Cultural Diversity for Virtual Characters (Extended Abstract)*

Birgit Endrass
Augsburg University
Germany
endrass@hcm-lab.de

Abstract

In human conversation, meaning is transported through several channels such as verbal and nonverbal behavior. Certain of these behavioral aspects are culturally dependent. Mutual understanding or acceptance is thus, amongst others, depended on the cultural background of the interlocutors. When designing virtual character behavior, culture should be considered as it may improve the character's acceptance by users of certain cultural backgrounds. This paper proposes a hybrid approach for the generation of culture-specific behaviors in a multiagent system. A computational model has been established by refining theoretical knowledge of culture-specific behavior with statistical data extracted from a video corpus of German and Japanese first-time meetings. Evaluation studies of such culturally enhanced virtual characters were conducted in both targeted cultures. Results indicate that human observers tend to prefer character behavior that was designed to resemble their own cultural background.

1 Introduction

Language is the most obvious barrier when people from different cultures communicate. According to Mehabian [1980] only 7% of human conversation is composed of the semantic content of verbal communication, while a vast part happens nonverbally (e.g., mimic or gestures) and paraverbally (e.g., accentuation or articulation). In addition to language, nonverbal and paraverbal behaviors are dependent on cultural background [Ting-Toomey, 1999]. Misunderstandings may occur due to different cultural backgrounds. An example of a culture-related behavioral difference includes the style in which conversations are managed. Interrupting the conversation partner, for example, is perceived as impolite in some cultures such as Germany, while it is judged positively as an increased interest in the ongoing conversation in other cultures, such as Hungary [Ting-Toomey, 1999]. A conversation between individuals of different cultures holds great potential for conflict or rejection of the communication partner. Similarly, virtual characters can be misunderstood due to different cultural backgrounds of the designer and the user. This can lead to decreased believability of the character or perceived empathy by the user. The work described here focuses on the generation of subtle culture-related behavioral traits for virtual characters rather than simulating obvious differences such as outer appearance or language.

Intelligent virtual agents are designed to resemble humans and to behave in a natural way. Through the great advancements in computer graphics and animation, virtual humans raise high expectations on human-like behavior. In the last decade considerable efforts have been made to enhance their believability by incorporating human factors such as personality [André *et al.*, 2000] or emotions [Gratch *et al.*, 2002]. Only recently, attempts have been made to integrate culture as a human factor. With the integration of cultural background, we aim on improving a character's acceptance by users of the targeted cultures.

The majority of approaches that investigate culture for virtual characters is model-based (e.g., [Johnson *et al.*, 2004; Aylett *et al.*, 2009; Nazir *et al.*, 2012; Kim *et al.*, 2009]), where theories from the social sciences are taken as a basis for computational models. That way, the causality of culture and corresponding behavior can be formalized in a generalizable manner. These approaches typically aim at creating cultural awareness by pointing out that behavior could be misinterpreted according to one's cultural background [Aylett *et al.*, 2009; Nazir *et al.*, 2012], or at training concrete culture-specific competencies such as communicating in foreign languages such as Iraqi [Johnson *et al.*, 2004] or negotiating effectively with foreign cultures [Kim *et al.*, 2009]. To test how simulated culture-specific differences influence the perception of human observers, studies were carried out, for example, by Koda *et al.* [2008] who investigate the different perception of facial expressions on avatars across cultures.

Data-driven approaches, on the other hand, use human data such as video corpora as a basis for simulation. So far, they have mainly been applied for the integration of individual factors [Martin *et al.*, 2005; Kipp *et al.*, 2007]. Data-driven approaches allow the extraction of concrete human behaviors but are hard to adapt to settings different from the ones recorded, as e.g., a certain individual was recorded in a cer-

*The dissertation on which this extended abstract is based was the recipient of the IFAAMAS-12 Victor Lesser Distinguished Dissertation Award [Endrass, 2012].

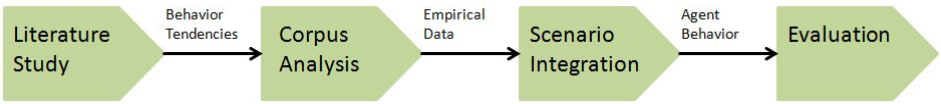

Figure 1: Workflow of the approach.

tain emotional state and later simulated with a virtual character. Applying this approach to culture is a challenging task as a large amount of data is needed to extract prototypical behavioral traits and distinguish cultural differences from other human factors such as personality.

