural dichotomies. Another cultural dichotomy introduced by Hall [18] is the division *high-context* versus *low-context* communication, focusing on differences in verbal communication patterns. In low-context communication, meaning is expressed through explicit utterances. The speaker is expected to construct clear messages that can be understood easily without the need to decode other aspects of behavior such as silence or tone of voice. In contrast, in high-context communication little information is explicitly encoded and the conversation relies mainly on physical context. Besides verbal utterances, meaning is transported through context or nonverbal clues. Germany is considered a low-context culture, while Japan counts as a high-context culture.

Following this dichotomy [17], Ting-Toomey [34] describes that conversation in high-context communication cultures relies mainly on physical context. Meaning is transported

through additional cues, such as silence or prosody of speech. In contrast, low-context communication cultures tend to explicitly code information. As we stated earlier in this section, Germany is considered a low-context culture, while Japan is a high-context culture. Thus, communication management behaviors should occur more frequently in Japanese conversations than in German ones.

Germany is considered an individualistic culture, while Japan is a collectivistic culture as we pointed out in the previous subsection. In collectivistic cultures, silence may occur in conversations without creating tension [21]. In addition, pauses can be a crucial feature of conversations in collectivistic cultures. These observations do not hold true for individualistic cultures. As a consequence, it should be more likely in the German culture that silence creates tension and, thus, pauses in speech are tried to be avoided. In Japanese conversations, vice versa, pauses can be considered a feature of the conversation and pauses reflect the thoughts of the speaker and can contain strong contextual meaning.

Similar findings are described by Trompenaars and Hampden-Turner [35]. According to the authors, communication in Western cultures is managed as follows: interlocutors start talking after the other conversation partner stopped. In contrast, in Asian cultures, silence is an important feature that can be considered a sign of respect and is more regularly used.

Verbal feedback is given in every culture in principle, but its meaning can vary with the communicative function expressed in the feedback. According to Ting-Toomey [34] in Japanese conversations communication partners explicitly communicate that they are listening by using the utterance “hai hai”, while the literal translation “yes yes” would communicate more than that. Overlapping speech is often considered as impolite in the sense of breaking in on the other person’s speech. However, feedback utterances are often performed without trying to get the turn while the other one is still speaking. As understanding signals are very common in Japanese conversation patterns, we expect an increased amount of overlapping speech caused by verbal feedback in Japanese conversations.

This idea is strengthened by another dimension in Hofstede’s model [21]. The power distance dimension describes the extent to which a different distribution of power is accepted by the less powerful members of a culture. Germany is considered a low-power distance culture, while Japan counts as a high power distance culture. High power distance cultures are described as soft-spoken and polite, while interpersonal synchrony is a lot more important than in low-power distance cultures, whose members tend to talk freely in any social context [34]. One possibility to achieve interpersonal synchrony in a conversation is to give feedback. This feedback often occurs during the discourse of the interlocutor, which should occur more often in the Japanese culture due to their higher value on the power distance dimension and higher need for interpersonal synchrony.

Summarizing our expectations on communication management behavior drawn from literature for the German and Japanese cultures, we state the following expectations:

- (1) More pauses in speech occur in Japanese conversations compared to German ones.
- (2) Overlapping speech occurs more often in Japanese conversations compared to German ones.

4.2.3 Empirical findings

To analyze communication management behaviors in our video data, we calculated the amount of silence (further called pauses in speech) as well as overlapping speech. These behavioral aspects were not annotated explicitly in the video corpus, but calculated from the annotation of the participants’ speech transliterations. Thus, we were able to analyze the

same subset of videos that was analyzed regarding verbal behavior, described earlier in this paper.

Regrading pauses in speech, we considered as a pause the parts of the conversation where none of the conversation partners spoke at a time. Thus, we analysed pauses on a quantitative level and did not consider semantics yet such as pauses during one's own turn and pauses between turn taking. For our analysis, we took into account pauses that lasted for more than one second and more than two seconds. In that manner, we sorted out very brief pauses that are used for breathing, for example. Please note that pauses over 2 s are also included in those that last for more than 1 s.

In line with our expectations, we found more pauses in speech in the Japanese conversations comparing the two cultures. Comparing the number of pauses in speech across the two cultures, using the independent two-tailed t test, we achieved significance for both, pauses that last for more than 1 s [$t(30)$, $p < .001$, $r = .92$] and pauses that last for more than 2 s [$t(30)$, $p < .001$, $r = .85$]. In the German videos, we found on average 6.24 pauses that lasted for more than one second, and 0.52 pauses on average that lasted for more than 2 s. In the Japanese videos, we observed 28 pauses on average that lasted over 1 s, and 9.18 pauses that lasted for more than 2 s.

Regarding overlapping speech, we considered time spans where both conversation partners spoke at the same time. Pragmatics such as the communicative meaning of the overlaps have not been taken into account yet. As stated earlier in this paper, the annotation of verbal behavior (that was used as a basis to calculate communication management behaviors) is very time consuming. Thus, we started with a preliminary analysis of a annotated subset of the video corpus. In this preliminary analysis, eight German and eight Japanese videos were taken into account (see Ref. [13]). We found significant differences between the cultures using the independent t test. In particular, we observed more overlaps in the Japanese conversations compared to the German ones ($t(14) = -1.85$, $p = .043$, $r = .44$). Investigating the full data set, however, we did not achieve any significant results comparing the frequency of overlaps across the two cultures.

4.3 Nonverbal behavior

Regarding nonverbal behavior, the questions asked for verbal behavior “what to say” and “how to say it” can arise as well. While some nonverbal behaviors are common in one culture, they are not necessarily understood in another culture as well. A prototypical example includes common forms of greeting. For example, a bow is used in Japan, a friendly handshake in Germany and a cheek kiss in France.

Besides different kinds of nonverbal behaviors, their performance of a gesture can also vary across cultures. Gesturing a lot and expressively is considered as being engaged in the conversations in Arab countries or Italy. However, it is considered inappropriate in other cultures such as Japan or Sweden. Regarding culture, how we exhibit a gesture can sometimes be more crucial to the observer's perception than the gesture itself.

