Culture-related differences in aspects of behavior for virtual characters across Germany and Japan

Birgit Endrass
Elisabeth André
Human Centered Multimedia
Augsburg University
Universitätsstr. 6a
D-86159 Augsburg, Germany
endrass,
andre@hcm-lab.de

Matthias Rehm
Department of Media
Technology
Aalborg University
Niels-Jernes Vej 14
DK-9220 Aalborg, Denmark
matthias@imi.aau.dk

Afia Akter Lipi
Yukiko Nakano
Dept. of Computer and
Information Science
Seikei University
Musashino-shi, Tokyo,
180-8633 Japan
y.nakano@st.seikei.ac.jp
afiaakhter@hotmail.com

ABSTRACT

Integrating culture as a parameter into the behavioral models of virtual characters to simulate cultural differences is becoming more and more popular. But do these differences affect the user's perception? In the work described in this paper, we integrated aspects of non-verbal behavior as well as communication management behavior into the behavioral models of virtual characters for the two cultures of Germany and Japan in order to find out which of these aspects affect human observers of the target cultures. We give a literature review pointing out the expected differences in these two cultures and describe the analysis of a multi-modal corpus including video recordings of German and Japanese interlocutors. After integrating our findings into a demonstrator featuring a German and a Japanese scenario, we presented the virtual scenarios to human observers of the two target cultures in an evaluation study.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents*; I.6.7 [**Simulation and Modeling**]: Model Development

General Terms

Human Factors, Design, Experimentation

Keywords

Virtual Agents, Multiagent Systems, Culture, Communication Management, Nonverbal Behavior

1. MOTIVATION

A vast part of our communication happens non-verbally. While we might be thinking about what we want to communicate verbally, we manage our non-verbal behavior mostly subconsciously. Thereby, we integrate our personality, emotional state and cultural background into our behavior. How this behavior is interpreted depends on the listener's social and personal background as well. Enormous effort has been done so far in integrating these personal or social factors into the behavior models of virtual characters.

Culture has come in focus lately as another important factor that influences the success of an interaction with a virtual character. How different culture-specific behavior patterns of virtual characters are perceived and interpreted across different cultures has not been studied so far. In this paper, we integrated findings about such culture-specific behavior patterns into the behavior model of virtual characters. Our main goal is to find out which aspects of behavior result in a positive or negative impression on the user. Thus, the interpretation of different behavioral aspects is tested in isolation, making use of the same underlying dialog. For the implementation and evaluation, we choose the basic behavioral dimensions of communication management in terms of pauses and overlaps between turns as well as gestural expressivity and body posture. The former have been shown to be basic structuring mechanisms for face to face communications [9], the latter have been shown to differ broadly between cultures [5]. In addition, all have been attributed as provoking misunderstandings in inter-cultural communications [27].

For that task, several challenges had to be solved. A standardized video corpus was collected in the participating cultures [26] and the data was analyzed in the target cultures simultaneously with equal quality [24], [8] and [20]. In order to integrate the findings into a multiagent system, on the one hand the virtual characters' appearances have to be adapted to their cultural background on the other hand different behavioral models have to be built in order to match the cultural-background. To evaluate these models, studies have been set up in the participating cultures. In our previous work, we concentrated on either the analysis of behavioral differences or evaluation studies in only one culture. The aim of this paper is to find out which behavioral aspects have an effect on the perception of human observers of the two target cultures and whether participants prefer agent behavior that was designed for their own cultural background.

This paper is organized as follows: In the following section (Section 2), we discuss related work in the research

Cite as: Culture-related differences in aspects of behavior for virtual characters across Germany and Japan, B. Endrass, M. Rehm, A.A. Lipi, Y. Nakano and E. André, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 441-448.

field of integrating culture into virtual agent applications. In the next chapter (Section 3), we introduce some theoretical background and state our expectations about differences in behavioral aspects for the two cultures of Germany and Japan drawn from the literature. In Section 4, we describe a video corpus, that was recorded in the above mentioned cultures as well as our analysis of culture-related differences. We focus on the above-mentioned basic behavioral dimensions of gestures, postures and communication management. Then, we describe the integration of our findings into a demonstrator (Section 5). Section 6 then gives details on our study, where we evaluated whether participants have preferences for agent behavior that was designed to match their own cultural background, before we conclude the paper (Section 7).

2. RELATED WORK

The aim of the work described in this paper is to integrate different aspects of culture-specific behaviors into a multiagent system in order to find out which behaviors affect the user's perceptions. In the following we summarize some related work on integrating culture into the behavioral models of virtual characters.

Only a few attempts have been made to integrate the aspect of culture into the behavioral models of virtual characters. An example includes the Tactical Language Training System (TLTS) by Johnson and colleagues [16]. In order to complete the tasks provided by the system, trainees have to learn a foreign language. So far, four versions of TLTS have been implemented: Iraqi, Dari, Pashto, and French [17]. Through interaction with the people in the virtual world, the learner is supposed to develop cultural sensitivity.

