Too Close for Comfort?

Adapting to the user’s cultural background

Matthias Rehm, Nikolaus Bee, Birgit Endrass, Michael Wissner, and Elisabeth André
Multimedia Concepts and Applications, University of Augsburg
Eichleitnerstr 30, 86159 Augsburg, Germany
{rehm|bee|endrass|wissner|andre}@informatik.uni-augsburg.de

ABSTRACT

The cultural context of the user is a largely neglected aspect of human centered computing. This is because culture is a very fuzzy concept and even with a computational model of culture it remains difficult to derive the necessary information to recognize the user’s cultural background. Such information is only indirectly available and has to be derived from the observable multimodal behavior of the user. We propose the usage of a dimensional model of culture that allows applying computational methods to derive a user’s cultural background and to adjust the system’s behavior accordingly. To this end, a Bayesian network is applied to allow for the necessary inferences despite the fact that the given knowledge about the user’s behavior is incomplete and unreliable.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors, human information processing*; I.5.5 [**Pattern Recognition**]: Implementation—*interactive systems*

General Terms

Human factors

Keywords

cultural computing, embodied conversational agents, bayesian network modeling

1. INTRODUCTION

Our cultural backgrounds largely depend how we interpret interactions with others, which aspects we find relevant, and what kind of behavior is deemed annoying or insulting. But is it really necessary to take the user’s cultural background into account for interactions with a computer system? Marcus and Gould [16] analyze websites from all over the world and show that they are tailored to cultural preferences and differ largely in the features that are deemed necessary for the entry point of a web presence.

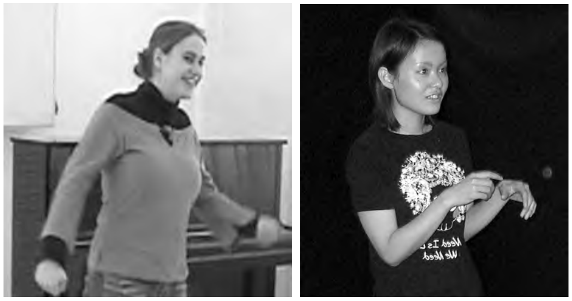

Figure 1: Comparing gestural activity of a German and a Japanese actress.

Whereas static presentations like e.g., web sites can be easily tailored to culture-specific demands during the design process (given that the designer recognizes the challenge, see e.g. [16]), dynamic generations of multimodal presentations of information cannot so easily be dealt with, because they are tailored on the fly depending on situational and contextual factors. To make these dynamic presentations sensible to cultural differences, one needs a set of parameters or rules that allow for influencing the generation process in the same way as the situational and contextual factors. Embodied conversational agents (ECAs) can be regarded as a special case of multimodal dynamic interaction systems. They promote the idea that humans, rather than interacting with tools prefer to interact with an artefact that possesses some human-like qualities at least in a large number of application domains. If it is true as Reeves and Nass’ Media Equation suggests that people respond to computers as if they were humans [19], 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.

In this article we present a model allowing to infer the user’s cultural background from his interactions with embodied conversational agents and to set the system’s cultural background resulting in culture-specific system behavior, i.e. in generating appropriate expressive behavior for the agents.

Cultural influences manifest themselves on a number of different channels like eye gaze or gestural expressivity and are thus inherently multimodal in nature. Figure 1 exemplifies different gestural activity of a German and a Japanese

HCM’07, September 28, 2007, Augsburg, Bavaria, Germany.
Copyright 2007 ACM 978-1-59593-781-0/07/0009

actress. But apart from the multimodal nature of cultural influences an additional challenge has to be taken into account. There is no one-to-one mapping between culture and expressive behavior. Gestural expressivity for instance is also influenced by a person's personality e.g. extroverts use more space and do more gestures. On the other hand, culture manifests itself not only in gestural expressivity but in many different behavior routines. Thus, instead of a one-to-one we have a many-to-many mapping resulting in the need of taking different channels into account to realize culture-specific behavior on the one hand and to infer the cultural background of the user on the other hand.

Before we consider the definition of culture and how this notion can be used in a computational way (Sec. 3, we present work that takes culture as an important contextual information into account (Sec. 2). Our approach to modeling cultural influences is presented in Section 4. In Section 5 we shortly present an observational study we conducted to ground the model in empirical data before we finish the article with conclusive comments (Sec. 6).