In this paper, nonverbal behaviors are, thus, classified according to their types as well as to their performance.

4.3.1 Theoretical background

The most well known categorization of gestures has been done by McNeill [28]. Gesture types are described by the following classification:

- Deictic: Deictic gestures are pointing or indicative gestures. They can be either concrete (e.g. pointing at something) or abstract (e.g. pointing to the left and the right to express a contrast).
- Beat: Beat gestures are rhythmic gestures that are often repeated and follow speech prosody.
- Emblem: Emblems have a conventionalized meaning and do not necessarily need to be accompanied by speech. A nod for example is valued as yes in most cultures, or bringing the tips of thumb and first finger together while extending the other fingers is decoded as an “OK”-sign in Northern America as described above.
- Iconic: Iconic gestures explain the semantic content of speech. This is done in a concrete manner such as forming a round shape with the hands while talking about a ball.
- Metaphoric: Metaphoric gestures accompany the semantic content of the speech as well. In contrast to iconic gestures, they visualize abstract concepts by the use of metaphors. Forming a box with both hands, for example, could be used while referring to a previous conversation (visualizing the “conversation”).
- Adaptor: Adaptors are hand movements towards other parts of the body. This is done, for example, to satisfy self or bodily needs or to manage emotions.

Similarly, body postures can be classified. Such a posture coding schema is provided by Bull [4]. The full coding scheme includes several parts of the body that can be used for the description of postures, such as head, arms, trunk or legs. In this paper, arm postures are taken into account. In total, 32 different arm positions are included in the scheme, such as PHEw—put hands on elbow or PHWr—put hands on wrist.

Besides the choice of nonverbal behavior itself, its performance can differ as well and vary vastly with cultural background. Differences in the execution of a nonverbal behavior can be described by the dynamic variation of a gesture.

Gallaher [16] investigated individual differences in nonverbal behavior and introduced so-called dimensions of style. The term ‘style’ is used intentionally to emphasize that also people that are less expressive have a unique nonverbal behavioral style. Following this idea, Hartmann et al. [20] describe six parameters that characterize a gesture’s expressivity. The *spatial extent* describes the arm’s extent toward the torso. The *speed* of the stroke phase of a gesture and the *power* of the arm before the stroke phase can vary as well. The *fluidity* describes the continuity between consecutive gestures, while the *repetivity* holds information about the repetition of the stroke. The last expressivity parameter, *overall activation*, counts the number of gestures that are performed. How gestures are executed can depend on individual and social factors such as personality, emotional state or culture.

4.3.2 Expectations for Germany and Japan

Nonverbal behavior is influenced by culture on several levels. On the one hand, the choice of behavior, e.g. which gesture or body posture to choose, can be influenced by culture. On the other hand, the style in which the nonverbal behavior is performed can vary with cultural background.

Regarding the German and Japanese cultures, nonverbal behavior should differ in both, selection of behavior as well as performance.

Earlier in this paper, we categorized Germany as a low-context culture and Japan as a high-context culture, based on Hall’s dichotomy [18]. In high-context communication little is encoded explicitly and the conversation mainly relies on physical context. Messages and symbols might seem relatively simple but contain a deep meaning. In contrast, low-context

communication explicitly codes information. Thus, symbols and messages are direct and to the point. We therefore expect a more frequent use of direct gestures (deictic and iconic gestures regarding McNeill's classification) in German conversations than in Japanese ones. Vice versa, we expect more metaphoric gestures in Japanese conversations than in German ones. Besides variations in gesture types, other nonverbal behaviors should also differ across the cultures, such as the choice of e.g. greeting style or body postures that are maintained during conversations.

Regarding a gesture's dynamic, it is sometimes more crucial to the observer how a gesture is exhibited than the gesture itself. The individualism dimension in Hofstede's model of culture [21] is related to the expression of emotions and the acceptable emotional display in a culture. In individualistic cultures such as Germany it is more acceptable to publicly show emotions than it is in collectivistic cultures such as Japan [10]. This also suggests that nonverbal behavior is expressed more emotionally in German conversations than in Japanese ones. We expect that displaying emotions more obviously should affect the expressivity in a way that the parameters introduced above, such as speed, power or spatial extent, are increased.

In sum, we expect the following differences in nonverbal behavior for the German and Japanese cultures:

- (1) Different nonverbal behaviors and gesture types should be observed across the cultures.
- (2) Gestures should be performed more expressively in German conversations compared to Japanese ones.

4.3.3 Empirical findings

For the analysis of nonverbal behavior, the whole video corpus was taken into account.

Gestures were annotated according to gesture type and the expressivity parameters. Each expressivity parameter was annotated using a seven-point scale, with one holding a small value and seven a large value for the parameter. With the exception of the parameter repetition, where the value denotes the exact number of repetitions of the stroke of a gesture. For the annotation of different postures, Bull's posture coding scheme was employed for arm postures.

Regarding gesture types, our analysis revealed that the overall number of gestures per minute is comparable in both cultures. For the frequencies of McNeill's gesture types, no statistically significant differences were found in the data. However, when classifying gestures into gestures that carry meaning and adaptors, significant differences were achieved. Our analysis revealed that the Japanese video data contained significantly more adaptors than the German one. This differentiation was suggested by the video material because we observed self-touching hand movements more frequently in the Japanese corpus.

The corpus data also showed differences in the usage of body postures [32]. Putting hands into the pocket occurred most frequently in the German data (40.6 %), followed by folding arms (18.9 %), putting a hand on the elbow (13.9 %), and putting hands behind the back (6.8 %). In the Japanese data set, joining the hands occurred most frequently (29.6 %), followed by putting the hands to the face (20.9 %), putting hands behind the back (14.8 %) and putting a hand to the wrist (12.2 %).

