

A data-driven approach to model Culture-specific Communication Management Styles for Virtual Agents

Birgit Endrass
Elisabeth André
Multimedia Concepts and Applications
Augsburg University
Universitätsstr. 6a
D-86159 Augsburg, Germany
{endrass, andre}
@informatik.uni-augsburg.de

Lixing Huang
Jonathan Gratch
Institute for Creative Technologies
University of Southern California
13274 Fiji Way
Marina del Rey, CA 90292 USA
lixinghu@usc.edu
gratch@ict.usc.edu

ABSTRACT

Virtual agents are a great opportunity in teaching inter-cultural competencies. Advantages, such as the repeatability of training sessions, emotional distance to virtual characters, the opportunity to over-exaggerate or generalize behavior or simply to save the costs for human training-partners support that idea. Especially the way communication is coordinated varies across cultures. In this paper, we present our approach of simulating differences in the management of communication for the American and Arabic cultures. Therefore, we give an overview of behavioral tendencies described in the literature, pointing out differences between the two cultures. Grounding our expectations in empirical data we analyzed a multi-modal corpora. These findings were integrated into a demonstrator using virtual agents and evaluated in a preliminary 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

Experimentation, Design, Human Factors

Keywords

Virtual Agents, Multiagent systems, Culture, Behavior, Simulation, Communication management

1. MOTIVATION

Language is the most obvious barrier when people from different cultures want to communicate. But even if both communication partners speak the same language fluently, misunderstandings may occur. Unfortunately, these misunderstandings are often not even perceived as such. Sub-

Cite as: A data-driven approach to model Culture-specific Communication Management Styles for Virtual Agents, Birgit Endrass, Lixing Huang, Elisabeth André and Jonathan Gratch, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 99–106
Copyright © 2010, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

consciously, the communication partners judge their interlocutor and interpret their words and behavior in a culture-specific way. Since people assume that their assumption is correct, a wrong interpretation stays unrecognized. Thus, a person might be considered as rude without the intention to appear in that way. In Hungary and France for example, interrupting the communication partner is interpreted as showing interest in the conversation, while in other cultures, such as Japan, this behavior is regarded as very rude and perceived as spelling into the conversations partner's speech or not waiting for his or her turn. Imagine two people from these cultures talking to each other. The conversation might lead to frustration on both sides, as one partner might think that the other is not interested in the conversation while the other thinks that the interlocutor is interrupting in an inappropriate manner. Although both conversation partners are not satisfied with the flow of the dialog, it is very unlikely that one of the two will address this problem directly, as this would seem awkward in both cases.

Being aware of the fact that the interpretation of certain behaviors is different across cultures is a first step towards inter-cultural understanding. This is in line with Hofstede's research [10], who names awareness as the first of three steps on the way to cultural understanding. These three steps are further described in the following:

1. Awareness: The first step of gaining inter-cultural competence is being aware that there are differences in behavior. According to Hofstede the hardest part of this step is to accept that there are no better or worse ways of behaving and especially that one's own behavior routines are not superior to others.
2. Knowledge: The second step in inter-cultural training is to learn where exactly differences in behavior are as well as knowing their importance and possible consequences in case of behaving differently.
3. Skills: Hofstede argues that the first two steps are crucial in order to avoid the most common inter-cultural misunderstandings. In case of integration into a certain culture a third step is necessary: the training of communication skills for that specific culture.

Especially for the first two steps of learning inter-cultural communication virtual agents are a powerful medium. By

observing their behavior, a trainee can learn about culture-specific differences without the need to embarrass real humans by watching them, or reading complicated explanations in textbooks. Following this idea, we want to use virtual agents in order to create cultural awareness and thus, prevent people from judging others without knowing that misunderstandings might be aroused by culture. However to achieve this goal, we first need to integrate certain culture-related differences in behavior into a multiagent system and test their impact on human observers. In the work described in this paper, we focus on the interpersonal management of a conversation. Therefore, we focus on the usage of silence, overlaps in speech as well as the frequency of giving verbal feedback in a culture-specific manner.