Aylett et al. [1] introduce an educational application that uses fantasy characters in order to develop intercultural empathy. Culture-related differences are expressed through different symbols and rituals. The agents adapt their behavior in a culture-specific way and interpret incoming events according to cultural background. Our aim, however, is the simulation of behavioral aspects in existing national cultures in order to find out which patterns affect the user's perception.

An approach that focuses on the perception of virtual characters simulating synthetic cultures is presented in Mascarenhas et al. [21]. For their simulation, two different groups of characters were created that differed in their rituals and cultural dimensions. A perception study showed that the subjects found significant differences in the cultures and were able to relate these differences to the phenomenon of culture.

Focusing on the different perception of virtual characters' appearances across cultures, Koda et al. [18] designed culture-specific comic-style agents to show different emotions to subjects from different cultures. The characters were perceived differently across cultures and emotions were interpreted more correctly in the corresponding culture. In [19], Koda et al. have a closer look at different regions of the face and conducted a cross-cultural study in Hungary and Japan in order to test the impact of facial regions as cues to recognize the emotions of virtual agents. In their results the authors report that Japanese subjects found facial cues in the eye region more important than Hungarians subjects, who vice versa concentrated more on facial cues in the mouth region.

An evaluation study that investigates the different perception of verbal and non-verbal behaviors is introduced by Iacobelli et al. [14]. In their work, the authors focus on ethnicity, by changing behaviors of the character and leaving the appearance constant. Ethnic identity and engagement were evaluated and their results reveal that users were able to relate the virtual agents correctly. This inspired our research and brought up the question which of the behaviors we plan on integrating affects the user's perception most.

An approach that deals with non-verbal behavior is presented in [15]. Jan et al. simulate cultural differences in non-verbals such as proxemics and gaze. In a user study, the authors evaluated whether their participants perceived differences between behaviors associated with their own cultural background and behaviors simulating a different cultural background. In a similar manner, we want to find out if users from Germany and Japan prefer behaviors that are built to match their own cultural background for the aspects of communication management behaviors, gestural expressivity and posture.

In the CUBE-G project [23], we aim on the integration of culture-specific behaviors for interaction with embodied conversational agents in order to build a training scenario for human users. Therefore, culture-specific behavior has been analyzed in a video corpus. So far, we analyzed non-verbal behaviors [25] and communication management behaviors [8] and integrated our findings into a demonstrator featuring a German and a Japanese dialog scenario. Furthermore, the impact of these behavioral differences has been partly evaluated by German observers. The studies showed that German participants preferred the German communication management scenario over the Japanese scenario. Japanese participants have not been considered yet. For non-verbal behavior, the question of whether observers prefer non-verbal behaviors in virtual scenarios that correspond to their own culture remaines still unanswered. The aim of this paper is to correlate the results in non-verbal behaviors and communication management behaviors with an evaluation study in both participating cultures.

3. THEORETICAL BACKGROUND

As we stated above, we are looking at different aspects of behavior in the two cultures of Germany and Japan. In particular, we focus on communication management behaviors (pauses in speech and overlaps), body posture and gestural expressivity. In this section, we introduce these behaviors and state our expectations in culture-related differences drawn from the literature.

As one aspect of behavior that might affect the perception of a particular conversation, we had a closer look at communication management behaviors. So-called regulators are used in order to manage communication [27]. *Vocalics* include verbal feedback signals (such as "uh-huh") as well as the usage of silence in speech or interruptions of the communication partner. Depending on the usage of these vocalics, a different rhythm of speech can evolve. *Kinesics* and *oculesics* comprise non-verbal regulators. According to [27], communication can be managed though hand gestures and body postures (kinesics) or eye and face gaze (oculesics).

These regulators are used to control the flow and pauses of a conversation and are considered culture-specific behaviors. In addition, regulators are used at a very low level of awareness since they are learned at a very young age [27].

Table 1: Hofstede's scores on the five dimensions of culture for the two cultures of Germany and Japan as well as the world average.

Culture / Dimension	Germany	World Average	Japan
PDI	35	55	54
IDV	67	64	46
MAS	66	48	95
UAI	65	61	92
LTO	31	41	80

We therefore consider regulators as an interesting aspect of behavior that might have an effect on the perception of a given conversation depending on the culture of the listener. This is in line with Ting-Toomey [27], who states that a discriminative use of regulators can cause intercultural distress or misunderstandings.