Table 2 summarizes the arm postures frequently observed in the video corpus along with a possible categorization. Interestingly, for the first three categories, postures that regularly occurred in one culture barely occurred in the other culture. The only example constitutes putting the hands behind the back, which was observed in both cultures. However, it was not

Table 2 Posture types frequently observed in the German and Japanese corpus

Category	Germany	Japan
Self-touch	PHIPt (put hands to pocket)	PHFe (put hand to face)
One-handed	PHEw (put hand to elbow)	PHWr (put hand to wrist)
Two-handed	FAs (fold arms)	JHs (join hands)
Hidden	PHB (put hands back)	



Fig. 3 Prototypical body postures identified from the corpus for a German cultural background

covered by the coding schema how the hands are put together behind the back, as provided for body postures that happen with the hands in front of the body. We thus consider self-touching postures as well as one- or two-handed postures best suited to demonstrate culture-related differences. Figures 3 and 4 exemplify these prototypical body postures with German and Japanese participants from the video corpus.

We further investigated differences in gestural expressivity. Comparing the two cultures, we found significant differences for all parameters using ANOVA. Results of this analysis are presented in Table 3. As expected, gestures were performed more expressively by German participants than by Japanese ones. In particular, gestures were performed faster and more powerfully in the German videos than in the Japanese ones. In addition, German participants used wider space for their gestures compared to Japanese participants who used less space. Gestures were also performed more fluently in the German conversations, while the stroke of a gesture was repeated more in the Japanese conversations.

5 Integration into a multiagent system

The culture-related aspects of behavior introduced above were integrated into behavior models for virtual characters taking into account our findings from literature and the corpus study.

There are several possible approaches to determine the behavior of virtual characters, such as plan-based approaches or the usage of finite state machines. Many topical discussions argue



Fig. 4 Prototypical body postures identified from the corpus for a Japanese cultural background

Table 3 Results of gestural expressivity analysis in the German and Japanese data

	Germany	Japan	<i>F</i>	<i>p</i>
Repetition	1.43	1.90	18.264	<0.01
Fluidity	3.96	3.48	68.434	<0.01
Power	3.50	2.75	57.998	<0.01
Speed	4.32	3.33	99.144	<0.01
Spatial extent	3.23	2.67	22.688	<0.01

which of the approaches is best suited for the implementation of virtual character behavior and conversational behavior in particular. Since the aim of our research is to model different culture-related aspects behavior for virtual characters, and not to identify the most appropriate approach for it (if that is possible after all), we do not provide a discussion on the suitability of the technical background that is used for our purpose. Instead, we follow approaches that already implemented solutions to similar problems successfully for other human factors although they have not been applied to culture yet.

For the simulation of our findings, the Virtual Beergarden scenario running in the AAA application [7] is used as a virtual environment. For the simulation of different cultural backgrounds, culture-specific characters were created where outward aspects such as skin type, hair color, shape of the face and clothing were adapted. Agent interaction can happen on two levels: verbally, achieved through text-to-speech, and nonverbally by applying animations such as gestures or body postures.

5.1 Verbal behavior

André et al. [1] introduced an approach that uses planning to generate dialog behavior for virtual characters. Their platform provides several agent components, each containing a behavior planner that controls the performance in the interactive conversation in a highly

dynamic way. The characters' behavior is determined by several attributes such as interests, attitudes or personality traits, which can be specified for each character individually.

Parameterizing virtual character behavior for different personality traits is very much related to our aim of generating culture-related behaviors. In Ref. [1], André et al. dealt with similar problems as the ones that have to be solved for our purpose and provided solutions that can be transferred to the integration of cultural background in an intuitive manner.

Thus, we follow up their work and apply a plan-based approach for the generation of culture-related dialogs. The hierarchical planner SHOP [29] was used for the realization of natural dialog behavior to generate appropriate dialog utterances in small talk conversations. To ensure autonomous behavior, the planner was adapted to a distributed system that generates intentional dialog behavior for each character individually.

Following our findings on differences in verbal behavior for the German and Japanese cultures, two tendencies were realized: (1) Less personal topics occur in Japanese dialogs than in German ones. (2) Topics are discussed in a more sequential manner in German dialogs compared to Japanese ones. Culture-specific rules have therefore been added to each agent's planning component, where topics as well as dialog utterance are selected based on cultural background.

To simulate non-monotonic behaviors, personal motivations were added to each agent's background that, in addition to cultural background, determine behavior selection. This motivation represents the agent's internal drive to talk about the particular topic, e.g. because of an increased personal interest in the topic or a general desire to talk a lot.

For realization of the first tendency, topics were categorized according to our findings from the literature review and corpus study. In that manner, each topic was tagged to either cover the immediate situation, external situation, or communication situation.

To distinguish the two cultures, thresholds were added to the knowledge bases that arise from cultural background and topic category based on observations from the corpus study. Following the results from our corpus analysis, the German threshold for the immediate situation is rather high, compared to the other categories, since this category is the least common for prototypical German conversations. Thresholds for topics covering the external or communication situation are equally high, as observed in our corpus data. For the Japanese culture, the threshold for topics covering the communication situation is highest, since it was least observed in our video data, followed by the immediate situation. Topics covering the external situation were observed frequently in the Japanese data, thus, the threshold is rather low.

Dependent on each character's values in the knowledge base, topics are selected differently. In particular, a higher personal motivation increases an agent's interest to talk about the particular topic, while a higher cultural threshold means that the category of the topic is more difficult to be addressed for the given cultural background.

For the process of topic selection, it is determined whether a topic is appropriate to be introduced, based on the cultural background of the character. A topic is considered appropriate if the agent's personal motivation is higher than the culture-specific threshold for the category of the topic. In that manner, every topic might, in principle, be introduced by every agent. However, as a result from the rules of the planner, it is less likely that a topic is selected by an agent that is not common in its cultural background.

To implement the second tendency, we took as a basis the prototypical small talk sequence described by Schneider [33]. To this end, for each topic four dialog acts were created: (1) question, (2) answer, (3) reverse-question and (4) reverse-answer.

To determine the sequence of a conversation, another culture-specific threshold was added to the knowledge base of each character. The threshold defines how difficult it is in a given

culture to drop a topic and thus interrupt the sequence by introducing a different topic. To simulate the tendency, the threshold is set higher for the Japanese culture, compared to the German one. In that manner, according to the cultural thresholds, the sequence is rather followed by a simulated German cultural background, while it is more likely that the sequence is interrupted by a character with a simulated Japanese cultural background.

