A Parameter-based Model for Generating Culturally Adaptive Nonverbal Behaviors in Embodied Conversational Agents

Lipi, Afia Akhter; Nakano, Yukiko; Rehm, Matthias

Published in:
Universal Access in Human-Computer Interaction

DOI (link to publication from Publisher):
10.1007/978-3-642-02710-9_70

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

A Parameter-Based Model for Generating Culturally Adaptive Nonverbal Behaviors in Embodied Conversational Agents

Afia Akhter Lipi[1], Yukiko Nakano[2], and Matthias Rehm[3]

[1] Dept. of Computer and Information Sciences,
Tokyo University of Agriculture and Technology, Japan
50007646211@st.tuat.ac.jp
[2] Dept. of Computer and Information Science, Seikei University, Japan
y.nakano@st.seikei.ac.jp
[3] Institute of Computer Science, Augsburg University, Germany
rehm@informatik.uni-augsburg.de

Abstract. The goal of this paper is to integrate culture as a computational term in embodied conversational agents by employing an empirical data-driven approach as well as a theoretical model-driven approach. We propose a parameter-based model that predicts nonverbal expressions appropriate for specific cultures. First, we introduce the Hofstede theory to describe socio-cultural characteristics of each country. Then, based on the previous studies in cultural differences of nonverbal behaviors, we propose expressive parameters to characterize nonverbal behaviors. Finally, by integrating socio-cultural characteristics and nonverbal expressive characteristics, we establish a Bayesian network model that predicts posture expressiveness from a country name, and vice versa.

Keywords: conversational agents, enculturate, nonverbal behaviors, Bayesian network.

1 Introduction

When we meet someone, one of the first things we do is to classify the person as "in-group" or "out-group". This social categorization is often based on the ethnicity [4]. When someone is identified as part of in-groups as opposed to out-group, she or he is perceived as more trustworthy. As the same way, does the ethnicity of Embodied Conversational Agents (ECAs) also matter? Findings in previous studies support a claim that ethnicity of embodied conversational agents effects users' attitudes and behaviors. Nass et al [8] found that users showed more trust and were more willing to take the agent's suggestion if the agent was of the same ethnic group or from the same cultural background.

Aiming at generating culture-specific behaviors, specifically postures, in ECAs, this study focuses on modeling cultural differences. Our method enables the user to experience exchanges of cultural specific posture expressions in human-agent interaction. However, defining culture is not an easy task and there are various definitions of this

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2009, LNCS 5615, pp. 631–640, 2009.

notion around, and descriptive and explanatory theories are not very useful for computational purposes. Thus, to generate culturally appropriate nonverbal behaviors in ECAs, we will propose a parameterized socio-cultural model which characterizes the group or the society using a set of numerical values, and selects agents' nonverbal expressions according to the parameter set using probabilistic reasoning facilitated by a Bayesian network.

As a data-driven approach, we have already collected comparative multimodal corpus for two countries; an Asian country Japan and a European country Germany, and extracted culture-specific posture shapes from the corpus [1]. In this paper, based on the results of our empirical study, we extend our research by employing a model-driven approach by introducing Hofstede model [7] as a theoretical basis of describing socio-cultural characteristics. Hofstede's theory is more appealing for establishing a computational model because Hofstede defines each culture using five dimensions each of which has quantitative nature. Integrating the Hofstede theory of culture [7] and the empirical data from our corpus [1], in this paper, we will implement a parameterized model which generates culture specific non-verbal expressions. Our final goal is not restricted to build a model for embodied conversational agents, but is to propose a general model which estimates nonverbal parameters for various cultures.

In the following sections, first we will discuss related work in section 2, and in section 3, explain the approach of this study in addition to a brief description of Hofstede model. Section 4 reports the empirical data in our corpus, and Section 5 proposes a Bayesian network which combines Hofstede theory and the empirical data. Section 6 describes a nonverbal decision module, and Section 7 gives conclusions and future work.

2 Related Work

As research in ethnicity of ECAs, Nass et al [15] examined the question: "Does the ethnicity of a computer agent affect users' attitudes and behaviors?" So they did a study of Korean subjects interacting with an American agent or with a Korean agent. They found out that ethnic similarity had significant effect on users' attitudes and behaviors. When the ethnicity of the subject was the same as the agent, the subject took the agent to be more trustworthy and convincing. Nass et al [15] claimed that user showed more trust and were more willing to take the agent's suggestion or more willing to give credit card number. These results suggests that culture adapted agents are more positively accepted, and will provide more successful outcome in E-commerce. Francisco et al. [11] assessed index of ethnicity on the basis of language and non-verbal features, not by physical appearance such as skin color, hair, style or clothing. They found that children had longer interaction with a virtual peer whose verbal and nonverbal behaviors matched with their own ones than with a ethnically mismatched virtual peer.