Another interesting aspect of behavior is the expressivity of non-verbal behaviors. How we exhibit a gesture can sometimes be more crucial for the observer's perception than the gesture itself. Differences in the dynamic variation can be described according to expressivity parameters [22]. The *spatial extend*, for example, describes the arm's extend toward the torso. The *speed* of a gesture and the *power* of the arm can vary as well. The *fluidity* parameter describes the continuity between consecutive gestures, while the *repetitivity* holds information about the repetition of the stroke. The last expressivity parameter, *overall activation*, counts the amount of gestures that are performed. How gestures are executed can depend on several individual and social factors such as personality, emotional state or culture.

Next in this study, we examined posture as another kind of non-verbal behavior. Posture is defined as a motion or position shift of the human body [3]. Based on previous studies, we defined four parameters to describe the characteristics of postures. The four parameters are *duration* till which a person remains in the same posture, *spatial extent* used in a posture, *rigidness* or relaxation apparent from the posture and *mirroring* as number of instances when an individual unconsciously imitates a partner's posture during a conversation. We already found that these parameters are useful in describing the culture variations in postures [20].

3.1 Culture-specific expectations

In the social sciences, culture is a well established research field. There are several approaches that define culture and describe differences in their behavior. A well-known model of culture was introduced by Hofstede [12], who built a five dimensional model in order to distinguish cultures. Over 20 different cultures were categorized in a broad empirical survey. Table 1 shows the scores of the two cultures of Germany and Japan, as published on Hofstede's web page [11]. Please note that these scores were normalized across cultures to stay between 0 and 100 in the first version and extended later, when more cultures were added and more extreme values were observed.

The Power Distance dimension (PDI) describes the extent to which a different distribution of power is accepted by the less powerful members of a culture. The Individualism dimension (IDV) 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. The gender or masculinity dimension (MAS) describes the distribution of roles between the genders and how masculine values are perceived. In feminine cultures, the roles differ less than in masculine cultures, while competition is rather accepted in masculine cultures and status symbols are of importance. In the uncertainty avoidance dimension (UAI), the tolerance for uncertainty and ambiguity is defined. It indicates to what extent the members of a culture feel either comfortable or uncomfortable in unstructured or unknown situations. The long-term orientation dimension (LTO) has been added afterwards, in order to explain differences between Asian and Western cultures. Values for long term orientation are, for example, thrift and perseverance; whereas examples for values for the short term orientation are respect for tradition, fulfilling social obligations, and saving one's face.

The positioning on these dimensions affects one's behavior. Taking a look at the cultural dimensions in isolation, Hofstede [13] introduces so-called synthetic cultures that find themselves on one of the extreme ends of each dimension. For these synthetic cultures he describes prototypical behavior norms. For the behavioral aspects investigated in our research, the individualism dimension and the power distance dimension are of special interest.

For collectivistic cultures, he states that silence may occur in conversations without creating tension. This observation does not hold true for individualistic cultures. In addition, he states that the usage of pauses can be a crucial feature in collectivistic cultures. Germany is a more individualistic culture than Japan (see Table 1, IDV). As a consequence, it should be more likely in the German culture that pauses in a conversation create tension and are thus tried to be avoided. In Japanese conversations, on the other hand, pauses can be considered a feature of the conversation.

Another behavioral aspect is affected by the power distance dimension. High-power cultures are described as verbal, soft-spoken and polite and interpersonal synchrony is much more important than in low-power cultures, whose members tend to talk freely in any social context [27]. One possibility to achieve interpersonal synchrony in a conversation is giving feedback. This feedback often occurs during the speaking floor of the interlocutor. This should occur more often in the Japanese culture due to their higher value on the power distance dimension (see Table 1, PDI). The individualism dimension is also related to the expression of emotions and the acceptable emotional displays in a culture. In individualistic cultures it is more acceptable to publicly display emotions than it is in collectivistic cultures [6]. This also suggests that non-verbal behavior is expressed more emotional in German conversations than in Japanese ones. We expect displaying emotions more obviously should affect the expressivity of gestures in a way that parameters such as speed, power or spatial extent are increased for a higher arousal in emotion.

Strengthening our expectations about the usage of silence in speech and overlapping speech, Ting Toomey [27] states that the beliefs expressed in talk and silence are culture-dependent. Following Hall's categorization of cultures [10] into high- and low context communication cultures, Ting Toomey [27] observes that conversation in high context com-

munication cultures relies mainly on physical context. Meaning can be transported through non-verbal cues, such as pauses, silence and prosody. In contrast, low context communication cultures tend to explicitly code information. Clear descriptions and a high degree of specificity are used commonly in these cultures. Germany is described as one of the most extreme low context cultures, while Japan finds itself on the extreme high context side [27]. Thus, communication management behaviors such as pauses in speech or overlapping speech, should occur more frequently in Japanese conversations. Verbal feedback is given in every culture but the meaning can vary with the communicative function expressed in the feedback. In Japanese conversations, for example, 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. Frequency and positioning of pauses and overlaps can vary across cultures, too. Overlapping speech is often considered as impolite. But feedback utterances are often performed while it is still the interlocutor's turn without wanting to gain the turn. As we stated above, acknowledgments are very common in Japanese conversations. Thus, we expect a high amount of overlapping speech in Japanese conversations that are short but frequent. In addition, Ting Toomey [27] states that silence serves as a critical communication-device in Japanese communication patterns. Pauses reflect the thoughts of the speaker and can contain strong contextual meaning.