5.2 Nonverbal behavior

For the realization of nonverbal behaviors, a Bayesian network was employed. This approach has already been taken successfully by Bergman and Kopp [3] for the generation of gestures taking into account individual patterns of gesture formulation. We consider this approach as equally well suited for the generation of nonverbal behaviors with different cultural patterns.

Dimensional models of culture seem especially well suited for this approach, since the dimensions can simply be represented by different nodes in the network. Mappings from dimensions to stereotypical behavior can be modeled by the transitions to different aspects of nonverbal behavior. Hofstede's dimensional model [21] was used as a basis, since the scores of a large set of cultures on the dimensions are available, as well as descriptions of stereotypical behavior.

Culture-related behavior is a non-deterministic concept. Individuals belonging to the same culture do not all show exactly the same behavior. In a simulated system, this would be very unrealistic as well. Bayesian networks cope well with uncertainty by making assertions about the probability of different performances. To cope with this issue, we may represent the system's uncertain beliefs, for example, by a probability distribution over different levels of a particular aspect of behavior.

The network was implemented using the GeNie modeling environment [8]. We modeled it top-down, from national cultures to the dimensions of culture, and from the dimensions to the aspects of nonverbal behaviors that are influenced. As analysed in our video data, we focused on gestural expressivity along the expressivity dimensions. In addition, the network allows to add other aspects of nonverbal behavior rather easily by following the suggestions from the research literature. This has been exemplified by modeling interpersonal distance and loudness of voice.

For the first layer of the network, Hofstede's scores [21] of the cultures on each dimension are used. To determine the probabilities of the Bayesian network, the scores on the dimensions were classified into three discrete values (low, medium, high). For the second layer, so-called synthetic cultures are used as a basis, which are abstract cultures that represent an extreme end on a dimension. Synthetic cultures help understand the correlation of the cultural dimensions to prototypical behavior, while behavioral tendencies are provided by Hofstede et al. [23] for each dimension separately.

The individualism dimension, for example, has the two opposite sides individualistic and collectivistic. This dimension is highly related to interpersonal distance behavior. It serves as an example here since this behavioral aspect is not strongly affected by other cultural dimensions. According to Hofstede et al. [23] descriptions of synthetic cultures, individualistic cultures tend to stand free in groups, whereas members of collectivistic cultures have a close physical distance, especially to in-groups. Thus, in our network, the distance increases with increasing individualism. Since Germany scores higher on the individualism dimension than Japan [21], the interpersonal distance resulting from our network is higher for a German cultural background compared to a Japanese one.

In some cases, culture-specific behaviors cannot be generated by customizing culture-neutral behaviors because culture is also reflected by specific gestures and postures that need



Fig. 5 Prototypical body postures for virtual characters simulating a German cultural background

to be accurately executed. To account for this fact, customizable behaviors were augmented by behaviors extracted directly from the corpus.

To exemplify the direct extraction of nonverbal behaviors, body postures were remodeled from the video corpus to enrich our animation set. As pointed out earlier, six body postures occurred regularly in the video corpus that seem to be very well suited to show differences between the two cultures. Hence, three stereotypical body postures were designed for each culture. Figure 5 shows screen shots of the prototypical German body postures displayed by stereotypical German virtual characters, while Fig. 6 shows the prototypical postures for the Japanese culture displayed by characters that simulate a Japanese cultural background.

6 Evaluation of behavioral aspects

A key problem is to identify which of the implemented aspects of behavior have an impact on human observers and thus, influence their perception, and which have not. To measure the impact of each aspect separately, they have to be evaluated in isolation. This is not a trivial task, since different aspects of behavior might correlate, e.g. the semantics of speech and speech accompanying gestures. To set up appropriate evaluation studies, these correlations have to be overcome. Thus, we were not able to evaluate verbal and nonverbal behavior in a single study and had to set up individual studies separating the different aspects.

We set up two different studies, one that investigates the content of the dialog (verbal behavior) [12] and another that investigates its form (communication management and nonverbal



Fig. 6 Prototypical body postures for virtual characters simulating a Japanese cultural background

behavior) [15]. The verbal study did not contain variations in nonverbal behavior such as gestures, while the nonverbal study contained Gibberish as a substitute for real speech to be able to integrate communication management behaviors.

For both studies, two different versions were created, one to be conducted in Germany and one to be conducted in Japan. For the study conducted in Germany, Western-looking characters were used, while for the study conducted in Japan, we used Asian-looking characters (see Fig. 7). In addition, we used language-specific text-to-speech systems for the Western and Asian characters (German and Japanese). In this vein, we assured that participants did not assume a cultural background different from their own. To avoid side effects evoked by gender, we chose a mixed gender combination for the agent conversations. The virtual characters in the studies interact in a dyadic conversation, showing an open conversational formation opened towards the camera. One behavioral aspect was studied at a time. All other aspects as well as aspects that were not included in our evaluation, such as interpersonal distance or loudness of voice, were left constant during the studies.

In each study, we hypothesized that human observers prefer agent behavior that is in line with observations made for their own cultural background. Thus, we stated the following two hypothesis for each behavioral aspect:

- H1 : German participants prefer the videos showing prototypical German dialog behavior over the Japanese versions.
- H2 : Japanese participants prefer the videos showing prototypical Japanese dialog behavior over the German versions.



Fig. 7 Screenshots of virtual character interaction as displayed during our evaluation studies in Germany (*left*) and Japan (*right*).

6.1 Content

In the evaluation study described in this subsection, we investigate culture-related differences in the content of small talk dialogs. In particular, topic selection was taken into account. In line with literature, we found in our video corpus topics covering the immediate and external situation significantly more often in the Japanese conversations compared to the German ones, while topics covering the communication situation occurred significantly more often in the German conversations.