2. RELATED WORK

Although embodied conversational agents are ideal candidates for integrating the cultural context of the user in the interaction as we have argued above, there are few approaches that actually consider this challenge. Rosis, Pelachaud, and Poggi [7] illustrate this problem by their survey of the Microsoft Agents web site which shows, that the appearance as well as the animations of the characters are all based on western cultural norms. They only found four non-western style agents, which moreover exhibited only a reduced set of animations. Sengers [22] emphasizes this problem as a "McDonaldization" of agents, if culture-specific aspects are disregarded in the design and behavioral modeling of agents. Apart from imposing western cultural standards on all users, the danger lies in a very low acceptance of such agents by users with different cultural backgrounds. This fact can be attributed to such fuzzy aspects like globalisation but as Nass and colleagues have shown, the cultural background and behavioral consistency of an agent matter. In one of their studies, Korean subjects were confronted with either an American or a Korean agent. The subjects trusted the agent which corresponded to their own cultural identity more. Thus, in an e-commerce scenario, e.g., the appropriate agent should lead to more successful transactions. In another study, Takeuchi, Katagiri, Nass, and Fogg [23] examine cultural effects for reciprocal behavior of American and Japanese subjects. In a variation of the desert survival task, they showed how subjects reacted towards a computer that helped them on the task. American subjects showed reciprocal behavior, if the same computer needed their help on a later task, the Japanese subjects on the other hand showed reciprocal behavior if the computer was a "member" of the same group as the previous computer. This group relation had to be established explicitly to invoke reciprocal behavior at all in the Japanese subjects. Allbeck and Badler [1] review culture-specific nonverbal communicative behavior taking into account facial expressions, gestures, body movements, posture, visual orientation (eye contact), physical contacts (handshakes, patting), spatial behavior (proximity, distance, positions), appearance, and nonverbal vocalizations. From a technical point of view, the problem arises of how to ensure consistency between verbal and nonverbal communicative behaviors. An agent that just stares at the interaction partner and does not show any appropriate eye movements or gaze behavior will create an awkward atmosphere which may well lead to a failure of the interaction. To prevent such failures of communication and make agents believable and consistent in their behavior, the EMOTE model by Allbeck and Badler seems to provide a promising starting point since it enables the generation of several variants for the same basic animation data depending on the settings of parameters, such as effort and shape. Even though they present a couple of interesting ideas regarding the use of the EMOTE model to capture cultural aspects, there is not yet any implementation of the approach so far. Noot and Ruttkay [18] define a specific markup language called GESTYLE which allows the user to vary an ECA's style both for verbal and nonverbal modalities. The style specifies when and how an agent uses certain gestures and how the speech is modulated. An important component of their approach is a style dictionary which guides the choice of appropriate behaviors. This dictionary could be the place where culture specific styles might be integrated. In contrast to them, Martin et al. [17] employ a copy-synthesis approach to specify expressivity dimensions for an embodied conversational agent. Based on annotated video recordings of human speakers, they manually define markups augmented by expressivity parameters which are then forwarded to an animation engine to generate individual behaviors. Johnson et al. [14] describe a language tutoring system that also takes cultural differences in gesture usage into account. The users are confronted with some prototypical settings and apart from speech input, have to select gestures for their avatars. Moreover they have to interpret the gestures by the tutor agents to solve their tasks. Core and colleagues [6] describe a training scenario for different negotiation styles which is set in a different culture than the trainees'. Although this setting might be regarded as a prototypical case for rendering the system's behavior culture specific, especially regarding different types of negotiation, this aspect is not integrated in the system. The authors acknowledge that it might be an appropriate move but do not make suggestions on how to integrate this aspect.

While the generation of individual behaviors is considered as an important prerequisite to realize culture-specific behaviors, none of the approaches so far is able to automatically extract the relevant parameters from the user's input. And even though there are a number of approaches to simulate culture-specific agents, a principled approach to the generation of cross-cultural behaviors is still missing.