Similar findings are described by Trompenaars and Hampden-Turner [28], who divide cultures into Western, Latin and Oriental cultures. While Germany is considered a Western culture, Japan would count as an Oriental culture (including Asian cultures). In line with Hofstede and Ting-Toomey, Trompenaars and Hampden-Turner describe Western cultures as verbal and state that their members get nervous when there are long pauses. In addition, they state that interruptions are considered as impolite. Thus, communication in Western cultures is managed as follows: interlocutors start talking after the other conversation partner stopped. In Oriental cultures silence is more important and can be considered a sign of respect. Pauses are used to process information or assure that the conversation partner gives away the speaking floor.

Summarizing our culture-specific expectations drawn from the literature, we expect more pauses in speech and overlapping speech such as in feedback behavior in Japanese conversations than in German ones. Gestures and postures should be more expressive in prototypical German behavior than in prototypical Japanese behavior.

4. EMPIRICAL VERIFICATION

Behavioral tendencies described in the literature are sometimes rather abstract. As we stated above, we expect more pauses in speech in Japanese conversations than in German ones, for example. In order to integrate our expectations into the behavior model of virtual characters, we need more details such as number or length of pauses. To answer these and other questions, we recorded and analyzed a video corpus in the two target cultures (see [23]). Three prototypical interaction scenarios were videotaped, while more than 20 subjects participated in each of the two cultures. In a total, around 20 hours of video material were collected. Subjects interacted with actors whom they did not know in advance in order to ensure that all subjects meet the same conditions and that all scenarios last for about the same time. For the first scenario, participants were asked to get acquainted with one another since they had to solve a task together later. Recordings started already during this time. The analysis described in the next section focuses on this first time meeting scenario, which lasted for around 5 minutes for each subject.

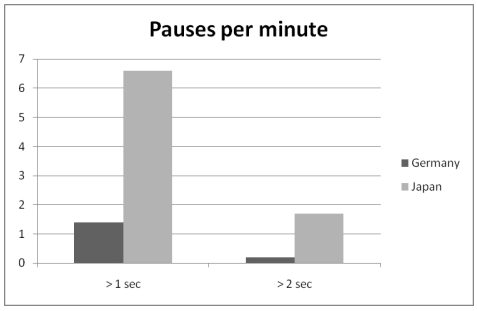
4.1 Analysis

As we stated above, we concentrate on several aspects of behavior such as the usage of pauses, overlapping speech, gestural expressivity and posture. The corpus described above was analyzed in order to find culture-related differences in these aspects [8] [24]. In the following section, we summarize our results:

For the analysis of pauses in speech, we considered as a pause the parts of the conversation where none of the conversation partners spoke and took into account the pauses that lasted for more than one second and more than two seconds respectively. In that manner, we sorted out very brief pauses that are used for breathing for example. Comparing the two cultures, we found more pauses in speech in the Japanese conversations. In the German videos, we found on average 7.1 pauses that lasted for more than one second and 1.3 pauses on average that lasted for more than 2 seconds. In the Japanese videos, we observed 31 pauses on average that lasted over 1 second and 8.4 pauses that lasted for more than 2 seconds. Figure 1 (left) shows the distribution of short (more than 1 second) and long pauses (more than 2 seconds) that were found on average per minute in each video. Comparing the amount of pauses in speech across the two cultures, using the two sided t-test, we achieved significance for both, pauses that last for more than 1 second ($p < 0.001$) and pauses that last for more than 2 seconds ($p < 0.001$).

Regarding overlapping speech, we considered time spans where both conversation partners spoke at the same time as overlapping speech. Pragmatics, such as using overlaps for feedback behavior, were not taken into account yet. The average occurrences of overlapping speech per subject per minute for the two cultures are shown in Figure 1 (right). We observed 6 overlaps per minute in German conversations on average, while in Japanese conversations 9 overlaps per minute occurred on average. Comparing the frequency of overlapping speech across the two cultures, we achieved significant results for the total amount of overlaps ($p = 0.04$). No significance was achieved for overlaps that last for more than 0.5 seconds ($p = 0.31$) and 1 second ($p = 0.12$). By trend, we observed more overlaps in the Japanese conversations for all lengths, which is in line with our expectations described above.