6.1.1 Study design

For the study, six English dialogs were created, three of them containing a prototypical German topic distribution and three of them containing a prototypical Japanese topic distribution. Based on the observed distribution in our video corpus, we equally integrated the immediate and external situation into the prototypical German small talk dialogs and all three categories into prototypical Japanese conversations, with an emphasis on the external situation. In that manner, we integrated two topics covering the external situation and two topics covering the communication situation in the German dialogs, and two topics covering the external situation, one topic covering the immediate situation and one topic covering the communication situation in the Japanese dialogs, while all dialogs lasted for approximately one minute. Topics within the dialogs were carefully chosen in agreement with our Japanese cooperation partners of the Cube-G project [31] to assure dialogs were principally appropriate in both cultures. This is particularly important, since topics that are safe in one culture are not necessarily safe in another culture [24]. The dialogs were later translated into the German and Japanese languages for the two evaluation studies in order to avoid effects due to the language barrier or culture-specific assumptions that might have been made for English-speaking characters.

Participants watched the videos in pairs, each containing a prototypical German and a prototypical Japanese conversation in alternating order. For each pair of videos they had to judge

- (Q1) which one is more appropriate,
- (Q2) which one is more interesting,
- (Q3) which conversation they would prefer to join and

(Q4) which pair of agents gets along with each other better,

and participants were able to either choose one of the videos or a button, indicating that none of the two videos was preferred. In addition, a comment box was provided that allowed participants to state an opinion on their choice.

6.1.2 Results

With the evaluation study, we wanted to find out whether human observers prefer agent conversations that reflect their own cultural background.

6.1.2.1 German study In the German evaluation study, 16 participants took part, six females and ten males, in an age range of 23–40 years. Since all participants observed three pairs of videos, we obtained a data set containing 48 judgments. For analysis, we conducted a chi-squared (χ^2) goodness-of-fit test.

Our results indicate that German participants significantly prefer videos with agent conversations that reflect prototypical German topic selection for all four questions. Table 4 summarizes the results from the German evaluation study. Thus, participants found German conversations more appropriate and interesting, would rather like to join the conversations and think that agents get along with each other better.

In line with our expectations, five out of ten participants who had explained their choice in the comment box, stated that they preferred the selected conversation because it was more personal and revealed more information about the interlocutors.

6.1.2.2 Japanese study In the Japanese evaluation study, 14 people participated, seven females and seven males, in an age range of 21–23 years. We thus obtained a data set containing 42 judgments. As for the German study, we conducted a chi-squared (χ^2) goodness-of-fit test. Our analysis revealed that the Japanese versions of small talk conversations were significantly preferred by Japanese participants for three out of the four questions. In Table 5, the results from the Japanese evaluation study are summarized. According to our study, Japanese participants found the Japanese versions of small talk conversations more appropriate and interesting and thought that agents were getting along with each other better. However, the study did not significantly indicate that participants would also rather like to join the Japanese conversations over the German ones. However, a trend into that direction can be observed.

Interestingly, and in line with our expectations, some Japanese participants showed that the immediate situation was of importance for them. For example, a Japanese video that included talking about the weather was judged positively by a participant because it “fit to the background image”, while another participant disliked a conversation since “the content in the video does not match to the background image”.

Table 4 Results from the perception study on content conducted in Germany [12]

Germany	German dialog	Japanese dialog	None	χ^2	df	p
Q1	33	5	10	27.875	2	<.001
Q2	37	4	7	41.625	2	<.001
Q3	34	4	10	31.5	2	<.001
Q4	28	7	13	14.625	2	.001

Table 5 Results from the perception study on content conducted in Japan [12]

Japan	German dialog	Japanese dialog	None	χ^2	df	<i>p</i>
Q1	11	22	9	7	2	.03
Q2	12	25	5	14.714	2	.001
Q3	11	21	10	5.286	2	.071
Q4	12	23	7	9.571	2	.008

6.1.3 Discussion

Regarding our evaluation study, the integration of different topic selections in culture-related small talk dialogs for virtual characters was very promising. Findings from literature were confirmed by our empirical study and integrated into the behavior of virtual characters. Our hypothesis that human observers prefer agent dialogs that were designed to reflect their own cultural background was confirmed. In particular, our results revealed that both German and Japanese participants preferred agent behavior that was in line with observations made for their own cultural background.

We thus claim that in applications where small talk is intended for virtual characters, e.g. to establish a positive personal relation with a human user, the integration of culture by selecting topics that are appropriate for the target cultural background can help improve the character's acceptance and should thus be taken into account by the designers of such a system.

A quite large number of participants did not decide for either of the conversations (see Tables 4 and 5). Looking at the comments, only two participants commented on their choice for none of the dialogs. One participant stated that the conversations were too short to make a decision; another stated that both conversations were quite interesting. From the second statement, we assume that our intended study design was appropriate in a way that our culture-specific conversations are feasible in both cultures and not constructed in a way that they only work in one culture.

6.2 Form

In this section, we describe our evaluation study that investigates aspects of communication management as well as nonverbal behavior for the two cultures of Germany and Japan. These aspects of behavior can be tested in a common study, since the correlation to verbal behavior needs to be carefully avoided for both aspects.

6.2.1 Study design

To avoid content in dialogs, we used Gibberish, a fantasy language that represents a language without any specific meaning of the words. To this end, words were generated that have the same statistical distribution of syllables as the words from the target language. In addition, we used language-specific text-to-speech systems to match the prosody of speech in the target culture. To get participants acquainted with the situation of listening to a Gibberish dialog, we showed a neutral conversation first.

In the study, each behavioral aspect is tested in isolation. Interrelations between communication management and nonverbal behavior themselves can be excluded by simply not

performing them during the cases where the other aspect is focused on. Regarding gestural expressivity, we implemented the two aspects of gestural speed and spatial extent since they seemed promising in a virtual character system. The other aspects were not evaluated yet, since power is much related to the speed parameter, fluidity is hard to judge when displayed by virtual characters and repetition seems to be dependent on the content of speech. In total, six pairs of videos were shown in random order, each lasting for half a minute and containing variations in one of the following aspects of behavior, based on our literature research and corpus analysis.

