Some Pitfalls for Developing Enculturated Conversational Agents

Matthias Rehm[1], Elisabeth André[1], and Yukiko Nakano[2]

[1] Faculty of Applied Informatics
University of Augsburg
{rehm,andre}@informatik.uni-augsburg.de
[2] Faculty of Science and Technology
Seikei University Tokyo
y.nakano@st.seikei.ac.jp

Abstract. A review of current agent-based systems exemplifies that a Western perspective is predominant in the field. But as conversational agents focus on rich multimodal interactive behaviors that underlie face-to-face encounters, it is indispensable to incorporate cultural heuristics of such behaviors into the system. In this paper we examine some of the pitfalls that arise in developing such systems.

Keywords: Embodied Conversational Agents, Cultural Heuristics, Multimodal Interaction.

1 Introduction

This paper argues that Embodied Conversational Agents (ECAs) [5] are prototypical devices for enculturating the human computer interface. It examines the standard development process for ECA systems and discusses at each step the pitfalls that arise from integrating culture as a computational parameter into the process. The paper is not going to argue for or against specific cultural theories, but relies on Hofstede's [11] dimensional theory of culture as a widely used example.

Embodied conversational agents can be regarded as a special case of multimodal dynamic interactive systems (see Figure 1 for some examples). They promote the idea that humans, rather than interacting with tools prefer to interact with an artifact that possesses some human-like qualities – at least in a large number of application domains. If it is true, as Reeves and Nass' [24] media equation suggests, that people respond to computers as if they were humans, then there are good chances that people are also willing to form social relationships with virtual personalities. As a consequence, it seems inevitable to take cultural aspects into account when creating such agents. Due to their embodiment, agents present complex multimodal systems with rich verbal and nonverbal repertoires. Additionally, the appearance of the agent might play an important role when taking cultural aspects into account.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2009, LNCS 5612, pp. 340–348, 2009.

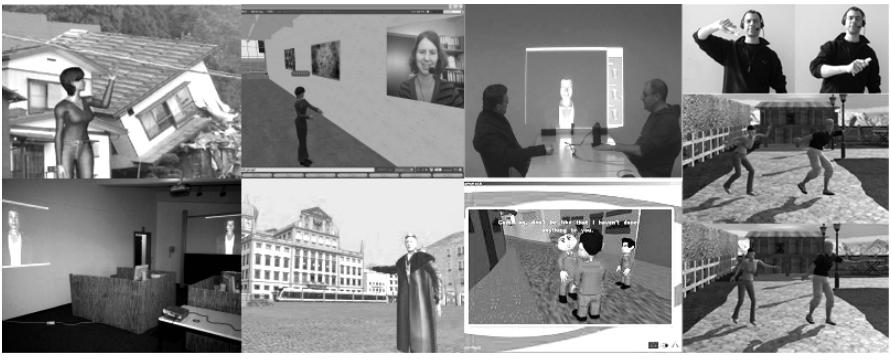

Fig. 1. Examples of Embodied Conversational Agent. Top row: the IPOC earthquake instructor [21], an autonomous bot in Second Life [27], the Gamble multiuser dice game [25], interacting with virtual dancers [30]. Bottom row: collaborating agents in edutainment [28], a virtual tourist guide, the FearNot! anti-bullying system [10].

Embodied Conversational Agents as an interface metaphor have a great potential to realize cultural aspects of behavior in several fields of human computer interaction:

1. Information presentation: By adapting their communication style to the culturally dominant persuasion strategy, agents become more efficient in delivering information or selling a point or a product.
2. Entertainment: Endowing characters in games with their own cultural background has two advantages. It makes the game more entertaining i.) by providing coherent behavior modifications based on the cultural background and ii.) by letting the characters react in a believable way to (for them) weird behavior of other agents and the user.
3. Education: For educational purposes, experience-based role-plays become possible, e.g. for increasing cultural awareness of users or for augmenting the standard language textbook with behavioral learning.