This paper is organized as follows: In the following section (Section 2), we summarize some related work in the research field of integrating culture into virtual agent applications. In the next chapter (Section 3), we introduce the theoretical background from the social sciences. Therefore, we start with so-called regulators that are used to manage interpersonal communication. Afterwards, we classify the two cultures that we chose for our study (America and Arabia) and explain our expectations regarding the usage of regulators in these two cultures, drawn from tendencies described in 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 in interpersonal coordination. We then describe the integration of our findings into a multiagent system (Section 5). In a preliminary study, we evaluated whether subjects perceive a difference between these behaviors (Section 6). In the last chapter (Section 7), we summarize and discuss our results and discuss possible future work in that area.

2. RELATED WORK

The aim of the work described in this paper is to integrate culture-specific behavior into a multiagent system in order to train cultural awareness for human users. Enculturating virtual agents for that purpose becomes more and more popular in the last years.

In [14] for example, Jan et al. present a computational model to simulate cultural difference, focusing on non-verbal behavior clues, such as proxemics and gaze. Evaluating their model, they showed that subjects perceive differences between behaviors associated with their own cultural background and behaviors simulating a different cultural background.

Another example is described by Core et al. [5]. The authors focus on training human users to negotiate with virtual agents that have a different cultural background than their own. The trainees gain language skills as well as knowledge about cultural-specific non-verbal behaviors, such as gestural emblems. In the role-playing scenario different negotiation styles have been implemented for the virtual agents. However, their behavior does not adapt to cultural background yet.

Instead of talking about culture as a national thing, Iacobelli et al. [13] focus on ethnicity and present a virtual peer that shows different verbal and non-verbal behaviors. By changing these behaviors and leaving the appearance of the virtual character constant, the authors tested human subjects' ethnic identity and engagement. Their evaluation suggests that users were able to relate the virtual agents cor-

rectly and that children engaged with the virtual peers in a promising way for educational applications.

Koda et al. [15] investigated the impact of virtual characters' culture-specific appearance and whether emotions are judged the same way across cultures. They therefore designed different appearances for virtual characters according to several cultures-specific comic-styles. Showing the designed emotions to subjects from different cultures, they found out the agents were perceived differently across cultures and that emotions are interpreted in a more correct manner in the same culture that the emotion was designed in.

Aylett et al. [2] introduce an educational application that uses virtual agents to develop inter-cultural empathy. In their system, they use fantasy characters that have their own symbols, rituals and culture. In order to simulate different cultures, the underlying agent architecture [6] was extended to allow cultural-specific adaptation of the agents and interpretation of incoming events in a culture-related way. However, their aim is not to simulate an existing national culture or to increase the awareness of certain cultural-specific features of human behaviors.

The CUBE-G project [16] integrates culture-adaptive behaviors for interaction with embodied conversational agents in order to build an edutainment application that aims on learning culture-specific behavior routines. In [7], we investigated communication management as one aspect of communication and in particular the usage of silence and overlapping speech. Therefore, we collected a video corpus in the two cultures Germany and Japan and analyzed it in regards of differences in communication management behavior. Two models reflecting prototypical Japanese and German behavior-styles were build and integrated into a multiagent application. In the system, prototypical Japanese behavior shows more pauses as well as more overlapping speech than the German version. In a pilot study, we found out that subjects perceive a difference between culture-specific agent dialogs.

For the work described in this paper, we follow that idea and want to add additional cultures to our research. Therefore, we analyze a video corpus collected in the US-American and Arabic cultures. In the future, we will thus be able to compare cultures belonging to the cultural groups of America, Arabia, Asia and Europe. Besides observing different cultures, we want to gain a deeper insight into how interpersonal communication is managed in different cultures. In particular, we aim on having a closer look at the quality of pauses and overlaps in speech in comparison to our previous work, where we conducted a quantitative analysis only. In addition, we want to have a closer look at communication management behaviors in combination with feedback behavior. We thus on the one hand follow the approach described in [7] in order to repeat the study with two additional cultures, on the other hand we enhance our computational models with a qualitative analysis in regards of where pauses and overlaps are located within the communication and how much verbal feedback is given in the different cultures.

3. THEORETICAL BACKGROUND

As we stated above, communication is managed differently across cultures. But how do we define communication management and where exactly do we expect culture-related differences? To clarify these questions, we explain in

the following subsections how interpersonal communication is coordinated and how different cultures can be classified according to well established definitions from the social sciences. Then we create profiles for the American and Arabic cultures and explain culture-related differences in communication management and feedback behavior for these two cultures.