- *Pauses in speech* Taking into account our corpus findings, the simulated dialogs reflecting typical Japanese conversations contain more pauses compared to the German version. German agent dialogs are designed to contain one pause that lasts for one second, while Japanese dialogs contain two pauses that last one second and one pause that lasts for two seconds.
- *Overlapping speech* Following our findings in overlapping speech, we integrated one overlap that lasts for 0.3 s and two overlaps that last for 0.5 s into the German dialog. The Japanese dialog contains three overlaps that last for 0.3 s, one that lasts for 0.5 s and one that lasts for one second.
- *Communication management* Videos reflecting different communication management behaviors contain both, pauses and overlaps in speech, as described above.
- *Speed of gestures* Our observations revealed that in the German corpus, gestures are performed faster than in the Japanese one. In the simulated dialogs, we thus customized the speed of gestures accordingly.
- *Spatial extent of gestures* Based on our findings, simulated gestures have a smaller spatial extent in the Japanese conversations compared to the German ones.
- *Postures* In the agent conversations, characters remained in a prototypical German (put hands to pockets, fold arms) or Japanese (join hands, put hand to wrist) body posture as observed in the video corpus.

In the study, participants stated their preference by rating on a six-graded scale. The scale contained three grades on each side, from “rather this video” over “this video” to “by any means this video”, while only one option could be selected by the participants. Thus, deciding for one video includes a non-preference for the other video.

6.2.2 Results

In the evaluation study, participants had to decide which of the videos they liked better, assuming that participants prefer behavior that was designed to reflect their own cultural background.

6.2.2.1 German study In the German evaluation study, 15 participants took part (six females and nine males) in an age range of 21–45, while with one exception all participants were students. In a chi-squared (χ^2) goodness-of-fit test, we tested, whether the perception of the observed pattern of behavior differs statistically.

Table 6 summarizes the results from the evaluation study conducted in Germany. Our results reveal that significantly more than 50% of the German participants had a preference for the German version of overlapping speech and spatial extent in gestures.

For the behavioral aspects pauses in speech, communication management and posture, we almost achieved significance, suggesting that German participants prefer the German dialog versions over the Japanese ones.

Table 6 Results from the perception study on form conducted in Germany [15]

Germany	German dialog	Japanese dialog	χ^2	df	<i>p</i>
Pause	11	4	3.267	1	.071
Overlap	13	2	8.067	1	.005
Communication management	11	4	3.267	1	.071
Speed	10	5	1.667	1	.197
Spatial extent	13	2	8.067	1	.005
Posture	11	4	3.267	1	.071

Table 7 Results from the perception study on form conducted in Japan [15]

Japan	German dialog	Japanese dialog	χ^2	df	<i>p</i>
Pause	11	6	1.471	1	.225
Overlap	12	5	2.882	1	.09
Communication management	7	9	0.25	1	.617
Speed	5	10	1.667	1	.197
Spatial extent	7	7	0	1	1
Posture	4	13	4.765	1	.029

For gestural speed, we did not get significant results. However, in line with our expectations, at least by trend, German participants showed a preference for the videos simulating prototypical German speed of gestures.

6.2.2.2 Japanese study In the Japanese evaluation study, 17 people participated (three females and 14 males) in an age range of 20–24, while all participants were students. For statistical analysis, a chi-squared (χ^2) goodness-of-fit test was conducted.

Table 7 summarizes the results from the evaluation study conducted in Japan. Results are less distinct compared to the German study.

Significantly more than 50 % of the Japanese participants had a preference for the version with Japanese posture behavior. For other behavioral patterns, we cannot claim any statistically significant evidence.

Although not significant, communication management and gestural speed of the Japanese version was preferred by the Japanese participants by trend.

The results for pauses in speech and overlapping speech were not as expected, as participants preferred the German videos over the Japanese ones, however not significantly.

6.2.3 Discussion

For the study, several aspects of human behavior were simulated with virtual characters and evaluated in isolation. Our results suggest that some of the aspects have a stronger impact on the perception of human observers than others.

Regarding communication management behaviors, pauses in speech and overlapping speech were considered. We found some trends suggesting that German participants favored the German dialog versions. The results suggest that in line with the research literature, by trend more pauses or more overlaps were liked less by German participants. For the Japanese

evaluation study, results indicate that also the German version was preferred, although not significant. This might be due to the missing semantics of the agent dialogs since we used Gibberish. The use of pauses and overlaps in the Japanese language can be closely related to the semantics of speech and is acceptable in one case and unacceptable in another. Thus, without having the necessary semantic clues at hand, Japanese participants might have been tempted to go for the “safe” solution and vote for the version with less pauses and overlaps.

For the integration of differences in nonverbal behavior, gestural expressivity was investigated by means of speed and spatial extent as well as body postures. Regarding gestural speed, our results indicate that by trend German participants preferred the faster version, while Japanese participants preferred the slower version. This suggests that the speed of gestures can be a helpful aspect when integrating cultural background. For the spatial extent of a gesture, we found a significant preference for wider gestures in the German study, while we did not find any trend in the Japanese study. Since gestures were designed in Germany, it is possible that lowering the spatial extent might not have the desired effect in other cultures. Regarding body postures, German participants preferred the German version by trend, while Japanese participants preferred the Japanese versions. We therefore think that considering culture-specific body postures can help increase a character’s acceptance.

7 Conclusions

Within the scope of this paper, different aspects of culture-specific behaviors were investigated. Based on a literature review as well as a corpus study, prototypical behaviors were extracted for the German and Japanese cultures. Therefore, literature from the social sciences was reviewed to describe culture and corresponding prototypical behavior. Theories from the social sciences are a good guideline to formulate hypotheses but are often too abstract to build computational models. In order to obtain statistical data on culture-related behavior, we analyzed a video corpus recorded in the target cultures with regards to our findings from literature. Simulating these behavioral trends with virtual characters, we evaluated which of the behavioral aspects have an impact on the perception of human observers of the target cultural backgrounds. Separate evaluation studies were designed for the different behavioral channels to exclude correlations between different aspects of behavior, to be able to test each of them in isolation. In particular, we investigated whether observers prefer behavior that was designed to resemble their own cultural background. With the studies, we gained a deeper insight into which of the aspects can help enhance a characters acceptance.