2 Approach

In this work, we present a hybrid approach that combines the advantages of a model-driven (e.g., the intuitive modeling of causality between culture and behavior) and a data-driven approach (e.g., extracting examples of concrete human behavior). By analyzing theoretical differences along concrete examples of human behavior, structured and statistical data is extracted that extends models described in the literature. Thus, social sciences teach us which behavioral aspects are of interest when building a model that describes culture-specific behavior, while we learn from empirical data how differences in these aspects manifest themselves. As outlined in Figure 1 several steps were taken for the approach.

As a first step, we studied literature from the social sciences to extract culture-related differences in behavior for the two cultures of Germany and Japan. In a second phase, a video corpus was recorded in the two cultures and analyzed taking into account the expectations on behavioral differences. Using our findings from literature and corpus studies, we subsequently built computational models that generate culture-specific behaviors and integrated them into a demonstrator including virtual characters. Finally, we evaluated the implemented aspects of behavior to investigate their impact on human observers' preferences.

3 Theoretical and Empirical Background

Often information about cultural differences is stated descriptively in the literature and is thus hard to formalize in computational models. Exceptions include dimensional approaches such as Hofstede's six-dimensional model [Hofstede *et al.*, 2010] and cultural dichotomies [Hall, 1966; 1983], as they classify culture and corresponding stereotypical behavior for given dimensions or groups. The Individualism dimension of Hofstede's model [Hofstede *et al.*, 2010], for example, explains that for individualistic cultures such as Germany the individual is important and that members tend to speak out their own mind and stand visually free in groups. For collectivistic cultures such as Japan, vice versa, the group is important and the opinion of the whole group counts. Members of such cultures tend to e.g., have close physical contact to in-group members.

We used different theories (e.g., [Hofstede *et al.*, 2010; Hall, 1966; 1983; Ting-Toomey, 1999; Trompenaars and Hampden-Turner, 1997]) to categorize the German and Japanese cultures. Based on these cultural profiles, we stated expectations on culture-related behavioral differences on a selected set of aspects of human behavior (see Table 1). The set covers aspects of verbal behavior, communication management and nonverbal behaviors, and was selected since the aspects are influenced by cultural background and meet the modalities that can be simulated with virtual characters.

behavioral aspect	Germany	Japan
topic selection	more private	less private
conversational flow	sequential	alternating
pauses	avoided	consciously used
overlaps	uncommon	common during feedback
gesture types	direct	metaphoric
expressivity	more expressive	less expressive

Table 1: Summary of behavioral expectations for the German and Japanese cultures.

For empirical grounding, a scenario of the video corpus recorded for the Cube-G project [Rehm *et al.*, 2007] was analyzed. The decisive question was whether concrete cultural characteristics can be extracted from a data set containing 20 first-time meetings recorded in each culture. Therefore, the set of behavioral aspects had to be parametrized for further processing. We developed an annotation schema based on established schemes and theories [Core and Allen, 1997; Schneider, 1988; McNeill, 1992; Hartmann *et al.*, 2006; Bull, 1987] containing variables for each of the relevant aspects. To analyze gestural expressivity, for example, we took the parameters spatial extent, speed, power, fluidity and repetition into account. After annotating the data set using the Anvil tool [Kipp, 2001], a statistical analysis was conducted. Our results [Endrass *et al.*, 2013] were in line with the expectations derived from the research literature and, except for overlapping speech, revealed statistically significant differences on all behavioral aspects of the selected set of behaviors: Regarding verbal behavior, we found significantly more personal topics (e.g., hobbies) in the German conversations, while we found significantly more topics covering the social (e.g., movies) and the immediate situation (e.g., setup of the study) in the Japanese conversations. In addition, the analysis revealed more topic shifts in the Japanese conversations compared to the German ones. The analysis of communication management behaviors showed significantly more pauses in speech in the Japanese conversations. Our analysis of gestural expressivity revealed that gestures were performed faster, more powerfully, more fluently and with a larger spatial extent in the German videos than in the Japanese ones, while the stroke of a gesture was repeated more often in the Japanese conversations. The corpus data also showed differences in the usage of body postures. Interestingly, postures that regularly

occurred in one culture barely occurred in the other culture.