Two main issues for enculturating embodied conversational agents are discussed in this paper:

1. Enculturating agents opens up a challenging research field because culture penetrates most of the above mentioned features (verbal and nonverbal behavioral, appearance) of an agent. Thus, enculturating such a system has to rely on a solid theoretical framework that is able to describe or even predict these influences.
2. Moreover, the developers' own cultural background provides them with implicit design heuristics for the system, which have to be challenged actively at every step of the process.

These issues are addressed in relation to the methodological approach for realizing ECA systems.

2 Designing ECA Systems

The methodological approach for modeling the behavior for embodied conversational agents is well exemplified by the following development steps:

- Study: To build a formal model for generating realistic agent behaviors, data of human interactions is necessary for two reasons: (i) it serves as an empirical foundation for the formal models of human agent interaction, and (ii) it serves as a benchmark against which these models are evaluated. In most cases, formal models are not built from scratch. Rather, the data analysis serves to refine existing models found in the literature. Such models often lack explicit information necessary for the integration in an agent system like synchronization and timing of modalities. Over the last decade, numerous work has established the area of multimodal corpus analysis to shed light on the specifics of multimodal interaction. To give some examples, [16] suggest an annotation scheme for gestures that draws on the distinction between the temporal course of a gesture and its type and relies on a gesture typology introduced by [20]. [1] as well as [7] annotate instead the expressivity dimensions of gestural activity focusing on how a gesture is accomplished and not on what kind of gesture is used. [26] describe an annotation scheme that analyzes gestures on a more abstract functional level. Their corpus captures the relation between linguistic and nonverbal strategies of politeness.
- Model: The data gathered in the previous step of the development process serves as the foundation of a formal model of human agent interaction. [4] give an account on how the data from such a corpus can be used to directly mirror the behavior of a human speaker with an agent. A similar approach is described by [16], who extract information of personal idiosyncrasies of the human speaker, which is then mimicked by the agent. [19] extract statistical rules from a corpus of natural dialogues that allow them to generate appropriate head and hand gestures for their agent that accompany the agent's utterances. Instead of rules, [26] have shown how statistical information can be extracted from a multimodal corpus and used as control parameters for a virtual character. To this end they analyzed what kind of relation exists between certain types of gestures and verbal strategies of politeness. The resulting models of human-human conversational behavior then serve as a basis for the implementation of ECAs that replicate the behaviors addressed by the models.
- Test: To evaluate the resulting system, experiments are set up in which humans are confronted with ECAs following the model. The data collected in the first step can serve as a baseline against which the resulting ECA implementation can be tested. [6] as well as [23] exemplify this use of multimodal corpora in developing agents that exhibit human turn taking behavior and human grounding behavior respectively. [25] uses instead a corpus of human agent interactions to exemplify how design guidelines can be derived on this basis for such interactive systems.

The above mentioned work concentrated on the challenge of realizing natural interaction behaviors for agent systems but did not acknowledge culture as a relevant parameter that might influence such interactions. As long as the cultural background of the users of such systems is identical to the developer's background, this does not pose a problem as both work with the same culturally determined heuristics for generating and interpreting behavior. In the next section, we present work that tries to explicate cultural influences in order to allow for adapting the behavior of agents to the user's cultural background.

3 Related Work

Compared to the systems described in the last section, culture adds another layer of complexity to the endeavor of modeling the behavior of conversational agents. Prominent approaches that embrace this challenge are so far primarily located in the area of intelligent tutoring systems.

The tactical language initiative [14] aims at coaching US soldiers in culture specific language skills. Obviously, the target domain is the Middle East. Users have to use the right phrases and select appropriate co-verbal gestures in order to achieve their goals, e.g. to persuade a doctor to move his hospital somewhere else. The same user group and cultural domain is focused in [18], who present a tutoring system to teach social norms in negotiation scenarios. [15] examine cultural differences in persuasion strategies and present an approach of incorporating these insights into a persuasive game for a collectivist society. Whereas the presented systems aim at simulating real cultures, the Orient system [3] considers a virtual culture instead, targeting teenagers as a user group and trying to increase the awareness of cultural differences in this age group.