Particularly, for those aspects where our hypotheses were confirmed, we consider their integration into the behavioral models of virtual characters with regards to cultural background as promising. We, therefore, think that designers of virtual characters can use our findings to adapt relevant aspects of behavior to the simulated cultural background. Hence, they can serve as a guideline on for related research, such as inter-cultural training.

As suggested by Cassell [6], findings from evaluation studies can serve as an input to refine the behavioral models of virtual characters. Thus, she presents a development cycle that can be iterated several times. In particular, our results on communication management behaviors across cultures did not reveal the intended results, as Japanese participants preferred the prototypical German version by trend. This points out a failure in our study that already occurred early in our approach. In the analysis of our empirical data, we only investigated the quantitative numbers of pauses in speech and overlapping speech. However, the communicative meaning of these behaviors should play a crucial role, e.g. regarding the difference of overlapping speech used as feedback behavior compared to the ‘barge-in’. This problem

should have been recognized at an earlier point in time, since our corpus analysis did not reveal significant results in this category. This issue should be taken as a warning for future research. After our experiments, an analysis on the semantics of speech was conducted by comparing the usage of understanding utterances (acknowledging that interlocutors understand each other). In line with the research literature, we found more understanding utterances in the Japanese corpus, which suggests a culture-related difference in feedback behavior. In our future work, we will consider these utterances in correlation with overlapping speech in order to get a better understanding on this behavioral category on a qualitative level.

Another issue that might have influenced our results in the form study is the rather low number of participants and the fact that only one pair of videos was recorded for each condition. Adding more videos in a follow up study can give a deeper insight into the users' preferences. In addition, recording several communicative situations for each aspect can help avoid the influence of the content of speech.

The method used in this paper combines the two approaches usually taken when integrating human factors into computational systems: top-down (theory-driven) and bottom-up (corpus-driven). In that manner, the social sciences are taken as a basis to formulate hypotheses about behavioral differences, while a corpus provides statistical data on the behaviors. The combined approach proves to be very well suited for the aim of integrating cultural background for virtual characters, as purely theory-driven approaches, for example based on dimensional models, sometimes cannot produce natural behavior for any given cultural background, while purely corpus-driven approaches are not general enough to be transferable to social situations differing from the context of the recordings.

In this paper, the integration of cultural background was exemplified for the German and Japanese cultures. The workflow constructed for this approach, however, is of a general nature and can be transferred to other cultural backgrounds. To exemplify the general nature of the approach, a study was carried out on the US American and Arab cultures for aspects of communication management behaviors [11]. In this study, we had a closer look at feedback behavior, rather than pure overlapping speech. Following our workflow, we started with a literature research to build cultural profiles and identify prototypical behavioral tendencies. As a next step, the expectations were grounded into empirical data using a speaker and listener scenario to obtain statistical insights into the target behavioral aspects for the two cultures. Findings were further integrated into our demonstrator to simulate prototypical culture-related behaviors with virtual characters. Our study design was adapted to avoid problems that occurred in our previous study and the implemented behaviors were investigated in perception studies with US American and Arab participants. Therefore, we think our approach can serve as a guidance for other culture-specific generation approaches.

In the current paper user preferences were evaluated investigating which culture-specific version of the modeled behavioral aspects are preferred by users of different cultural backgrounds. With it, we want to contribute to the field by giving a deeper insight in culture-related user preferences of virtual characters' dialog behavior. However, we did not investigate the cultural perception of the behaviors yet to see whether human users actually perceive the cultural differences. For our future work, we plan a follow up experiment in which we will evaluate which of our prototypes is perceived as most German or most Japanese by users of the target cultural backgrounds.

Acknowledgements This work was partly funded by the European Commission within the 7th Framework Program under Grant agreement eCute. This paper constitutes a summary of a PhD thesis which was partly covered by a Grant from the Elitenetzwerk Bayern (Elite Network Bavaria). Earlier parts of the work were done in scope of the Cube-G project funded by the German Research Foundation (DFG).