Isbister [5] pointed out the importance of non-verbal communicative behaviors which is largely culture-specific. She reviewed a number of features of nonverbal communication such as eye gaze and gestures. Arabs treat sustained eye contact as a sign of engagement and sincerity where as Japanese interpret sparse use of direct eye contact as a sign of politeness. Another example is a simple head nod which is interpreted as a sign of agreement in Germany but indicates only attention in Japan. The

frequency, the manner, and the number of gestures are also culturally dependent. Mediterranean people have far more gestures than North American do. Italians tend to use big gestures and do gestures more frequently than English or Japanese. Southern Europeans, Arabs, and Latin Americans use animated hand gestures where as Asians and Northern Europeans use quieter gestures [5].

As studies in learning systems, Johnson et al. [10] described a language tutoring system that also takes cultural differences in gesture usage into account. Maniar and Bennett [12] proposed a mobile learning game to overcome cultural shock by making the user aware of the cultural differences. The eCIRCUS project (Education through characters with emotional intelligence and role playing capabilities that understand social interaction) [13] is aiming at developing models and innovative technologies that support social and emotional learning through role-plays. For example, children become aware of social sensitive issues such as bullying through virtual role-plays with synthetic characters.

3 Describing Socio-cultural Characteristics

As a theoretical approach, we employ Hofstede theory to describe socio-cultural characteristics. Then, from an empirical approach, we propose several nonverbal expressive parameters to characterize the posture expressiveness. These two layers will be integrated into a Bayesian network model to predict either behavioral characteristics or a culture.

We start with introducing Hofstede theory [7]. Hofstede theory defines culture as a dimensional concept, and consists of the following five dimensions which are based on a broad empirical survey.

1. Hierarchy/Power Distance Index: This dimension describes the extent to which different distribution of power is accepted by the less powerful members. More coercive and referent power is used in high power distance societies and more reward, legitimate, and expert power in low power distance societies.

2. Identity: This is 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 herself/himself. On the collectivist side, people are integrated into strong and cohesive groups.

3. Gender: The gender dimension describes the distribution of roles between genders. Feminine cultures place more value on relationships and quality of life whereas in masculine cultures competition is rather accepted and status symbols are of importance.

4. Truth/Uncertainty: The tolerance for uncertainty and ambiguity is defined in this dimension. It indicates to what extent the members of a culture feel either uncomfortable or comfortable in unstructured situations which are novel, unknown, surprising, or different from usual.

5. Virtue/Orientation: This dimension distinguishes long and short term orientation. Values associated with long term orientation are thrift and perseverance whereas values associated with short term orientation are respect for tradition, fulfilling social obligations, and saving one's face.

Since cultural characteristics in Hofstede theory are synthetic, a set of parameter values indicates the cultural profile. Table 1 gives Hofstede's ratings for three countries [2]. For example, in Identity dimension, Germany (67) is more individual culture than Japan (46), and US (91) is the most individual among three.

Table 1. Hofstede ratings for three countries

	Hierarchy	Identity	Gender	Uncertainity	Orientation
Germany	35	67	66	65	31
Japan	54	46	95	92	80
US	40	91	62	46	29

4 Characterizing Nonverbal Behaviors

4.1 Defining Posture Expressive Parameters

To define parameters that characterize posture expressivities, we reviewed previous studies. To describe cultural differences in gestures, Efron [14] proposed parameters such as spatio-temporal aspects, interlocutional aspects, and co-verbal aspects. Using a factor analysis, Gallahar [15] revealed four dimensions; expressiveness, expansiveness, coordination, and animation. Based on these previous studies, Hartmann et al. [16] defined gestural expressivity using six parameters such as repetition, activation, spatial extent, speed, strength, and fluidity

Based on our literature study, we came up with five parameters which define the characteristics of posture. The five parameters are spatial extent, rigidness, mirroring, frequency, and duration. In the next section, the details of deriving values for each behavioral expressive parameter are explained.

4.2 Assigning Values

Since we found that the cultural difference in posture shifts is very clear in arm postures [1], we focus on predicting arm postures. Among the five expressive parameters we proposed in section 4.1, we got the value of frequency and duration from our previous empirical study [1]. To find the values for spatial extent and rigidness, we will conduct an experiment. Then to derive the numerical value for mirroring, we will analyze our video data.