3.1 Interpersonal coordination

Communication is managed using so-called regulators consisting of *vocalics*, *kinesics* and *oculesics* [17]. The first category (vocalics) includes verbal feedback signals: 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. Non-verbally, communication can be managed though hand gestures and body postures (kinesics) or eye and face gaze (oculesics).

According to Ting-Toomey [17], regulators are considered as culture-specific behaviors, that “act as the nonverbal traffic signs to control the flow of and pauses of conversations”. In addition, regulators are learned at a very young age and are thus used at a very low level of awareness. Ting-Toomey also states that discriminative use of regulators often causes inter-cultural distress and misunderstandings. However, people from contrastive cultures may not be able to name the reason for their frustration, since regulators are used subconsciously.

For the work described in this paper, we focus on verbal regulators, namely the usage of silence, overlaps in speech and verbal feedback behavior. Again, Ting Toomey [17] states that the beliefs expressed in talk and silence are highly dependent on culture. 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. The utterance “hai hai” simply expresses that the conversation partner is listening, while the literal translation “yes - yes” would transport an additional meaning. Besides the function of verbal feedback, the frequency and positioning within the conversation can vary across cultures.

3.2 Cultural profiles

In the social sciences culture is a well established research field. However, for our purposes some descriptions of culture are rather vague or abstract. In order to find more concrete definitions of culture, we are concentrating on cultural theories that explain cultures according to categories or dimensional models and that describe culture-related differences in behavior accordingly. As we stated earlier, for the work described in this paper, we focus on the American and Arabic cultures. In that vein, we concentrate on distinctions for these two cultural groups and introduce our expectations regarding differences in the management of communication.

The most well known dimensional model of culture was presented by Hofstede [11], whose theory is based on a broad empirical survey in which over 20 different cultures were categorized into a five dimensional model. Each dimension contains two extreme sides, for which he clearly defines stereotypical behavior norms. In [9], the Arabic World is described as one culture including the countries of Egypt,

Table 1: Hofstede’s scores on the dimensions Power Distance (PDI) and Individualism (IDV) for the two cultures America and Arabia.

Culture / Dimension	Arabic World	World Average	USA
PDI	80	55	40
IDV	38	64	91

Iraq, Kuwait, Lebanon, Libya, Saudi Arabia, and the United Arab Emirates. The scores for this culture on Hofstede’s five dimensions differ noticeable from the scores describing the United States as a culture. Most obvious are differences on the dimensions *Power Distance* and *Individualism*.

The Power Distance dimension describes the extent to which a different distribution of power is accepted by the less powerful members of a culture. This dimension is predominant in the Arabic culture and indicates a high level of inequality of power and wealth within the society. America on the other hand scores low on this dimension, which supposes greater equality between societal levels, including government, organizations and families.

The Individualism dimension 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. This is the predominant dimension for the United States, who are extreme individualistic, while it is the lowest dimension for the Arabic world. Table 1 shows the rankings on the Power Distance Index (PDI) and Individualism dimension (IDV) as defined by Hofstede for the American and Arabic cultures as well as the world’s average scores. Please note that these scores were normalized across all cultures to stay between 0 and 100.

Differences on Hofstede’s dimensions between the American and Arabic cultures are vast. But how does this correlate with their way of communicating and with the usage of regulators and verbal feedback in particular? In [12], Hofstede takes a look at the cultural dimensions in isolation and describes prototypical behavior for cultures that find themselves on one of the extreme sides. For collectivistic cultures, he states that silence may occur in conversations without creating tension, which does not hold true for individualistic cultures. Furthermore, the tone of voice as well as the usage of pauses can be a crucial feature in collectivistic cultures. We thus expect more silence in Arabic conversations than in American ones. High-power cultures are described as very verbal but usually soft-spoken and polite, while low-power cultures tend to talk freely in any social context. We therefore expect both American and Arabic cultures to speak a lot. In high-power cultures interpersonal synchrony is much more important than in low-power cultures, see [17]. To achieve interpersonal synchrony in a conversation, communication partners try to adapt to one another. We therefore expect more verbal feedback in Arabic conversations, as interlocutors want to show attentiveness while listening.