These systems explicitly model cultural behavior of the agents for a given domain. What is lacking so far is a principled approach of considering culture as a computational parameter. [13] present a first approach of modifying the behavior of characters by cultural variables relying on Hofstede's [11] dimensions. The variables are set manually in their system to simulate the behavior of a group of characters. [29] aim at automatically adapting to the user's cultural background by setting appropriate parameters for the nonverbal behavior of the agents. To this end they employ Bayesian networks that model the causal relations between cultural dimensions and nonverbal behavior. [12] investigate the relative importance of appearance and verbal as well as nonverbal behavior to attribute a specific culture to an agent and find evidence that consistent behavior can override the cultural background implied by the agent's appearance.

To sum up, the importance of cultural influences on the interactive behavior of agents have been acknowledge but there are only few approaches that go beyond the explicit modeling of a specific culture for a clearly determined application scenario. One reason might be the difficulty of pinning down the influences of culture on the development process of agent-based systems. In the remainder of this paper, we exemplify the problems of developing truly enculturated agents.

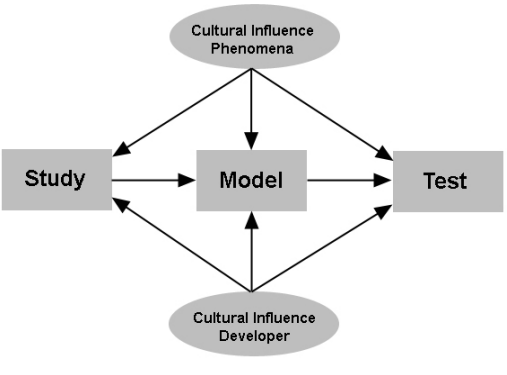

Fig. 2. Cultural influences on the development of ECAs

4 Culture in the Development Process

Introducing culture into the development process poses challenges on two levels. (i) The development should be grounded in a theoretical framework that is able to explain and ideally predict behavior based on the features of a specific culture. This would allow realizing a parameter-based model of cultural influences in order to simulate culture-specific behavior without having to develop a completely new agent for every culture. (ii) The developer(s) own culture has to be kept at bay as it provides implicit design heuristics that have to be actively challenged at every step of the development process.

The rest of this section will address these two challenges for the first two steps of the development process. Figure 2 exemplifies the interrelation of cultural aspects that interfere with this process either from the phenomena under examination or from the developers own cultural background.

4.1 Study

The most appropriate theoretical framework for our endeavor seems to be a theory that defines culture as norms and values, i.e. heuristics for behavior. A number of approaches exist for this line of thinking ([9], [17], [31]). One that is prominent and widely used is Hofstede's [11] dimensional theory of culture. Based on a broad empirical survey Hofstede defines culture as a dimensional concept, where a given culture occupies a certain area on each dimension. Correlated with the locations on these five dimensions are heuristics on how to behave "properly" in the given culture. The five dimensions are hierarchy, identity, gender, uncertainty, and orientation. We will not go into detail here but give an example on possible correlations between dimension and behavioral heuristics. The identity dimension e.g. is tightly related to the expression of emotions and the acceptable emotional displays in a culture. Thus, it is more acceptable in individualistic cultures like the US to publicly display strong emotions than it is in collectivistic cultures like Japan [8]. Uncertainty avoidance like identity is directly related to the expression of emotions. In uncertainty accepting societies, the facial expressions of sadness and fear are easily readable by others whereas in uncertainty avoiding societies the nature of emotions is less accurately readable by others, which was shown by [2]. It has to be noted that Hofstede's theory is not without controversy. His theory is based on a large-scale questionnaire study with IBM employees, which constitutes a strong selection bias on the results. Nevertheless, Hofstede's theory has a great appeal because of its quantitative nature. Although the theory describes certain correlations between cultural dimensions and correlated behavioral heuristics, this attribution is not unambiguous as the correlated heuristics might contradict each other on different dimensions. Consider for instance the following example dealing with proxemics. High power distance (hierarchy) might result in standing further apart in face-to-face encounters whereas collectivism (identity) generally means standing closer together in the same situation. Both attributions hold true for the Japanese culture. Thus, what will be the result of these correlations if they are combined? Solutions of different complexity can be thought of. Interlocutors could position themselves simply in a mean distance. Or we could define a hierarchical relation between the dimensions resulting in some information being overridden or weighted differently. More sensible would be a contextual adaptation

that takes the semantics of the dimensional position into account. If a culture has a high power distance then there could be differences in proxemics behavior that are related to social status, for instance resulting in standing further away from high status individuals but closer together with peers.