As we stated above, we analyzed gestures according to expressivity parameters (see Section 3). Each parameter was coded using a seven-point scale. Analyzing the two cultures, we found significant differences for all parameters (using ANOVA with $p < 0.01$ for all parameters). Figure 2 (left) shows the average ratings of the expressivity parameters for the two cultures of Germany and Japan. Gestures were performed faster and more powerfully in the German videos than in the Japanese one's. In addition, German subjects used wider space for their gestures compared to Japanese subjects who used less space. Gestures were also performed



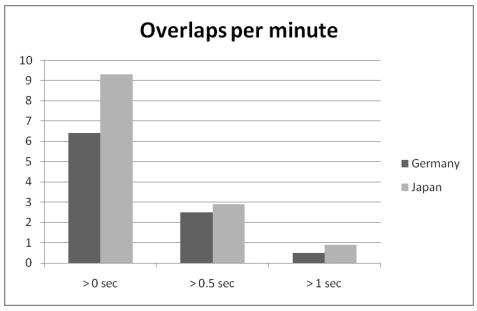
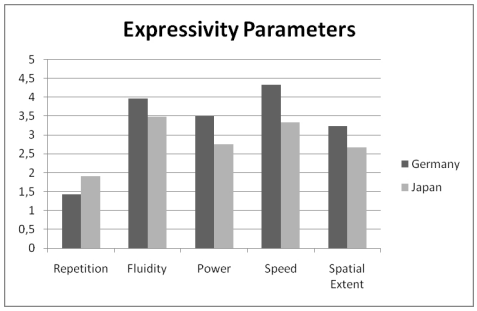


Figure 1: Pauses (left) and overlaps in speech (right) per minute, averaged over participants.



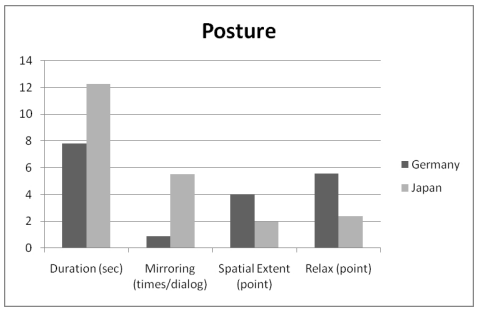


Figure 2: Ratings of expressivity parameters (left) and posture characteristics (right) in German and Japanese culture.

more fluently in the German conversations and the stroke of a gesture was repeated less in the Japanese conversations. For the analysis of posture, we used Bull's coding scheme [2] to label the posture type/shape. Figure 2 (right) shows the arm posture changes that were extracted from studying the corpus data of German and Japanese subjects. The value for duration was derived by calculating the average number of posture shifts observed in the data. To get the score for mirroring, we looked at the total number of common posture shapes of both interactors in each turn. The value for spatial extent and rigidness were assigned based on the average of 7 point scale ratings done by multiple annotators. We used the opposite word relax instead of rigidness to make the word easy to understand. Figure 2 (right) indicates that Japanese subjects remained in the same posture longer, engage in more frequent mirroring, take up less space, and display a more rigid posture in comparison to German subjects.

The postures most frequently observed in the German videos (folding the arms in front of the trunk (FAs) and putting the hands in the pockets of the trousers (PHIPt)) and in the Japanese videos (joining both hands in front of the body (JHs)) are exemplified in Figure 3 (left and middle). It is notable that ratings for postures frequently observed in the German corpus such as PHIPt and FAs were rated higher in spatial extent and lower in rigidness, compared to postures frequently observed in the Japanese data such as JHs and PHB (put hands back). Details of how values of each of the posture traits in relation to culture were obtained, are provided in [20].

5. SIMULATION

In order to simulate the behavioral tendencies described in the literature and verified by our empirical corpus study for the German and the Japanese cultures, we use the Virtual Beergarden scenario [4]. In the scenario, an arbitrary number of agents can be loaded that are able to move around in the scenario freely, exhibit gestures and communicate with each other. For the simulation of different cultures, culture-specific characters were modeled. Thus, we created prototypical German looking and prototypical Japanese looking characters (see Figure 3, left and middle) whose appearances (skin, hair or shape of the face) have been adapted to their cultural background.

Verbal behavior is realized by a text-to-speech component. For the different characters, different voices can be used, e.g. German, English or Japanese speech synthesis. Non-verbal behaviors are divided into gestures, postures and movement animations. Gestures can be culture-specific or not. An example of a culture-specific gesture is, for example, a bow for the Japanese greeting. Examples for culture-specific postures are shown in Figure 3 (left and middle). General gestures such as beat gestures can be exhibited by every agent. The performance of these gestures, however, can be customized and thus be performed in a culture-specific way. To this end, every gesture is divided into three phases: preparation, stroke and retraction. The preparation and retraction phases are used to blend the animations. A gesture could, for example, be chosen while the agent already performs another gesture or stands in a certain posture. The stroke phase can be performed in different ways taking into account the expressivity parameters. The parameter speed, for example, can be varied by playing the animation faster or slower; the parameter repetition can be changed by playing the stroke phase several times.