Another framework distinguishing cultures is proposed by [18], where cultures are divided into the three groups: Western, Latin and Oriental cultures. While America is considered a Western culture, the Arabic world would count as an Oriental culture, while in the Latin group we would find

countries such as Mexico for example. Western cultures are described as verbal cultures where members get nervous and uneasy when there are long pauses. Interrupting the conversation partner is considered as impolite. Thus, turn taking is managed in a way that one starts talking after the other stopped. Latin cultures are described as being even more verbal and interruptions are seen as showing interest in the conversation. Thus, interrupting the interlocutor will occur quite frequently. In Oriental cultures silence is much more important. While in Western cultures silence might be interpreted as failure to communicate, in Oriental cultures it is considered as a sign of respect. A pause can be used to process the information and assure that the conversation partner has finished his or her turn. We therefore presume pauses to be a turn-taking signal in Arabic conversations, while this does not hold true for American dialogs.

4. EMPIRICAL EVIDENCE

As we stated above, our aim is to build a multiagent system that demonstrates culture-specific communication styles for the two cultures America and Arabia. Findings in the literature (see Section 3.2) suggest that in stereotypical American conversations pauses in speech occur rarely, in contrast to Arabic conversations where pauses are used to handle turn-taking. Furthermore, we assume that in Arabic conversations verbal feedback is given more often, as interpersonal synchrony is more important in this culture.

But as these findings are rather tendencies than clear rules describing behaviors that can be implemented in a multiagent system, we analyzed a video corpus in the two cultures America and Arabia. That way, we want to gain some deeper insight in how communication is managed in these cultures.

4.1 Video Data

To ground our expectations about culture-related differences in communicative behavior into empirical data, we analyzed the video corpus recorded as part of a NSF-funded effort to explore cultural differences in dyadic communication (see acknowledgments). For the acquisition of the video corpus, subjects were invited in pairs in the American and Arabic cultures. The American setting took place at an American university, where mostly students participated. The Arabic corpus was recorded in native Arabic countries. The subject pairs knew each other in advance and had a personal relationship with each other, such as classmates or siblings. Since the video data was collected in every subject's county of origin, conversations were held in each's mother tongue.

In order to obtain a lot of feedback in the dialogs, one of the subjects watched a video in advance that he or she had to explain to the interlocutor during the recordings. The video contained the pear story which is established in intercultural research [4].

In that vein, a situation with a narrator and a listener evolved. To assure that the listener was active, subjects were told that the listener had to tell the story to the experimenter afterwards. In a total, we recorded more than 70 dialogs in all possible gender combinations. Figure 1 shows two examples of the recordings. Three videos were taped for each conversation, one showing both communication partners at once and a close-up on each of the participants' faces.



Figure 1: Example interactions from the video corpus (upper: Arabic; lower: America).

4.2 Analysis

In order to analyze the conversations from our video corpus, we used the audio signals. As Figure 1 shows, the subjects' voices were recorded using head-mounted microphones. Thus, we were able to separate the audio into two signals, one containing the speaker's speech and another containing the listener's speech. Since both speech-signals belonged to the same video originally (the dyad-video, see Figure 1 left) they were perfectly aligned and thus allowed the analysis of time-critical features such as silence (non of the subjects talking) or overlapping speech (both subjects talking at the same time).

The audio signals for speaker and listener were processed incrementally to find out if there is either speech or no speech using the following procedure. First, Praat [3] was used to extract the intensity of the speech signal frame by frame. As a next step, we built an intensity histogram with the intensity of each frame and updated a threshold from it incrementally. The speech-or-no-speech feature is determined by comparing the intensity of each frame with the threshold: if the intensity is larger than the threshold, it is considered speech; otherwise, it is not. Since there may be some short silent traces within an utterance, the frame whose intensity is less than the threshold might still be decided as speech, if such frames are within a speech segment and the number of them is less than a predefined number. In our experiment, we allowed gaps of up to 300 ms within speech segments.

Due to problems with the recordings, we had to discard some of the videos, e.g. videos where the speaker's voice was noticeable in both audio-channels. In that case, we did not get reliable features and therefore had to reduce the data-set for further analysis. In a total, we used 44 videos for our analysis.