3. THE FUZZY NOTION OF CULTURE

To allow culture to be used in a computational way, it is necessary to build on a concept of culture that features a way to measure the impact of different cultures on behavior or expressivity. The definition of culture is not an easy task and there are many fuzzy definitions of this notion around. Nevertheless there is one theoretical school which claims that culture can be defined as values and norms that members of a certain group adhere to. Kluckhorn and Strodtbeck [15] e.g. distinguish between five different value orientations ranging from people and nature over time sense to social relations. Although this is a first classification of possible values, the impact on behavior is more of an anecdotal character not allowing for an operationalizable model.

Dimension	Extreme	Sound	Space
Hierarchy	High power	soft	far
	Low power	loud	close
Identity	Individualistic	loud	far
	Collectivistic	soft	close
Gender	Masculinity	loud	close
	Femininity	soft	close
Uncertainty	Uncertainty avoidance	loud	far
	Uncertainty tolerance	soft	close
Orientation	Long-term orientation	soft	far
	Short-term orientation	soft	close

Table 1: Correlation between positions on the Hofstede dimensions and nonverbal behavior.

A more recent representative of this line of thinking is Hofstede [10] who defines culture as a dimensional concept. His theory is based on a broad empirical survey that gives the most detailed insight in differences of value orientations and norms so far. Hofstede defines five dimensions on which cultures vary. Thus, a given culture is defined as a point in a five-dimensional space.

1. *Hierarchy:* This dimension describes the extent to which different distribution of power is accepted by the less powerful members. According to Hofstede more coercive and referent power (based on personal charisma and identification with the powerful) is used in high-H societies and more reward, legitimate, and expert power in low-H societies.

2. *Identity:* Here, the degree to which individuals are integrated into a group is defined. On the individualist side ties between individuals are loose, and everybody is expected to take care for himself. On the collectivist side, people are integrated into strong, cohesive in-groups.

3. *Gender:* The gender dimension describes the distribution of roles between the genders. In feminine cultures the roles differ less than in masculine cultures, where competition is rather accepted and status symbols are of importance.

4. *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. Whereas uncertainty avoiding cultures have rules to avoid unknown situations, uncertainty accepting cultures are more tolerant of opinions different from what they are used to and they try to have as few rules as possible.

5. *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.

According to Hofstede [10], nonverbal behavior is strongly affected by cultural affordances. The identity dimension e.g. is tightly related to the expression of emotions and the acceptable emotional displays in a culture: "(...) individualist cultures tolerate the expression of individual anger more easily than do collectivist cultures. The same holds for the expression of fear, which is easily recognized in individualist cultures but which only some observers in collectivist cultures are able to identify" (Hofstede, 2001:232). 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. Argyle [2] reports a cross-cultural study about the recognition of emotional expressions from English, Italian, and Japanese. Subjects from each culture had to identify the emotional expressions from people of each of the three cultures. English and Italian subjects were able to recognize the emotional expressions from their own and each others culture. But both failed to recognize the expressions of the Japanese (E: 38%, I: 28%). The Japanese subjects on the other hand scored fairly well in recognizing the expressions from English and Italian people but also encountered difficulties with the expressions of people from their own culture (45%).

	Hierarchy	Identity	Gender	Uncert.	Orient.
Arabia	80	38	52	68	*
China	80	20	66	30	118
Germany	35	67	66	65	31
Israel	13	54	47	81	*
Japan	54	46	95	92	80
Sweden	31	71	5	29	33
Thailand	64	20	34	64	56
US	40	91	62	46	29

* denotes missing values for the given culture

Table 2: Hofstede's ratings for eight selected countries.

Hofstede, Pedersen, and Hofstede (2002) expliciteley examine the differences that arise in the use of sound and space for the five dimensions. They give a summary of these features for the two extremes of each cultural dimension (see Table 1) which they call synthetic cultures. Hofstede collected and analyzed a large data base among more than 70 countries to define these five dimensions and rated 56 countries and regions [9]. For the work presented here, we choose 8 of them as exemplary cases, which differ extremely in all five dimensions. Table 2 gives Hofstede's ratings for these countries.