Figure 3: Culture-specific agents in the Virtual Beergarden (left: Germany; middle: Japan) and during the evaluation study (right: Japanese agents showing Japanese vs. German postures).

6. EVALUATION

Most misunderstandings in inter-cultural communication are caused by differences in non-verbal behavior [27]. In an evaluation study, we investigate whether the culture-related differences that we found in the literature and in our video corpus are perceived by human observers during agent interaction.

6.1 Design

In order to find out which of the behavioral aspects do have an impact on the user's perception, we simulated them in isolation. For the study conducted in Germany, the German looking characters were used and for the study conducted in Japan, we used the Japanese looking characters. In addition, we used language specific text-to-speech systems for the Western and Asian characters (German and Japanese) to match the prosody of the speech of the target culture. Thus, participants should not assume a cultural background different from their owns.

For each behavioral dimension, participants were shown two videos with face to face dialogs. In one video, the characters performed prototypical German behavior, in the other prototypical Japanese behavior for the specific behavioral aspect. In the study, participants had to state their preference by providing ratings on a 6 graded scale, containing three grades on each side, starting from "rather this video" to "by any means this video". For the two parted study, we stated the following two hypotheses:
H1: For each behavioral dimension, German participants prefer the videos showing German behavior over the Japanese versions.
H2: For each behavioral dimension, Japanese participants prefer the videos showing Japanese behavior over the German versions.

In order to avoid side effects evoked by gender, we showed mixed gender combinations in the videos. That is, one female and one male character interacted with each other in both cultures. To avoid preference for one of the videos due to the semantics of speech, 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. The same dialog was retained during the whole study changing only aspects of the non-verbal and communication management behaviors. Keeping the dialog consistent also assured that the users' perceptions are not influenced by other linguistic features, such as the length of the sentences.

In order to get participants acquainted with the situation of listening to a Gibberish dialog, we showed a neutral conversation first. In this video, the dialog described above was performed without any non-verbal behavior or any pauses in speech or overlapping speech. After this neutral video, six pairs of videos were shown in random order, each lasting for half a minute and containing differences in one of the following aspects of behavior (see Figure 3 (right) for a sreenshot of the evaluation study as it was conducted in Japan):

- Pauses in speech: As we observed more pauses in the Japanese corpus, the simulated dialogs reflecting typical Japanese conversations contain more pauses as well. Taking into account our corpus findings, German agent dialogs contained one pause that lasted one second, whereas the Japanese version contained two pauses that lasted one second and one pause that lasted two seconds.

- Overlapping speech: Following our analysis of overlapping speech, we integrated one overlap that lasted 0.3 seconds and two overlaps that lasted 0.5 seconds into the German dialog. The Japanese dialog contained three overlaps that lasted 0.3 second, one that lasted 0.5 seconds and one that lasted one second.

- Communication management: Videos showing communication management behavior contained both: pauses and overlaps as described above.

- Speed of gestures: Our findings showed that in the German corpus gestures are performed faster than in the Japanese one. Thus, in one pair of the videos the gestures were customized according to speed. Three gestures were shown in both videos, but played faster in the German and slower in the Japanese behavior model.

- Spatial extent of gestures: Similar to gesture speed, another screen in the study contained two videos showing gestures with a different spatial extent. According to our findings, gestures had a smaller spatial extent in the Japanese models.

- Postures: The posture evaluation does not take the results on mirroring into account yet, but looks only into the interpretation of dominant body postures found in our corpus study for the two cultures.

6.2 Results and Discussion

As we stated earlier, we designed two different versions of our evaluation study. One utilizing the German-looking characters and a German text-to-speech system and another one using the Japanese-looking characters and a Japanese text-to-speech system, each showing both behavioral models. Instruction texts as well as preference questions matched the participants' mother tongue. In the German evaluation study, 15 participants took part (6 female and 9 male), while in the Japanese study 17 people participated (3 female and 14 male). All subjects were students (with one exception in the German study) in an age range between 20 and 45. In the evaluation study, participants had to decide which of the videos they liked better, assuming that participants prefer videos showing virtual characters that behave in a way that was designed for their own cultural background. In a goodness-of-fit test, we tested whether the observed pattern of events significantly differed from what we might have expected by chance alone.

Significantly more than 50% of our German participants had a preference for the version with German overlapping speech and spatial extent in gestures (both with $chi^2 = 8{,}067$ and p = 0.005 with df = 1). For pauses in speech, communication management and posture, we almost achieved significance (with $chi^2 = 3.26$ and p = 0.071 with df = 1 for all three aspects). However, by trend German participants showed a preference for the videos simulating prototypical German behavior for all aspects of behavior.