Frequency and duration. Frequency and duration can be assigned by referring the results of our previous empirical study [1]. Average frequency of arm posture shifts in German data is 40.38 per conversation and 22.8 in Japanese data. On the other hand, average duration of each posture is 7.79 sec in German data and 14.8 sec in Japanese data. Thus, Japanese people like to keep one posture longer than German people.

Measuring impression for spatial extent and rigidness. By spatial extent, we mean the amount of physical space required for a certain posture. As the term rigidness seems more tricky type, we used the opposite word *relax* instead of rigidness to make the term simple to the subjects.

Study Design: We extracted 15 video clips of postures from Japanese video data and 15 posture video clips from German data, and asked 7 Japanese subjects and 10 German subjects to rate each video clip. The rating was made using a questionnaire which asked the subjects to rate impressions on the shape of arm, lower body, and whole body using 7 point scales where 1 is meant to be the least value and 7 is the top most value. For each video clip, the subjects answered their impression in two dimensions; spatial extent and relax. Before starting the experiment, each subject was handed an explanation form which explained how each subject should rate the video clips.

Result: The rating results are shown below. Table 2 shows that German do more relaxed postures than Japanese, and Japanese do smaller postures than German.

Table 2: Non-verbal expressive parameters from Experimental data

Country	Spatial Extent	Rigidness
Japan	5.25	8.58
Germany	7.33	7.62

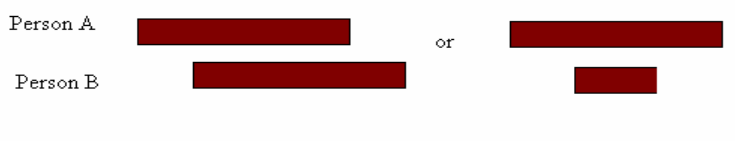
Analyzing mirroring. Mirroring refers to an interpersonal phenomenon in which people unknowingly adjust the timing and content of their behavioral movements such that they mirror the behavioral cues exhibited by their social interaction partner. Mirroring has positive effects on interaction, and enhances the relationship between the conversants.

Study Design: We analyzed videos of 10 Japanese pairs and 7 German pairs (both speaker and listener) acting the first time meeting scenario [1] where two people met first time and had a conversation to know each other. The dyadic conversation took place for 5 minutes.

After annotating posture shifts of both speaker and listener using Bull's Coding Scheme [3], we counted the number of postures common in both parties, speaker and listener, by using two *conditions* below;

(1) If person A shifts to a new posture while speaking, and within five seconds person B also changes to the same posture as person A, vice versa.

(2) If person A shifts to a new posture, and soon person B also does the same posture which is overlapped with person A's posture.



Result: The average number of mirroring for Japanese is 6.2 per conversation, and for German is 0.57. This result suggests that Japanese are more likely to synchronize with the conversation partner than German people. So they like to be in a group and more collective in nature.

5 Combining Theoretical and Empirical Approach to Develop a Parameter-Based Model

Based on Hofstede theory of culture, we proposed a model where culture is connected to Hofstede dimensions which are also connected with nonverbal expressive parameters for postures.

5.1 Reasoning Using a Bayesian Network

To build this parameter based model, we employ Bayesian network technique. Figure 1 shows our Bayesian network which models relationship between socio-cultural aspects and behavioral expressiveness. Bayesian networks are acyclic directed graphs in which nodes represent random variables and arcs represent direct probabilistic dependences among them. Bayesian networks [2] handle uncertainty at every state. This is very important for our purpose as the linkage between culture and nonverbal behavior is a many to many mapping. In addition, since the network can be used in both directions, it can infer the user's cultural background as well as simulate the system's (agent's) culture specific behaviors.

5.2 Parameter Based Socio-cultural Model

In order to build a Bayesian network for predicting socio-cultural aspects in posture expressiveness, the GeNie[6] modelling environment was used.

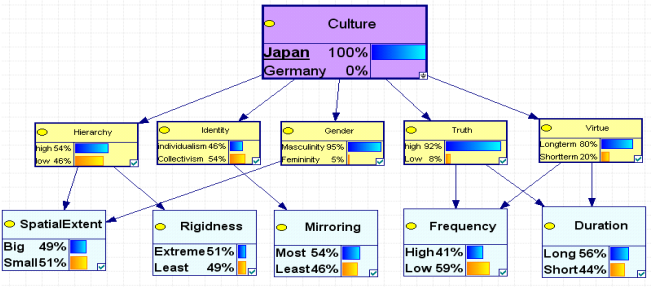


Fig. 1. Bayesian network model predicting Japanese posture expressiveness parameters

First Layer: The first part of the network is quite simple. The entry node of the Bayesian network is a culture node which is connected to Hofstede's dimensions. Currently we have inserted two countries, Germany and Japan.