Although all participants were given the same instructions, the length of the videos differed significantly across the two cultures. On average, American subjects talked one third longer than Arabic subjects. Consequently, all of the following results were normalized by the length of the videos.

As we were interested in feedback-behavior, we analyzed the listener's audios in isolation first. In that vein, we wanted to find out whether there are culture-related differences in

Table 2: Differences found in the corpus analysis for the two cultures America and Arabia

average occurrence (per minute)	America	Arabia	p
listener speech (total)	10.62	14.01	0.10
feedback (<= 1 sec)	8.02	10.86	0.08
silence (> 0.5 sec)	6.35	7.75	0.15
silence (> 1 sec)	1.34	2.16	0.09
same speaker	0.99	1.62	0.08
overlaps (per minute)	11.23	13.84	0.21
overlaps (per speech-floor)	1.08	0.99	0.16

the quantity of listener-activity. Our analysis revealed that Arabic listeners had the speaking floor more often than American listeners. In order to gain a deeper insight in how this tendency is related to feedback behavior, we had a closer look at very short segments of the listener’s speech. Every time listener-activity was detected in the audio-signal, the length was computed. We then categorized the speech-activity into feedback (<= 1 second) and speech (> 1 second). Although feedback does not necessarily have to be shorter than one second, this categorization helps filtering out short speech-segments such as "uh-huh" or "mm-hmm". Comparing the two cultures on those short segments, we found an even clearer tendency that Arabic subjects were more active than American subjects, although not significant (with a p-value = 0.08, using the two sided t-test). Table 2 shows a summary of the behavioral differences found in our corpus analysis.

To exclude the possibility that these findings were aroused by gender, we performed an inner-cultural analysis as well. Therefore, we divided the videos into female vs. male listeners and same vs. mixed gender combinations and compared the two groups within the two cultures. We did not find any significant results, either for the genders or for the gender combinations looking at both cultures separately. However, comparing the different groups across cultures (e.g. female American listeners vs. female Arabic listeners) we found the same tendency as for our analysis taking into account the whole data-set for each culture.

As a next step, we took into account the speakers’ audio as well. In order to find out how communication is managed between the interlocutors, we had a closer look at *silence* and *overlapping* speech as well as their correlations with observations made for verbal feedback as described above.

Analyzing silence in speech, we computed those traces where neither the listener nor the speaker spoke at a time. Whenever such a silent trace was detected, we examined the length. Following [7], we sorted out pauses in speech that result from breathing or the delay for deciding whether to give feedback or not. Consequently, for analysis we only took into account those pauses that lasted for more than 0.5 seconds. Comparing the two cultures we found a tendency that more silent traces occurred in the Arabic conversations than in the American ones. This is consistent with tendencies described in the literature (Section 3.2). Having a closer look at longer pauses (those that lasted for more than 1 second), the tendency was even stronger and almost significant (with a p-value = 0.09, using the two sided t-test, see Table 2).

After the quantitative analysis of counting the amount of

pauses in speech, we had a closer look at the quality of the detected pauses and their communicative function in particular. To find out whether a pause is used as a turn-taking signal, we analyzed *who* breaks the silence. Therefore, we distinguished if either the same speaker that had the speaking floor before took it again (pause within a turn), or if the other conversation partner took the floor (pause between two turns). Comparing the two cultures we found a tendency that it is much more common in American conversations that the same speaker breaks the silence than in the Arabic conversations (almost significant; with a p-value = 0.08, using the two sided t-test).

As we were interested in communication management, we also had a look at time spans where the interlocutors spoke simultaneously. We therefore computed those parts of the conversations where we detected both speaker and listener audio-signal at the same time, and refer to it as overlapping speech. Regarding culture related differences in the usage of overlapping speech, observations are less obvious than for pauses. As we stated above, the videos differed significantly by length across the two cultures. Normalizing the detected overlaps by the length of the videos, we found more overlaps by trend in the Arabic conversations than in the American conversations. However, normalizing by speech-floor (since we already found out that Arabic listeners give more feedback than American listeners), results were controversial and we found more overlaps in the American conversations than in the Arabic ones.