Results in the Japanese study are less strong. Significantly more than 50% of our Japanese participants had a preference for the version with Japanese posture behavior (with $chi^2 = 4.675$ and p = 0.029 with df = 1). For other behavioral patterns, we cannot claim any evidence. The results for pauses in speech and overlapping speech, however, were a bit surprising for the Japanese study as participants seemed to favor the German videos over the Japanese ones (although not significant). We attribute the missing semantics of the Gibberish dialogs as the main reason for this result, based on the following observations: On the German side pauses are generally viewed as somewhat awkward and overlaps as rude regardless of the semantic content of utterance. On the other hand, as discussions with our Japanese project partners showed afterwards, the use of pauses and overlaps in the Japanese language seems to be tight to the semantics of the utterances 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.

This "failure" highlights a very important aspect of cross-cultural interaction in research teams. Despite frequent discussions and experience in cross-cultural projects, the developer's own cultural expectations are always present and sometimes interfere with the development. In this case, the seemingly good solution of using Gibberish for the tests, due to the arguments given above, lead us to missing an important feature of Japanese dialogs, i.e. its high context nature as Hall puts it [10].

Interestingly, the results for communication management behavior seem to be more related to the results from pause behavior than the results from overlapping behavior. We made similar observations in [7], where we considered communication management behaviors for the two cultures of Arabia and US America. The analyses suggested too, that the impact of pause behavior was stronger than the impact of overlapping behavior to human observers.

Although, we only had a limited number of participants in our study, for some cases we have significant results suggesting that behavioral patterns are preferred that were designed for the participants cultural background. However, for none or the behavioral patterns, we found evidence that more than 50% of our participants prefered behavior that did not match their cultural background.

7. CONCLUSION

In this paper, we investigated different behavioral dimensions for the two cultures of Germany and Japan, in order to find out which of these aspects affect the human observer. Focusing on parts of communication that are performed rather subconscious and where the influence of culture can play a crucial role without even realizing it, we concentrated on aspects of non-verbal behavior and communication management and did not consider semantics of speech yet. Culture-related differences have been extracted from the literature for the two cultures of Germany and Japan and strengthened by a empirical corpus study in the two target cultures. Results have been integrated into a multiagent system that demonstrates the simulation of cultural patterns of behavior.

For our evaluation study, behavioral aspects were tested in isolation. In that manner, we wanted to find out which of these aspects affect the perception of the user. Our preliminary evaluation study in Germany revealed that subjects significantly preferred the version that resembled behavior observed for their own cultural background for some of the behavioral aspects (overlapping speech and spatial extent of gestures). For all other aspects participants seemed to prefer the German versions at least by trend. In the Japanese evaluation study, we found out that Japanese subjects significantly preferred postures designed for their cultural background. Only for pauses in speech and overlapping speech we observed a controversial trend. One reason for this outcome might be the missing semantics of the shown dialogs. Since the Japanese version contained both more pauses and more overlaps in speech, but lacked the context in which they occur, participants chose the safe solution, i.e. the version with less pauses and overlaps. As a consequence, we think that pauses and overlaps need to be placed very carefully and in relation to the actual dialog.

Reflecting on our findings, we plan to refine our models in communication management by adding context. In another step, we want to combine all the aspects of behavior that we investigated in isolation and build a scenario with virtual characters that behave according to their cultural background on different channels. In that way, we want to believably simulate different cultural backgrounds and create an awareness for these differences on the user's side.

8. ACKNOWLEDGMENTS

The first author of this paper was supported by a grant

from the Elitenetzwerk Bayern (Elite Network Bavaria). This work was also partly funded by the European Commission within the 7th Framework Program under grant agreement eCute (education in cultural understanding, technologically enhanced) and the Japan Society for the Promotion of Science (JSPS) under a Grant-in-Aid for Scientific Research (C) (19500104) and (S) (19100001).

9. REFERENCES

[1] R. Aylett, A. Paiva, N. Vannini, S. Enz, E. André, and L. Hall. But that was in another country: agents and intercultural empathy. In Decker, Sichman, Sierra, and Castelfranchi, editors, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Budapest, Hungary, 2009.

[2] P. Bull. *Posture and Gesture*. Pergamon Press, Oxford, 1987.

[3] J. Cassell, Y. Nakano, T. Bickmore, C.Sidner, and C. Rich. Non-verbal Cues for Discourse Structure. In *The 39th Annual Meeting of the Association for Computational Linguistic (ACL 01)*, pages 106–115, 2001.

[4] I. Damian, P. Huber, B. Endrass, and N. Bee. Advanced Agent Animation. In *IVA Gala 2010*, 2010.

[5] D. Efron. *Gesture, Race and Culture*. Mouton and Co, 1972.