What is apparent from these examples is one obvious conclusion. To adapt the behavior of agents to cultural heuristics it is indispensable to gain insights into how these differences manifest in face-to-face encounters. Unfortunately, there is a lack of reliable cross-cultural data as the information in the literature is often of an anecdotal character, or lacks technical information that is necessary to realize an interactive system. One way to deal with this problem is to gather data in a standardized way, tailored to the modeling endeavor. In [22] in this volume, we describe such an approach for the German and Japanese culture.

Whereas sometimes the developer's intuition might work due to the fact that the developer can take his own actions as a model for building the interactive behavior of an ECA, this is quite problematic if designing for a different culture. The developer's own cultural norms and heuristics hinder this process in making quite specific aspects of behavior relevant that might be irrelevant in a different culture. Consider the following example. If studying turn-taking behavior in Germany, the effect will be to consider an ordered exchange between interlocutors with little overlap as the basic form of discussion. But in other cultures it is more common to have strong overlaps and simultaneous turns in discussions to emphasize one's interest in the topic [32]. Thus, investigating turn-taking behavior in Italy might result in a completely different model of turn-taking behavior. Thus, even when being aware of cultural differences does not necessarily help in identifying relevant behaviors. An obvious solution to this problem would be to always involve developers from the targeted cultures in the development process. This might only be feasible for large-scale projects. A low-budget solution would be to discuss most of the design choices as often as possible with someone from the target country. To do so, it is important to make one's own design choices explicit. As the underlying heuristics are implicit and generally interpreted as the "natural" way to do things, this might not be too easy. One way of solving this problem could be to develop some best-practice advices on how to check for cultural issues in the design of the system.

4.2 Model

If we roughly sketch the process of behavior selection and generation in an agent system, it becomes obvious that culture penetrates most stages of this process. Figure 3 gives a simplified impression of some of the main processing steps. In the planning stage, culture provides scripts and rituals for interactions. One of the most fundamental situations in this respect is a first meeting encounter. According to [2], a first meeting is a ritual that follows pre-defined scripts. [32] follows this analysis by denoting a first meeting as a ceremony with a specific chain of actions. Behavior selection is concerned with enriching the dialogue step with suitable verbal and non-verbal behavior. Consider the use of gestures as an example. Culture influences the selection process on different levels. On the one hand, it is necessary to choose the right gesture type and animation for the utterance. This repertoire of available gestures is at least partially culture-specific as there are sets of language and thus

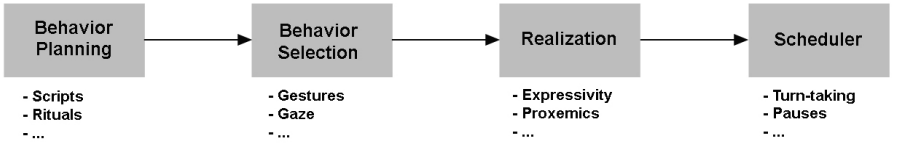

Fig. 3. Culture influences during the behavior generation process

culture-specific emblematic gestures. On the other hand, if and how many gestures are employed in an utterance differ widely between cultures. The Italian culture for instance has a rich repertoire of emblematic gestures and gestures in general are used frequently in face-to-face encounters. Quite the opposite is true for the German culture. In the realization stage another influence of culture comes into play. Consider again gestural activity. Whereas one culture gestures fast and frequently, taking much space in doing so, other cultures make only use of infrequent gestures that do not intrude the space of the interlocutor. The scheduling stage at last is necessary to ensure appropriate timing in turn-taking of the interlocutors, which again is culture-specific. For instance in the above mentioned study on German and Japanese behavior [22] we found that German interlocutors are generally uncomfortable with longer pauses in conversations compared to the Japanese samples.