Middle Layer: The middle layer defines Hofstede's five dimensions. We already integrated all five dimensions: hierarchy, identity, gender, uncertainty, and orientation. Hofstede ratings for each country shown in Table1 are used as the probabilities in each node.

The Lowest layer: The lowest layer consists of a number of different behavioral parameters that depend on a culture's position on Hofstede's dimensions. We draw a connection between the cultural dimensions and the nonverbal behaviors. Lowest level consists of five nodes whose values were specified in section 4.2.

a) Spatial Extent: Spatial extent describes the amount of physical space required for a certain posture. From our experimental data, we found that German do bigger posture than Japanese. When we compare the postures between male and female subjects, we found that, Japanese female do smaller posture than male, and the difference is bigger than in German data. So, we can say Japanese society is more masculine than German. Moreover, hierarchy affects the spatial extent. In high hierarchical societies, people seem to make small postures [5].

b) Rigidness: How stiff the posture is. Our experimental data revealed that Japanese people seem to do more rigid postures than German, and German seem to be more relaxed than Japanese. In high hierarchical society, people are stiffer than low hierarchical society. Thus, we assume a linkage between hierarchy and rigidness.

c) Mirroring: Since mirroring is to copy the conversation partner's postures during a conversation, we assume that frequency of mirroring correlates with collective nature. In our corpus study in section 4.2, Japanese people actually more frequently did mirroring than German people.

d) Frequency: German people change their posture more frequently than Japanese. According to Hofstede theory, Japanese culture is of long term orientation therefore we set links from truth and virtue to frequency.

e) Duration: Japanese people stay at a single posture for a long period of time than German people. Thus, we assume that both truth and virtue affect duration.

For each node in the Bayesian network, probability is assigned based on the data that we reported in section 4. For example, since the posture shift frequency of German data (40.38) is 1.77 times of Japanese data (22.8), as probability values, we assigned 0.66 and 0.34 to each country respectively.

Output: When a country is chosen at the top level as evidence, behavior expressive parameters are estimated. For instance, as shown in Figure 2, when *Japan* is chosen as evidence, the results of estimations are; spatial extent is small (51%), rigidness is extreme (51%), mirroring is most (54%), frequency is low (59%), duration is long (56%). In the same way, when *Germany* is given as evidence, the estimation results are; spatial extent is big (51%), rigidness is least (53%), mirroring is least (66%), frequency is high (52%), and duration is short (52%).

5.3 Evaluation of the Model

As an evaluation of our model, we tested whether this model can properly predict posture expressiveness of other countries. When the Hofstede scores for US shown in Table 1 are applied, the model predicts that spatial extent for US is big (51%), rigidness is least (52%), mirroring is least (90%), frequency is high (53%), and duration is short (53%). This prediction suggests that American postures are less rigid (in other words more relaxed), and this supports what Ting Toomey has reported [5].

6 Posture Selection Mechanism

This section presents our posture selection mechanism which uses the Bayesian network model as one of its components. A simple architecture is given in Figure 2. Basically it is divided in to three main modules. The input to the mechanism is a country name and a text that the agent speaks.

6.1 Probabilistic Inference Module

The Probabilistic Inference Module takes country name as input and outputs the nonverbal parameters for that country. To generate outputs, the module refers our Bayesian network model. We used Netica API of JAVA version as an inference engine. The outputs of this module are values of nonverbal expressive parameters of each culture: spatial extent, rigidness, duration, and frequency.

6.2 Decision Module

This module is the most important module. This module has two sub-modules.

b1: Posture computing module: This module takes the estimation results from the Bayesian network as inputs, and uses them as weights for each empirical data. Then, it calculates the sum of all the weighted values. For example, the score for a posture, PHFe (Put hand to face), which is frequently observed in Japanese data, is shown below. 0.5183, 0.507, 0.58, and 0.56 are weights for spatial extent, rigidness, frequency, and duration respectively, which are given by the Bayesian network. 4.19, 4.4, 2.725, and 1.01 are values obtained in empirical studies in section 4[1].