So far, only the quantity of time spans where both conversation partners spoke at the same time were analyzed. Pragmatics were not yet taken into account. As a next step, we had a closer look at feedback in combination with overlapping speech. Therefore, we divided the listener-feedback into overlapping feedback, feedback after a pause and other feedback. Comparing feedback behavior between the cultures, our analysis revealed that overlapping feedback is more common in American conversations than in Arabic ones. However, an inner-cultural analysis showed that overlapping feedback is the dominant behavior in both cultures. Since these observations do not reveal a clear tendency that can be assigned to cultural background, we did not consider overlapping speech in our computational model yet.

To simulate culture-related differences in communication management behavior, we thus focus on the amount of feedback as well as on the amount and positioning of pauses in speech (see Table 2, results highlighted in bold).

5. INTEGRATION INTO A MULTIAGENT SYSTEM

To simulate the tendencies described in the literature and verified by our empirical corpus study for the American and the Arabic cultures, we use the Virtual Beergarden running on the Horde3D GameEngine [1]. The multiagent application presents a virtual scenario where an arbitrary number of virtual agents are able to interact with each other.

For the demonstrator, we use virtual agents whose appearance is culturally as neutral as possible and agents could be either from an American or Arabic cultural background. Figure 2 shows a screenshot of the demonstrator. Both agents are dark-haired and not dressed in a culture-specific way. As a full body-view is shown, faces can not be observed in enough detail to estimate a certain cultural back-



Figure 2: Female and male agent interacting in the Virtual Beergarden.

ground. To avoid side effects aroused by gender, a mixed gender combination was chosen for the agent conversations. In that vein, a female and a male character interacted in the simulation.

To assure that human observers are not distracted by the virtual agents' non-verbal behaviors or assign meanings to these behaviors, they are reduced to a minimum. In that vein, the posture shown in Figure 2 is retained unchanged during the conversation and no hand gestures are used.

Another major distraction from the different models of interpersonal communication coordination could be the semantics of the speech. Thus, the content of the conversation needed to be as neutral as possible. Following [7], we used Gibberish for the evaluation study to analyze communication management behavior for virtual agents, generated by the Gibberish Generator [8]. Randomized output is generated out of an input text, that has the same statistical distribution of alphabetic characters or combinations of characters. As we recorded a speaker-listener situation for our corpus collection, we chose a well known fairy tale as sample input-text.

As we stated earlier, the analysis of our corpus revealed that more pauses in speech were found in Arabic conversations than in American ones. Accordingly, we placed more silence in the Arabic dialog model than in the American model. Besides this quantitative difference, we also found out that pauses can rather be found between turns in Arabic conversations and within turns in American communications. Following that tendency, we placed pauses in the agent dialogs between the turns in the Arabic version (other speaker is taking the turn) and within the turns in the American version (same speaker is continuing). In our demonstrator, the Arabic behavior model contains 3 pauses that are placed between the speaker turns and the American model contains two pauses that occur within a speaker's turn.

Analyzing feedback behavior in our corpus, we found more verbal feedback in the Arabic conversations than in the American ones. By trend, about double the verbal feedback was given in Arabic conversations. In our behavior models, we integrated verbal feedback twice in the American version and double the verbal feedback in the Arabic version (4x feedback).

The analysis for overlapping speech revealed more overlapping feedback in American conversations than in Arabic

ones. However, since this behavior is dominant in all cultures, we decided not to integrate these findings into our demonstrator yet.

Summing up, the two culture-specific behavior models contain differences in pause behavior and the usage of verbal feedback. In order to find out which of these differences do actually have an impact on human observers we decided to simulate them in isolation as well as in combination. Therefore, we created different versions of the culture-specific behavior models: one containing only the pause behavior, one containing only the feedback behavior and one containing both behaviors. In that vein, we are able to test our findings in isolation and refine our model if necessary.

6. PILOT STUDY

To obtain a first evidence whether the integrated findings of culture-related differences in communication management are appropriate and have the favored effect on human users, we designed a preliminary online evaluation. For the study, we recorded six videos of the same Gibberish dialog. As we stated earlier, the conversation was set up in a similar manner as the corpus recordings. To this end, one agent told a (Gibberish) story to another agent, that acts in the role of an active listener. Depending on the version of the behavior model, either American or Arabic pause or feedback behavior or both is added to the dialog. The plain dialog contains 9 sentences for the storyteller and 3 questions for the listener.