One suggestion to deal with this ubiquitous influence of culture is presented in [29]. By modeling the causal relations between a culture's location on Hofstede's dimensions and correlated behavior in a probabilistic network, it becomes possible to extract different types of cultural influences from this network. The different layers of the network can serve as influence at different steps of the generation process. Another idea is presented in [3]. In a close analogy to the Chomskian ideas of language use, a universal behavior selection process is realized, which is augmented with culture-specific transformation rules for perceptions and actions.

Again, the cultural background of the developer supplies heuristics on what is interpreted as relevant or typical behavior. At this point, this check is necessary on different layers of abstraction. The developer's background may bias how the data derived in the previous step is used to model the behavior of an ECA. The definition of objective criteria is a necessary prerequisite for a reliable analysis. Actually building the ECA based on the analysis and the model suffers from the same pitfalls as before. What is an unimportant variation in gestural expressivity in one culture might lead to severe misunderstandings in a second culture. The same suggestions presented in the previous section apply here.

5 Conclusion

In this paper we examined a number of pitfalls one could stumble in while developing enculturated conversational agents and presented some ideas on strategies to prevent these pitfalls. These pitfalls originate from different sources. On the one hand trying to integrate cultural aspects in interactive systems poses the challenge of modeling these aspects based on a sound theoretical framework that is able to explain or even predict behavioral heuristics of different cultures. On the other hand the developer's

own cultural background has to be kept in check in the process as it provides implicit design heuristics for the system that might go easily unnoticed. Consequently, there are different strategies for avoiding these pitfalls. Hofstede's dimensional theory of culture is very popular but the difficulties in applying this theory have been noted. Thus, it remains to be shown if there are more suitable models for this endeavor. For keeping the developer's own cultural background in check, a solution was presented that involves establishing best practice guidelines for the integration of cultural aspects in interactive systems.

Acknowledgements

The work described in this paper was partially supported by the German Research Foundation (DFG) with research grant RE2619/2-1, the Japan Society for the Promotion of Science (JSPS) with a grant-in-aid for scientific research (C) (19500104), and by the European Community (EC) in the eCIRCUS project IST-4-027656-STP.

References

1. Abrilian, S., Martin, J.-C., Buisine, S., Devillers, L.: Perception of movement expressivity in emotional TV interviews. In: HUMAINE Summerschool (2006)
2. Argyle, M.: Bodily Communication. Methuen & Co. Ltd., London (1975)
3. Aylett, R., Paiva, A., Vannini, N., Enz, S., André, E., Hall, L.: But that was in another country: agents and intercultural empathy. In: Proceedings of AAMAS (2009)
4. Caridakis, G., Raouzaiou, A., Bevacqua, E., Mancini, M., Karpouzis, K., Malatesta, L., Pelachaud, C.: Virtual agent multimodal mimicry of humans. Language Resources and Evaluation 41, 367–388 (2007)
5. Cassell, J., Sullivan, J., Prevost, S., Churchill, E.: Embodied conversational agents. MIT Press, Cambridge (2000)
6. Cassell, J., Nakano, Y., Bickmore, T.W., Sidner, C.L., Rich, C.: Non-Verbal Cues for Discourse Structure. Meeting of the Association for Computational Linguistics, 106–115 (2001)
7. Chafai, N.E., Pelachaud, C., Pelè, D.: Analysis of gesture expressivity modulations from cartoon animations. In: Proceedings of the LREC Workshop on Multimodal Corpora (2006)
8. Ekman, P.: Telling Lies — Clues to Deceit in the Marketplace, Politics, and Marriage, 3rd edn. Norton and Co. Ltd., New York (1992)
9. Hall, E.T.: The Hidden Dimension. Doubleday (1966)
10. Hall, L., Woods, S., Aylett, R., Newall, L., Paiva, A.: Achieving Empathic Engagement Through Affective Interaction with Synthetic Characters. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784, pp. 731–738. Springer, Heidelberg (2005)
11. Hofstede, G.: Cultures Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Sage Publications, Thousand Oaks (2001)
12. Iacobelli, F., Cassell, J.: Ethnic identity and engagement in embodied conversational agents. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS, vol. 4722, pp. 57–63. Springer, Heidelberg (2007)
13. Jan, D., Herrera, D., Martinovski, B., Novick, D., Traum, D.: A Computational Model of Culture-Specific Conversational Behavior. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS, vol. 4722, pp. 45–56. Springer, Heidelberg (2007)