The corpus analysis extended findings from the literature with structured data as well as statistical information on the quantity of the occurrence of the behavioral aspects. For example, we derived from the literature that silence in speech is creating tension in some cultures such as Germany and is thus tried to be avoided, while it is used as a sign of respect in Japan where it is consciously used [Ting-Toomey, 1999]. From the corpus analysis, we learned e.g., that pauses lasting for over two seconds occurred more often in the recorded Japanese conversations (on average 9.18 times in conversations that lasted for five minutes) compared to the German ones (on average 0.52 times in conversations that lasted for five minutes) [Endrass, 2012].

4 Design and Implementation

For the generation process, culture had to be integrated as a parameter of behavior. We identified approaches that had already overcome similar challenges successfully for other human factors such as personality, although they have not been applied to culture yet. Incorporating culture is a challenging task as group effects have to modeled while keeping individual factors.

In the domain of dialog generation, plan-based approaches have been proven to be very well suited for the production of goal-directed dialog utterances and have, for example, been applied to the parametrization of personality [André *et al.*, 2000]. Plan-based approach provide certain advantages such as flexibility and expandability. By planning the dialog flow, we are able to structure utterances on an abstract level independent from the concrete wording. In that manner templates for additional topics that are tagged according to the categorization can simply be added to the knowledge base, while the general planning algorithm stays the same. For the implementation of verbal aspects, we adapted a hierarchical planning system [Nau *et al.*, 1999] for our needs to allow autonomous dialog planning for each character individually. For the realization of culture-specific topic selection during first-time meetings, possible topics were categorized based on our finding from the literature as well as the empirical study. Our planing algorithm determines the likeliness of a certain topic category dependent on an agent's cultural background to e.g., generate more private topics for German conversations. To ensure non-monotonic behaviors, we added personal motivations that also influence the process of behavior selection. In addition, the flow of a conversation is influenced by cultural background. In particular, we defined how difficult it is with a given cultural background to drop a topic and with it interrupt the conversational flow by e.g., introducing a new topic.

Regarding the adaption of nonverbal behaviors, approaches using Bayesian networks have been successfully applied, e.g., for the generation of gestures supporting spatial information [Bergmann and Kopp, 2009]. For the integration of cultural differences, we consider this approach especially well suited as it allows dealing with uncertain knowledge resulting from the fact that there is no deterministic mapping between cultural background and nonverbal behaviors. Thus, it enables us to customize behaviors to a particular culture without giving up a certain amount of variability that is necessary to ensure that a character is perceived as an individual. We modeled a Bayesian network to determine the most probable nonverbal expressivity using as a basis Hofstede's dimensional model of culture [Hofstede *et al.*, 2010]. For simulation in a virtual environment [Damian *et al.*, 2011], animations are customized to match different levels of expressivity by e.g., playing them with a higher speed or spatial extent. In some cases, culture-specific behaviors cannot be generated by customizing behaviors because culture is also reflected by specific nonverbal behaviors that need to be executed accurately. To account for this fact, we augmented customizable behaviors by behaviors extracted directly from the corpus. This was, for example, realized for prototypical body postures. Figure 2 illustrates prototypical body postures as observed in the empirical study and modeled in our virtual scenario.

5 Evaluation

To measure the impact of each implemented aspect on the users' preferences, evaluation studies were conducted in Germany and Japan [Endrass *et al.*, 2013]. We designed distinct studies for the different behavioral channels to exclude correlations, for example, by using a fantasy language to simulate communication management and nonverbal behaviors without adding any semantic content of speech.

Two different versions of each evaluation study were designed, where the outer appearance of the characters as well as the text-to-speech systems were adapted to meet the participants' cultural background. Due to the similarity principle [Byrne, 1971], which states that interaction partners who perceive themselves as being similar are more likely to like each other, we hypothesized in each of the studies that human observers prefer character behavior that was designed to resemble their own cultural background.