In order to get subjects acquainted with the virtual agents and the unusual situation of listening to a Gibberish dialog, we showed a neutral conversation first. In this video, the dialog described above was held without performing any culture-specific behavior. Although held in Gibberish, the same dialog was retained during the whole study to assure that the users' perceptions are not influenced by other linguistic features, such as the length of the sentences.

After watching the neutral video, subjects were shown three pairs of videos in alternating order. One pair contained the prototypical usage of pauses in speech for American versus Arabic communication. Another pair contained the different usage of giving verbal feedback and a third pair contained a combination of both behaviors.

For each pair of videos subjects were asked to judge:

1. Q1: Which video seemed the most natural?
2. Q2: Which group of agents would you rather like to join?
3. Q3: Which video appeared more unrealistic?
4. Q4: Which pair of agents do you think liked each other better?
5. Q5: Which pair of agents was more friendly with each other?

As these questions were asked in forced choice, subjects had to decide on their preferences for one of the simulated dialogs. In addition, we asked subjects to give a short statement on why they decided that way. Figure 3 shows a screenshot of our evaluation study, containing a pair of videos simulating two different culture-specific versions of the Gibberish dialog, as well as the questions and a field for further comments.



Figure 3: Screenshot of the evaluation study showing two videos with different behavioral models.

In a preliminary study, we evaluated 10 subjects: 8 American and 2 Arabic of whom 6 male and 4 female, between 20 and 40 years old. Table 3 shows an overview of the American subjects' ratings for the five questions mentioned above. American subjects significantly preferred the prototypical American pause behavior (with a p -value < 0.01 , using the two sided t -test) and thus judged this version as more realistic, more friendly and so forth (rating the Arabic version as more unrealistic respectively). For the second pair of videos, displaying different feedback behavior, tendencies are less obvious. To the contrary, American subjects preferred the Arabic version by trend (see Table 3). The combined version however, showed exactly the same tendency as for pause behavior. Thus, American subjects significantly preferred videos displaying the prototypical American version (with a p -value < 0.01 , using the two sided t -test). This suggests that the perception of different pause behavior was dominant over feedback behavior.

Since American subjects perceived the American version of pause behavior as more realistic, this tendency could result from the fact that the Arabic version was just not realistic in general. Although we only had two Arabic subjects (both born and raised in Egypt), interestingly both rated the Arabic pause version as more realistic (both for all five questions). Thus, the American pause version was not superior. We take this as a first evidence that our behavior model for the usage of pauses in speech does evoke different perceptions. As for American subjects, the Arabic subjects rated the feedback version designed for the other culture as more realistic and preferred the American version (both subjects for all five questions). In line with the observations for American subjects, Arabic subjects judged the videos showing the combined behavior the same way that they rated the videos showing the pause behavior. Thus, both subjects preferred the Arabic version for all five questions. This strengthens the idea that the pause behavior model effected subjects' estimations much more than our feedback behavior

Table 3: American subjects' preferences in our pilot study

	pause		feedback		both	
	USA	Arab	USA	Arab	USA	Arab
Q1	6	2	3	5	6	2
Q2	4	4	3	5	5	3
Q3	6	2	2	6	6	2
Q4	6	2	5	3	6	2
Q5	6	2	3	5	5	3
Sum	28	12	16	24	28	12
%	70%	30%	40%	60%	70%	30%

model.

Another interesting observation was, that subjects perceived the intensity of the agents' behavior very differently. While one of the subjects stated that the differences were almost too obvious and that there is a clear difference observable, another subject claimed that he could not see any differences in the videos. Although we can not assign this to culture, we take these comments as an evidence that there are differences in the perception of communication coordination behavior.

Summarizing our preliminary pilot study, we draw the conclusion that there are differences in the perception of different pause behaviors. Observations for estimations of feedback behavior lead to the conclusion that we need to analyze our data more carefully and refine our feedback behavior models. With this knowledge we can build a more sophisticated model of culture-specific communication management behaviors and evaluate them in a larger scaled study.

7. CONCLUSION

Following the idea of creating a platform where human observers can learn about different cultures through observation of selected behaviors, we integrated culture-related behaviors into a virtual agents system.