14. Lewis Johnson, W.: Serious use of a serious game for language training. In: Proceedings of the International Conference on Artificial Intelligence in Education, pp. 67–74 (2007)
15. Khaled, R., Biddle, R., Noble, J., Barr, P., Fischer, R.: Persuasive interaction for collectivist cultures. In: Piekarski, W. (ed.) The Seventh Australasian User Interface Conference (AUIC 2006), pp. 73–80 (2006)
16. Kipp, M., Neff, M., Kipp, K.H., Albrecht, I.: Towards Natural Gesture Synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS, vol. 4722, pp. 15–28. Springer, Heidelberg (2007)
17. Kluckhohn, F., Strodtbeck, F.: Variations in value orientations. Row, Peterson (1961)
18. Chad Lane, H., Hays, M.J.: Getting down to business: Teaching cross-cultural social interaction skills in a serious game. In: Workshop on Culturally Aware Tutoring Systems (CATS), pp. 35–46 (2008)
19. Lee, J., Marsella, S.: Nonverbal Behavior Generator for Embodied Conversational Agents. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS, vol. 4133, pp. 243–255. Springer, Heidelberg (2006)
20. McNeill, D.: Hand and Mind — What Gestures Reveal about Thought. The University of Chicago Press, Chicago (1992)
21. Nakano, Y., Nishida, T.: Awareness of Perceived World and Conversational Engagement by Conversational Agents. In: Proceedings of the AISB 2005 Symposium on Conversational Informatics for Supporting Social Intelligence & Interaction (2005)
22. Nakano, Y., Rehm, M.: Multimodal Corpus Analysis as a Method to Ensure Cultural Usability of Embodied Conversational Agents. In: Proceedings of HCI International (2009)
23. Nakano, Y., Reinstein, G., Stocky, T., Cassell, J.: Towards a Model of Face-to-face Grounding. In: Proceedings of the Association for Computational Linguistics (2003)
24. Reeves, B., Nass, C.: The Media Equation – How People Treat Computers, Television, and New Media Like Real People and Place. Cambridge University Press, Cambridge (1996)
25. Rehm, M.: She is just stupid – Analyzing user-agent interactions in emotional game situations. Interacting with Computers 20(3), 311–325 (2008)
26. Rehm, M., André, E.: More Than Just a Friendly Phrase: Multimodal Aspects of Polite Behavior in Agents. In: Nishida, T. (ed.) Conversational Informatics, pp. 69–84. Wiley, Chichester (2007)
27. Rehm, M., Rosina, P.: Second Life as an Evaluation Platform for Multiagent Systems Featuring Social Interactions. In: Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS) (2008)
28. Rehm, M., André, E., Conradi, B., Hammer, S., Iversen, M., Lösch, E., Pajonk, T., Stamm, K.: Location-based interaction with children for edutainment. In: André, E., Dybkjær, L., Minker, W., Neumann, H., Weber, M. (eds.) PIT 2006. LNCS, vol. 4021, pp. 197–200. Springer, Heidelberg (2006)
29. Rehm, M., Bee, N., André, E.: Wave like an Egyptian – Accelerometer based gesture recognition for culture specific interactions. In: Proceedings of Britisch HCI (2008)
30. Rehm, M., Vogt, T., Bee, N., Wissner, M.: Dancing the Night Away – Controlling a Virtual Karaoke Dancer by Multimodal Expressive Cues. In: Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS), pp. 1249–1252 (2008)
31. Schwartz, S.H., Sagiv, L.: Identifying culture-specifics in the content and structure of values. Journal of Cross-Cultural Psychology 26(1), 92–116 (1995)
32. Ting-Toomey, S.: Communicating Across Cultures. The Guilford Press, NewYork (1999)