The analysis of the evaluation studies [Endrass *et al.*, 2013] revealed statistically significant results for the following behavioral aspects:

Results of the study investigating verbal behavior in first-time meetings by means of differences in topic selection revealed that German observers found German versions of the dialogs significantly more appropriate and interesting, would rather like to join the conversation and think that agents get along with each other better. The Japanese part of the evaluation study revealed that Japanese participants found the Japanese versions significantly more appropriate and interesting and thought that agents were getting along with each other better. Regarding communication management behaviors and nonverbal behaviors, differences in the usage of pauses in speech, overlapping speech, gestural speed, spatial extent and postures were taken into account, each aspect tested in isolation. The evaluation conducted in Germany revealed that participants significantly preferred the version that resembled behavior observed for their own cultural background for overlapping speech and the spatial extent of gestures. For all other aspects, participants preferred the German versions at least by trend. In the Japanese study, we found that Japanese participants significantly preferred postures designed for their

Figure 2: Prototypical body postures observed in the Japanese (left) and German (right) video data (upper) and simulated with virtual characters (lower).

cultural background. We only observed a controversial trend for pauses in speech and overlapping speech, indicating that Japanese participants preferred the German version (although not significant).

6 Conclusions

This extended abstract presents the integration of culture-related behaviors into a multiagent system using a hybrid approach, along with its impact on user preferences of simulated behaviors.

For the behavioral aspects where our hypotheses were confirmed, we consider their attention when designing virtual character behavior as promising. We aim on contributing to the field of intelligent virtual agents by providing our findings that can help improve a character's acceptance by users of certain cultural backgrounds.

To demonstrate the general nature of our approach, a study was carried out on the US American and Arab cultures in the context of communication management behaviors [Endrass *et al.*, 2010]. Results confirmed that human observers seem to prefer agent behavior designed for their own cultural background.

In our future work, we aim on refining our models based on our previous findings. The Japanese evaluation study on overlapping speech, for example, did not show the intended results. One reason for this outcome might be the missing semantics of the agent dialogs which was chosen for the study design to avoid different factors influencing each other. We thus aim on adding semantic context such as feedback behavior. The set of behavioral aspects described here contains different modalities of behavior but remains limited. We aim on extending the set by adding more aspects that are considered as culture-depended such as head nods or leg and feet postures.

So far the behavioral models are based on statistical findings of the annotated corpus only. As a next step, we also aim on applying machine learning approaches by, e.g., learning the most probable culture-specific behavior from our video data for a Bayesian network that was structured based on tendencies derived from the research literature.

Acknowledgments

First and foremost, I would like to thank my supervisor, Prof. Dr. Elisabeth André, for her constant support and encouragement during this dissertation [Endrass, 2012]. My special thanks go to Prof. Dr. Matthias Rehm, my co-supervisor, for his constant help throughout the years. Parts of this dissertation were conducted in scope of the CUBE-G project funded by the German research foundation (DFG) other parts were funded by a scholarship of the Elite Network Bavaria (Elitenetzwerk Bayern). Last but not least, I want to express my thanks to our Japanese cooperation partners Yukiko Nakano and Toyoaki Nishida.

References

[André *et al.*, 2000] Elisabeth André, Martin Klesen, Patrick Gebhard, Steve Allen, and Thomas Rist. Integrating Models of Personality and Emotions into Lifelike Characters. In *Affective Interactions*, pages 150–165. Springer, 2000.

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

[Bergmann and Kopp, 2009] K. Bergmann and S. Kopp. Bayesian Decision Networks for Iconic Gesture Generation. In *Proc. of 9th Int. Conf. on Intelligent Virtual Agents (IVA 2009)*, pages 76–89. Springer, 2009.

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

[Byrne, 1971] D. Byrne. *The attraction paradigm*. Academic Press, New York, 1971.

[Core and Allen, 1997] Mark Core and James Allen. Coding Dialogs with the DAMSL Annotation Scheme. In *Working Notes of AAAI Fall Symposium on Communicative Action in Humans and Machines*, 1997.

[Damian *et al.*, 2011] Ionut Damian, Birgit Endrass, Peter Huber, Nikolaus Bee, and Elisabeth André. Individualized Agent Interactions. In *Proc. of 4th Int. Conf. on Motion in Games (MIG 2011)*, 2011.

[Endrass *et al.*, 2010] 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)*, pages 99–108, 2010.

[Endrass *et al.*, 2013] Birgit Endrass, Elisabeth André, Matthias Rehm, and Yukiko Nakano. Investigating culture-related aspects of behavior for virtual characters. *Autonomous Agents and Multi-Agent Systems*, 2013.

[Endrass, 2012] Birgit Endrass. *Cultural Diversity for Virtual Characters: Investigating Behavioral Aspects across Cultures*. PhD thesis, Augsburg University, Germany, 2012.