References

1. André, E., Klesen, M., Gebhard, P., Allen, S., & Rist, T. (2000). Integrating models of personality and emotions into lifelike characters. In A. Paiva (Ed.), *Affective interactions* (pp. 150–165). Berlin: Springer.
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 C. Sierra, C. Castelfranchi, K. S. Decker, & J. S. Sichman (Eds.), *Proc. of 8th int. conf. on autonomous agents and multiagent systems (AAMAS 2009)*. Budapest: AAMAS.
3. Bergmann, K., & Kopp, S. (2009). Bayesian decision networks for iconic gesture generation. In Z. Ruttkay, M. Kipp, A. Nijholt, & H. H. Vilhjálmsson (Eds.), *Proc. of 9th int. conf. on intelligent virtual agents (IVA 2009)* (pp. 76–89). Amsterdam: Springer.
4. Bull, P. (1987). *Posture and gesture*. Oxford: Pergamon Press.
5. Byrne, D. (1971). *The attraction paradigm*. New York: Academic Press.
6. Cassell, J. (2007). Body language: Lessons from the near-human. In J. Riskin (Ed.), *Genesis redux: Essays in the history and philosophy of artificial intelligence* (pp. 346–374). Chicago: University of Chicago Press.
7. Damian, I., Endrass, B., Huber, P., Bee, N., & André, E. (2011). Individualized agent interactions. In *Proc. of 4th int. conf. on motion in games (MIG 2011)*, Edinburgh.
8. Decision Systems Laboratory. (2005–2007). GeNIe and SMILE. <http://genie.sis.pitt.edu/>. Last viewed: 03.12.2012
9. eCute Project. (2012). eCute: Education in cultural understanding, technologically-enhanced. <http://ecute.eu/>. Last viewed: 03.12.2012
10. Ekman, P. (1992). *Telling lies—clues to deceit in the marketplace, politics, and marriage* (3rd ed.). New York: Norton and Co.
11. Endrass, B., Huang, L., André, E., & Gratch, J. (2010). A data-driven approach to model culture-specific communication management styles for virtual agents. In W. van der Hoek, G. A. Kaminka, Y. Lespérance, M. Luck, & S. Sen (Eds.), *Proc. of 9th int. conf. on autonomous agents and multiagent systems (AAMAS 2010)* (pp. 99–108). Toronto: AAMAS.
12. Endrass, B., Nakano, Y., Lipi, A., Rehm, M., & André, E. (2011). Culture-related topic selection in SmallTalk conversations across Germany and Japan. In H. H. Vilhjálmsson, S. Koop, S. Marsella, & K. R. Thórisson (Eds.), *Proc. of 11th int. conf. on intelligent virtual agents (IVA 2011)* (pp. 1–13). Berlin: Springer.
13. Endrass, B., Rehm, M., & André, E. (2009). Culture-specific communication management for virtual agents. In K. S. Decker, J. S. Sichman, C. Sierra, & C. Castelfranchi (Eds.), *Proc. of 8th int. conf. on autonomous agents and multiagent systems (AAMAS 2009)*. Budapest: AAMAS.
14. Endrass, B., Rehm, M., & André, E. (2011). Planning small talk behavior with cultural influences for multiagent systems. *Computer Speech and Language*, 25(2), 158–174.
15. Endrass, B., Rehm, M., Lipi, A. A., Nakano, Y., & André, E. (2011). Culture-related differences in aspects of behavior for virtual characters across Germany and Japan. In L. Sonenberg, P. Stone, K. Tumer, & P. Yolum (Eds.), *Proc. of 10th int. conf. on autonomous agents and multiagent systems (AAMAS 2011)* (pp. 441–448). Richland: AAMAS.
16. Gallaher, P. E. (1992). Individual differences in nonverbal behavior: Dimensions of style. *Journal of Personality and Social Psychology*, 63(1), 133–145.
17. Hall, E. T. (1959). *The silent language*. New York: Doubleday.
18. Hall, E. T. (1966). *The hidden dimension*. New York: Doubleday.
19. Hall, E. T. (1983). *The dance of life*. New York: Doubleday.
20. Hartmann, B., Mancini, M., & Pelachaud, C. (2006). Implementing expressive gesture synthesis for embodied conversational agents. In S. Gibet, N. Courty, & J. F. Kamp (Eds.), *Gesture workshop (GW 2005)*. LNAI (Vol. 3881, pp. 188–199). Heidelberg: Springer.
21. Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organisations. Software of the mind. Intercultural cooperation and its importance for survival*. New York: McGraw Hill.
22. Hofstede, G. J. (2005). Role playing with synthetic cultures: the evasive rules of the game. In *9th International workshop of the IFIP Wg 5.7 SIG on experimental learning in industrial management*, Helsinki.
23. Hofstede, G. J., Pedersen, P. B., & Hofstede, G. (2002). *Exploring culture—exercises, stories and synthetic cultures*. Yarmouth: Intercultural Press.
24. Isbister, K., Nakanishi, H., Ishida, T., & Nass, C. (2000). Helper agent: Designing an assistant for human–human interaction in a virtual meeting space. In *Proceedings of CHI 2000* (pp. 57–64). New York: ACM Press.
25. Johnson, W. J., Marsella, S., & Vilhjálmsson, H. (2004). The DARWARS tactical language training system. In *Interservice/industry training, simulation, and education conference*, Orlando.

26. Kim, J., Hill, R. W., Durlach, P., Lane, H. C., Forbell, E., Core, M., Marsella, S., Pynadath, D., & Hart, J. (2009). BiLAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education*, 19, 289–308. IOS Press.
27. Koda, T., Rehm, M., & André, E. (2008). Cross-cultural evaluations of avatar facial expressions designed by western designers. In H. Prendinger, J. Lester, & M. Ishizuka (Eds.), *Proc. of 8th int. conf. on intelligent virtual agents (IVA 2008)* (pp. 245–252). Berlin: Springer.
28. McNeill, D. (1992). *Hand and mind—what gestures reveal about thought*. Chicago/London: University of Chicago Press.
29. Nau, D., Cao, Y., Lotem, A., & Munoz-Avila, H. (1999). SHOP: Simple hierarchical ordered planner. In *Proc. of 16th int. joint conf. on artificial intelligence (IJCAI 1999)* (pp. 968–973). Stockholm: IJCAI
30. Rehm, M., André, E., Bee, N., Endrass, B., Wissner, M., Nakano, Y., et al. (2009). Creating standardized video recordings of multimodal interactions across cultures. In M. Kipp, J. C. Martin, P. Paggio, & D. Heylen (Eds.), *Multimodal Corpora. LNAI* (Vol. 5509, pp. 138–159). Berlin: Springer.
31. Rehm, M., André, E., Nakano, Y., Nishida, T., Bee, N., Endrass, B., et al. (2007). The CUBE-G approach—coaching culture-specific nonverbal behavior by virtual agents. In I. Mayer & H. Mastik (Eds.), *ISAGA 2007: Organizing and learning through gaming and simulation*. Delft: Eburon.
32. Rehm, M., Nakano, Y., Koda, T., & Winschiers-Theophilus, H. (2012). Culturally aware agent communication. In M. Zacarias & J. de Oliveira (Eds.), *Human–computer interaction. SCI* (Vol. 396, pp. 411–436). Berlin: Springer.
33. Schneider, K. P. (1988). *Small talk: analysing phatic discourse*. Marburg: Hitzeroth.
34. Ting-Toomey, S. (1999). *Communicating across cultures*. New York: The Guilford Press.
35. Trompenaars, F., & Hampden-Turner, C. (1997). *Riding the waves of culture—understanding cultural diversity in business*. London: Nicholas Brealey Publishing.
36. Yin, L., Bickmore, T., & Cortés, D. E. (2010). The impact of linguistic and cultural congruity on persuasion by conversational agents. In J. M. Allbeck, N. I. Badler, T. W. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Proc. of 10th int. conf. on intelligent virtual agents (IVA 2010)* (pp. 343–349). Berlin: Springer.