[6] P. Ekman. *Telling Lies - Clues to Deceit in the Marketplace, Politics, and Marriage*, volume 3rd edn. Norton and Co., New York, 1992.

[7] B. Endrass, L. Huang, E. André, and J. Gratch. A data-driven approach to model Culture-specific Communication Management Styles for Virtual Agents. In *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, 2010.

[8] B. Endrass, M. Rehm, and E. André. Culture-specific communication management for virtual agents. In Decker, Sichman, Sierra, and Castelfranchi, editors, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Budapest, Hungary, 2009.

[9] C. Goodwin. *Conversational Organization — Interaction between Speakers and Hearers*. Academic Press, New York, 1981.

[10] E. T. Hall. *The Silent Language*. Doubleday, 1959.

[11] G. Hofstede. http://www.geert-hofstede.com/.

[12] G. Hofstede. *Culture's Consequences - Comparing Values, Behaviours, Institutions, and Organizations Across Nations*. Sage Publications, 2001.

[13] G. J. Hofstede, P. B. Pedersen, and G. Hofstede. *Exploring Culture - Exercises, Stories and Synthetic Cultures*. Intercultural Press, Yarmouth, United States, 2002.

[14] F. Iacobelli and J. Cassell. Ethnic Identity and Engagement in Embodied Conversational Agents. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, editors, *Proc. of Conf. on Intelligent Virtual Agents (IVA 2007)*, pages 57–63. Springer, 2007.

[15] D. Jan, D. Herrera, B. Martinovski, D. Novick, and D. Traum. A Computational Model of Culture-Specific Conversational Behavior. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, editors, *Intelligent Virtual Agents (IVA 2007)*, pages 45–56. Springer, 2007.

[16] W.-J. Johnson, S. Marsella, and H. Vilhjálmsson. The DARWARS Tactical Language Training System. In *Interservice / Industry Training, Simulation, and Education Conference*, 2004.

[17] W.-L. Johnson and A. Valente. Tactical Language and Culture Training Systems: Using Artificial Intelligence to Teach Foreign Languages and Cultures. In *Innovative Applications of Artificial Intelligence (IAAI 2008)*, pages 1632–1639. Association for the Advancement of Artificial Intelligence (AAAI), 2008.

[18] T. Koda, M. Rehm, and E. André. Cross-cultural evaluations of avatar facial expressions designed by western designers. In H. Prendinger, J. Lester, and M. Ishizuka, editors, *Proc. of Conf. on Intelligent Virtual Agents (IVA 2008)*, pages 245–252. Springer, 2008.

[19] T. Koda, Z. Ruttkay, Y. Nakagawa, and K. Tabuchi. Cross-cultural study on facial regions as cues to recognize emotions of virtual agents. In T. Ishida, editor, *Culture and Computing*, pages 16–27. Springer, 2010.

[20] A.-A. Lipi, Y. Nakano, and M. Rehm. Culture and Social Relationship as Factors of Affecting Communicative Non-verbal Behaviors. *Japanese Society of Artificial Intelligence*, 25(6):712–722, 2010.

[21] S. Mascarenhas, J. Dias, N. Afonso, S. Enz, and A. Paiva. Using rituals to express cultural differences in synthetic characters. In Decker et al., editor, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, 2009.

[22] C. Pelachaud. Multimodal expressive embodied conversational agents. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 683–689, 2005.

[23] M. Rehm, E. André, Y. Nakano, T. Nishida, N. Bee, B. Endrass, H.-H. Huan, and M. Wissner. The CUBE-G approach - Coaching culture-specific nonverbal behavior by virtual agents. In I. Mayer and H. Mastik, editors, *ISAGA 2007: Organizing and Learning through Gaming and Simulation*, 2007.

[24] M. Rehm, Y. Nakano, E. André, T. Nishida, N. Bee, B. Endrass, M. Wissner, A.-A. Lipi, and H.-H. Huang. From observation to simulation: generating culture-specific behavior for interactive systems. *AI & Society*, 24(3):209–211, 2009.

[25] M. Rehm, Y. Nakano, E. André, and T. Nishida. Culture-specific first meeting encounters between virtual agents. In *Intelligent Virtual Agents 2008 (IVA 2008)*, pages 223–236, 2008.

[26] M. Rehm, Y. Nakano, H.-H. Huang, A.-A. Lipi, Y. Yamaoka, and F. Grueneberg. Creating a standardized corpus of multimodal interactions for enculturating conversational interfaces. In *IUI-Workshop on Enculturating Interfaces (ECI)*, 2008.

[27] S. Ting-Toomey. *Communicating across Cultures*. The Guilford Press, New York, United States, 1999.

[28] F. Trompenaars and C. Hampden-Turner. *Riding the waves of culture - Understanding Cultural Diversity in Business*. Nicholas Brealey Publishing, London, 1997.