3.1 Towards culture as a computational term

Cultural influences manifest themselves on different levels of behavior as we have seen above. Thus, the information about the cultural background of an interlocutor is only indirectly available and has to be derived from observations of other variables. To this end, the user's multimodal communicative behavior like eye gaze, spatial behavior, or gestural expressivity has to be analyzed.

Fortunately, there are already quite sophisticated recognition methods available for different modalities on which the inference of the user's cultural background can rely. Nevertheless, the necessary knowledge for this inference is unsure and unreliable because on the one hand recognition engines are far from perfect, on the other hand there might be a prototypical behavior for a given culture but still a specific user might deviate from this behavior. Thus, the model has to

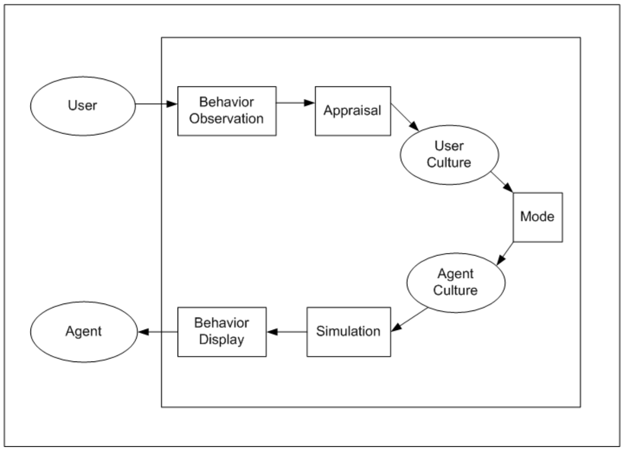

Figure 2: Overview of processing steps

cope with this unreliable information which makes Bayesian networks well suited for the task.

Bayesian networks as described in [13] are a formalism to represent probabilistic causal interactions. They have already been successfully applied to model emotions for virtual agents ([3], [4]). In the domain of culture they are also very suitable, for the following reasons:

1. Bayesian networks 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. By using a rule based system, we would get in trouble if one individual is not acting exactly in a way coherent to his cultural background. A Bayesian net, however, is making assertions about the probability of different performances, which is very adequate to model cultural behavior.

2. As the links in a Bayesian network represent the coherences between causes and effects, they are intuitively meaningful. The theoretical effect of the gender dimension of culture on the loudness of the voice, for example, is represented by a link between these two nodes. The impact, that with increasing masculinity the loudness of the voice is also rising, is easy to realize. The exact probabilities may still be difficult to define, but as we use relatively isolated effects and their relations with the cultural dimensions, we can use tendencies of behavior described in the cultural science, especially in Hofstede's synthetic cultures.

3. Bayesian networks on the one hand can be used to calculate the most probable outcome due to changes in the causal nodes and, on the other hand, to determine the most likely causes for observed effects. This is especially important for our purpose, as we can use our network in both directions, to infer the user's cultural background and to simulate the system's culture specific behavior.

4. MODELING CULTURE FOR HUMAN CENTERED COMPUTING

To integrate culture as a contextual factor into the human computer interaction, two tasks have to be solved. On the one hand, the system's behavior has to be adapted to the user's cultural background. Therefore, culture specific system behavior has to be defined. On the other hand, the user's cultural background must be known to the system either by telling it directly or by inferring this background from the interaction. As we have argued above, modeling the contextual influence of culture by the use of a Bayesian network allows us on the one hand to estimate the user's culture, on the other hand to simulate the agents' culture specific nonverbal behavior. For a general overview, the necessary processing steps are shown in Fig. 2.

- *Behavior observation:* First, the user's nonverbal behavior is detected by several sensors. If, for example the user steps further away from the screen, this could be observed by a camera.

- *Appraisal:* The collected data from the first step is used to estimate the user's culture in a diagnostic way. Therefore the context of culture is prepared and coherence between culture and nonverbal behavior is modeled.

- *Mode:* Depending on the intended application, the user's cultural background and the agents' culture are linked. In the work presented here, we simulate a cultural mirror, which means the agents adapt their behavior to the inferred cultural background of the user.

- *Simulation:* According to the constituted agent culture the nonverbal behavior is calculated. In this step we reuse the model built for the appraisal in a probabilistic way.

- *Behavior display:* Some features of nonverbal behavior are shown by our virtual agents. For example the quantity of gestures or the physical distance between the agents are adapted to the given culture.