*PHFe={(0.5183 *4.19)+(0.507 *4.4)+(0.58 *2.725)+ (0.56*1.01)} * 10 = 65.49*

b2: Posture distinguishing module: This sub-module separates typical postures of each cultures as German like postures, Japanese like postures, or common postures (used in both Germany and Japan). It is judged by checking the thresholds for each country. If the text is Japanese, and the posture value falls within the range of Japanese posture, it sends the posture to the Generation phase as a posture candidate.

[1] Since various kinds of measures were used in the empirical data, they were normalized them into 1 to 7.

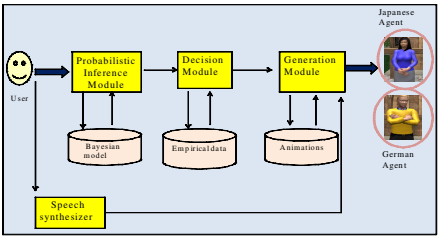


Fig. 3. A simplified architecture of the system

6.3 Generation Module

This module takes postures recommended by the decision module and looks for the animation file for that posture in animation database. Then, Horde3D animation engine generates the animation file. We use Hitachi HitVoice for TTS which converts the text into a wav file, and then the agent speaks with appropriate culture-specific postures.

7 Future Work and Conclusions

Employing Bayesian network, we combined Hofstede model of socio-cultural characteristics with posture expressive parameters that we proposed, and found that our model estimates cultural specific posture expressiveness quite well. As future work, we plan to apply our posture generation mechanism to language exchange application on the web where two users from different countries log on the service, and teach their own language to the partner, and learn a foreign language from her or his partner. In this application, the system not only helps the user teach a language, but also makes the learner familiar with the culture-specific nonverbal behaviors.

Acknowledgment. This work is funded by the German Research Foundation (DFG) under research grant RE 2619/2-1 (CUBE-G) and the Japan Society for the Promotion of Science (JSPS) under a Grant-in-Aid for Scientific Research (C) (19500104).

References

1. Rehm, M., et al.: Creating a Standardized Corpus of Multimodal Interactions for Enculturating Conversational Interfaces. In: Proceedings of Workshop on Enculturating Conversational Interfaces by Socio-cultural Aspects of Communication, 2008 International Conference on Intelligent User Interfaces (IUI 2008) (2008)
2. Rehm, M., et al.: Too close for comfort? Adapting to the user's cultural background. In: Proceedings of the 2nd International Workshop on Human-Centered Multimedia (HCM), Augsburg (2007)
3. Bull, P.E.: Posture and Gesture. Pergamon Press, Oxford (1987)
4. Nass, C., Isbister, K., Lee, E.: Truth is Beauty Researching Embodied Conversational Agents. In: Cassell, J., et al. (eds.) Embodied Conversational Agents, pp. 374–402. The MIT Press, Cambridge (2000)
5. Ting-Toomey, S.: Communication Across Culture. The Guildford Press, New York (1999)

6. GeeNIe and SMILE, http://genie.sis.pitt.edu/
7. Hofstede, http://www.geert-hofstede.com/hofstededimensions.php
8. Lee, E.-J., Nass, C.: Does the ethnicity of a computer agent matter? An experimental comparison of human-computer interaction and computer-mediated communication. In: Prevost, S., Churchill, E. (eds.) Proceedings of the workshop on Embodied Conversational characters (1998)
9. Isbister, K.: Building Bridges through the Unspoken: Embodied Agents to facilitate intercultural communication. In: Payr, S., Trappl, R. (eds.) Agent Culture: Human –Agent Interaction in a Multicultural World, pp. 233–244. Lawrence Erlbaum Associates, Mahwah (2004)
10. Johnson, W., et al.: Tactical Language Training System: Supporting the Rapid Acquisition of Foreign Language and Cultural Skills. In: Proc. of InSTIL/ICALL - NLP and Speech Technologies in Advanced Language Learning Systems (2004)
11. 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 (2008)
12. Maniar, N., Bennett, E.: Designing a mobile game to reduce cultural shock. In: Proceedings of ACE 2007, pp. 252–253 (2007)
13. http://www.e-circus.org/
14. Efron, D.: Gesture, Race and Culture. Mouton and Co. (1972)
15. Gallaher, P.E.: Individual Differences in Nonverbal Behavior; Dimension of style. Journal of Personality and Social Psychology 63819, 133–145 (1992)
16. Hartmann, B., Mancini, M., Buisine, S., Pelachaud, C.: Design and evaluation of expressive gesture synthesis for embodied conversational agents. In: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems, pp. 1095–1096 (2005)