[Gratch *et al.*, 2002] J. Gratch, J. Rickel, E. André, N. Badler, J. Cassell, and E. Petajan. Creating Interactive Virtual Humans: Some Assembly Required. *IEEE Intelligent Systems*, 17(4):54–63, 2002.

[Hall, 1966] Edward T. Hall. *The Hidden Dimension*. Doubleday, 1966.

[Hall, 1983] E.-T. Hall. *The Dance of Life*. Doubleday, New York, 1983.

[Hartmann *et al.*, 2006] Bjorn Hartmann, Maurizio Mancini, and Catherine Pelachaud. Implementing expressive gesture synthesis for embodied conversational agents. In *Gesture Workshop (GW 2005)*, number 3881 in LNAI, pages 188–199. Springer, 2006.

[Hofstede *et al.*, 2010] Geert Hofstede, Gert-Jan Hofstede, and Michael Minkov. *Cultures and Organisations. SOFTWARE OF THE MIND. Intercultural Cooperation and its Importance for Survival*. McGraw Hill, 2010.

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

[Kim *et al.*, 2009] J. Kim, R.-W. Hill, P. Durlach, H.-C. Lane, E. Forbell, M. Core, S. Marsella, D. Pynadath, and J. Hart. BiLAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education*, 19:289–308, 2009. IOS Press.

[Kipp *et al.*, 2007] Michael Kipp, Michael Neff, Kerstin-H. Kipp, and Irene Albrecht. Towards Natural Gesture Synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In *Proc. of 7th Int. Conf. on Intelligent Virtual Agents (IVA 2007)*, pages 15–28. Springer, 2007.

[Kipp, 2001] M. Kipp. Anvil - A Generic Annotation Tool for Multimodal Dialogue. In *Eurospeech 2001*, pages 1367–1370, 2001.

[Koda *et al.*, 2008] T. Koda, M. Rehm, and E. André. Cross-Cultural Evaluations of Avatar Facial Expressions Designed by Western Designers. In *Proc. of 8th Int. Conf. on Intelligent Virtual Agents (IVA 2008)*, pages 245–252. Springer, 2008.

[Martin *et al.*, 2005] Jean-Claude Martin, Sarkis Abrilian, Laurence Devillers, Myriam Lamolle, Maurizio Mancini, and Catherine Pelachaud. Levels of Representation in the Annotation of Emotion for the Specification of Expressivity in ECAs. In *Proc. of 5th Int. Conf. on Intelligent Virtual Agents (IVA 2005)*, pages 405–417. Springer, 2005.

[McNeill, 1992] David McNeill. *Hand and Mind What Gestures Reveal about Thought*. University of Chicago Press, Chicago, London, 1992.

[Mehabian, 1980] A. Mehabian. *Silent messages: Implicit communication of emotions and attitudes*. Wadsworth, 2 edition edition, 1980.

[Nau *et al.*, 1999] Dana Nau, Yue Cao, Amnon Lotem, and H. Munoz-Avila. SHOP: Simple Hierarchical Ordered Planner. In *Proc. of 16th Int. Joint Conf. on Artificial Intelligence (IJCAI 1999)*, pages 968–973, 1999.

[Nazir *et al.*, 2012] Asad Nazir, Christopher Ritter, Ruth Aylett, Eva Krumhuber, Aleksandra Swiderska, Nick Degens, Birgit Endrass, Colette Hume, John Hodgson, and Samuel Mascarenhas. ECUTE: DIFFERENCE IS GOOD! In *Proc. of the IADIS International Conference e-Learning*, pages 425–429, 2012.

[Rehm *et al.*, 2007] Matthias Rehm, Elisabeth André, Yukiko Nakano, Toyoaki Nishida, Nikolaus Bee, Birgit Endrass, Hung-Hsuan Huan, and Michael Wissner. The CUBE-G approach - Coaching culture-specific nonverbal behavior by virtual agents. In *ISAGA 2007: Organizing and Learning through Gaming and Simulation*, 2007.

[Schneider, 1988] K. P. Schneider. *Small Talk: Analysing Phatic Discourse*. Hitzeroth, Marburg, 1988.

[Ting-Toomey, 1999] S. Ting-Toomey. *Communicating across cultures*. The Guilford Press, New York, 1999.

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