In the following the two essential steps of setting the system's cultural background and of analyzing the user's cultural background are discussed in detail and a comprehensive example is given to illustrate the overall interaction flow.

4.1 Setting the system's cultural background

To modify the agents' behavior, it is necessary to derive appropriate behavioral parameters for the target culture. The entry node of the Bayesian network thus is a culture node which is connected to Hofstede's dimensions. The lowest layer consists of a number of different behavioral parameters that depend on a culture's position on Hofstede's dimensions. Because a given culture is represented in Hofstede's model as a point in the five-dimensional space, it would have sufficed to model only two layers, dimensions and behavioral parameters. To make the information about the target culture more accessible to the human user, the additional top-level node of culture was added.

Fig. 3 shows our Bayesian Network, the width of the connections indicates the strength of the influence between the corresponding nodes.

As we mentioned above, the concept of culture is hard to formalize. To model the connection from culture to nonverbal behavior, we rely on the ideas of cultural dimensions and synthetic cultures. The first part of the network is quite simple, as we could use Hofstede's estimations to classify our

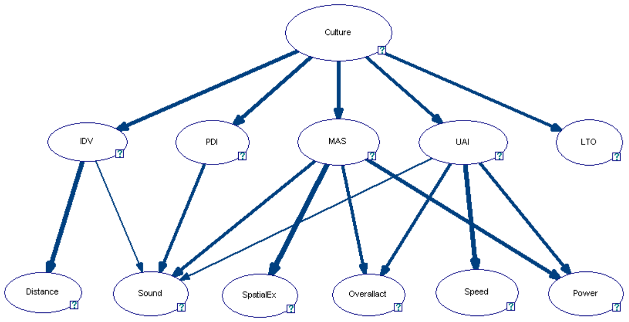

Figure 3: Bayesian network to model culture specific nonverbal behavior

eight cultures (see Table 2) on the five dimensions by using three discrete values (low, medium, high). For the second part, the linkage between the cultural dimensions and the resulting nonverbal behavior, we used the so called synthetic cultures [11], which are defined as an extreme on one of the cultural dimensions. As there are five dimensions (Hierarchy, Identity, Gender, Uncertainty, and Orientation), there are 10 synthetic cultures (one on both ends for each dimension). Hofstede et al. explicitly name tendencies of nonverbal behavior for these synthetic cultures. For the Identity dimension, for example, they state, that the so called *Indivs* (extremely high on the Individualism dimension) are verbal and likely to stand out visually, when in groups. *Collecs* (the other extreme on this dimension), in contrast, can be very silent and are physically very close within in-groups. Taking these tendencies into account, we draw a connection between the cultural dimension Identity and the nonverbal behavior clues distance and sound, with the meaning that with an increasing value on the Identity dimension the physical distance between individuals and the loudness of the voice also increase. Similar connections are established between all cultural dimensions and the corresponding behaviors.

The output level of the network, i.e. the behavioral parameters that are taken into account in the current version of the model consists of nodes for distance, sound, spatial extension, overall activation, speed, and power. All nodes are used to set the agents' culture specific behavior. Distance specifies how far apart agents stand while they interact, sound regulates how loud they speak, and the expressivity parameters influence the gestural behavior of the agents.

4.2 Analyzing the user's cultural background

As we have argued above, culture is an indirect context information and has thus to be derived from other sources of observable behavior like gaze, speech, or gestural expressivity. Here we restrict ourselves to the analysis of the user's gestural expressivity.

Bevacqua et al. [5] used the following six features to define the expressivity of gestures: *overall activation* is the number of gestures in a specific time, *spatial extent* describes how much space a gesture needs, temporal extent is the *speed* of movements, *fluidity* is the smoothness of movements, *power* is the strength of a gesture and *repetition* is the amount of repeated parts of a gesture. Four of these features have been integrated in the first version of our cultural model, *overall activation*, *spatial extent*, *speed*, and *power*.