In this paper, we investigated interpersonal communication management as one aspect of culture-specific interaction. After defining culture and interpersonal communication management according to literature from the social sciences, we pointed out differences in these behaviors for the American and Arabic culture. A corpus analysis with human participants from the two cultures revealed culture-related differences in the usage of pauses in speech and in the amount of verbal feedback.

We integrated these findings into a demonstrator with virtual agents, that show either prototypical American or Arabic communication management behavior. Presenting our demonstrator to human participants, we have a first indication that the virtual agents' behavior is perceived differently. Although our observations are very preliminary and can not be assigned to be aroused by cultural-background yet, we claim that culture-related differences in behavior and their integration into multiagent systems is a promising research field.

So far, tendencies in the perception of different usage of silence in speech are promising. However for feedback behavior, tendencies are less clear. We therefore need to analyze our data again in order to build a more realistic model to simulate differences in feedback behavior. After refining our culture-specific behavior model, we plan on carrying out a

larger-scaled evaluation study in order to draw conclusions on the preferences of members of different cultures. In particular, we want to investigate whether human participants prefer communication management behaviors in dialogs between virtual agents that are in line with observations made for their own cultural background.

As we stated earlier, we were following up on the culture-specific perception study described in [7] and enhanced it with additional cultures as well as with a quantitative analysis, such as the location of silence in speech. As future work, we aim on combining the results from the two different corpus studies, as well as their integration into culture-specific behavior models for virtual agents. In that vein, we want to get a deeper insight into the differences in communication management behavior and the perception of agent behavior across several cultures that are located on different continents and therefore compare our findings for an American, an Asian, a European and a Middle-Eastern culture.

We consider our work as one step that helps to integrate cultural factors into virtual humans. In a long term view, the simulation of different behavior types can help human trainees to learn cultural awareness and in particular how to interpret behavioral clues shown by others as well as gaining knowledge that one's own behavior might be judged differently by others.

8. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 0729287 and the U.S. Army Research, Development, and Engineering Command (RDECOM). The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

The first author of this paper was also supported by a grant from the Elitenetzwerk Bayern (Elite Network Bavaria).

9. REFERENCES

- [1] Augsburg University. <http://mm-werkstatt.informatik.uni-augsburg.de/projects/gameengine>.
- [2] 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.
- [3] P. Boersma and D. Weenink. Praat: doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/>.
- [4] W. Chafe. The Pear Film. <http://www.linguistics.ucsb.edu/faculty/chafe/pearfilm.htm>.
- [5] M. Core, D. Traum, H. C. Lane, W. Swartout, J. Gratch, M. V. Lent, and S. Marsella. Teaching negotiation skills through practice and reflection with virtual humans. *Simulation*, 82(11):685–701, 2006.
- [6] J. Dias and A. Paiva. Feeling and Reasoning: a Computational Model. In Bento, Cardoso, and Dias, editors, *12th Portuguese Conf. on Artificial Intelligence (EPIA 2005)*, pages 127–140. Springer, 2005.
- [7] 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.
- [8] K. Enevoldsen. <http://thinkzone.wlonk.com/gibber/gibber.htm>.
- [9] G. Hofstede. <http://www.geert-hofstede.com/>.
- [10] G. Hofstede. *Cultures and Organizations - Intercultural Cooperation and Its Importance for Survival*. Profile Books, London, UK, 1991.
- [11] G. Hofstede. *Culture's Consequences - Comparing Values, Behaviours, Institutions, and Organizations Across Nations*. Sage Publications, 2001.
- [12] G. J. Hofstede, P. B. Pedersen, and G. Hofstede. *Exploring Culture - Exercises, Stories and Synthetic Cultures*. Intercultural Press, Yarmouth, United States, 2002.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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, *Proc. of 8th Int. Conf. on International Simulation and Gaming Association (ISAGA 2007): Organizing and Learning through Gaming and Simulation*, 2007.
- [17] S. Ting-Toomey. *Communicating across Cultures*. The Guilford Press, New York, United States, 1999.
- [18] F. Trompenaars and C. Hampden-Turner. *Riding the waves of culture - Understanding Cultural Diversity in Business*. Nicholas Brealey Publishing, London, UK, 1997.