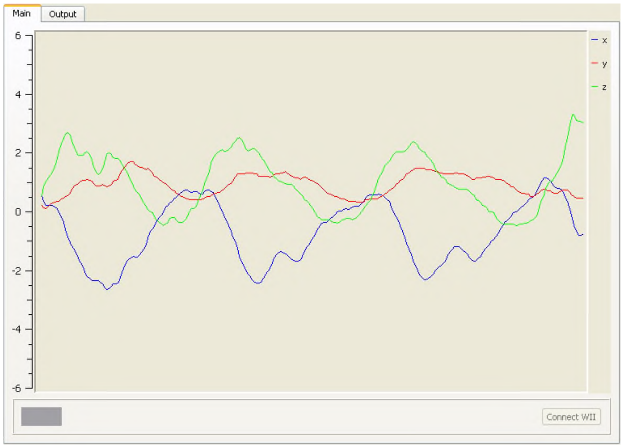

Figure 4: Typical signal of Wii remote accelerators.

Unlike Bevacqua et al. we do not analyze the user's expressivity from video streams but instead supply the user with Nintendo's Wii remote controller. The Wii Remote uses accelerometers to sense its movements in 3D space. The controller is able to connect via Bluetooth to a common PC. We gather the acceleration data from the Wii Remote for each direction (a_x: left/right, a_y: back/forth, a_z: up/down) with 100Hz and normalize it to 0. Figure 4 gives an impression of a typical signal for the three accelerometers. A gesture is recorded while pressing the front button (button B) of the Wii Remote. After releasing the recording button the expressivity features are automatically calculated and sent to our cultural model. The application is developed in C++ and uses the SMILE API from the DS Laboratory from University of Pittsburgh [8] to calculate the probabilities in our Bayesian network.

To obtain the selected expressivity features from the user's gesture in a format we can use in the Bayesian network, we first must transform it from raw data to feature values. Unfortunately we cannot calculate the expressivity features the way Bevacqua et al. [5] calculate them, as they use video processing for acquiring the hand position of gestures. And the accuracy of the acceleration sensor in Nintendo's Wii Remote controller is not accurate enough to transform the acceleration data back to its absolute position in space.

For a better readability, we define two variables S and L that are calculated on the raw data:

$$S = \sum_{i=0}^{n} a_x^2 + \sum_{i=0}^{n} a_y^2 + \sum_{i=0}^{n} a_z^2 \qquad (1)$$

$$L = \sum_{i=0}^{n} |a_x| + \sum_{i=0}^{n} |a_y| + \sum_{i=0}^{n} |a_z| \qquad (2)$$

Power (3) is derived straightforward as it is equivalent with energy and can thus be calculated in the usual way.

$$Power = \frac{1}{n} S, \text{ where n denotes the signal length} \qquad (3)$$

To find the formulas for the expressivity parameters spatial extent and speed we used an experimental approach. We recorded 20 gestures, 10 with big and 10 with low spatial extent, to find a reliable formula. We found that the signal's power (3) divided by the signal's sum of its absolute values

	Sp.Ext.		Speed		Power	
	high	low	high	low	high	low
Rec. rate	97.2%	72.2%	100%	97.2%	100%	94.4%

Table 3: Evaluation results for Wii remote expressivity recognition

is a good representation of the spatial extent derived from acceleration data (see formula 4).

$$SpExt = \frac{S}{L} \qquad (4)$$

For finding the formula for speed, we also recorded 20 gestures, 10 with fast speed and 10 with low speed. We found that a light variation of the formula for spatial extent (4) by multiplying instead of dividing the signal's power (3) by the sum of its absolute values, gives a good approximation of the gestures speed (see formula 5).

$$Speed = \frac{1}{n^2}SL, \text{ where n denotes the signal length} \qquad (5)$$

Our Bayesian network uses discrete values for representing different evidence conditions. We defined following conditions to which all input and output channels must be adapted: low, medium and high. To be able to calculate such discrete values, we must normalize the calculated features. Our previously recorded gestures give us a good representation of maximum and minimum values. We use these values to normalize our expressivity features to 0 and thresholds for low, medium and high features.

An evaluation of recognition accuracy was done with three subjects. The results are given in Table 3. Each subject had to do 12 gestures for each of the six conditions. The recognition rates are very good for speed and power, the recognition of low spatial extent turned out to be more difficult.

To infer the user's cultural background from his expressive behavior, the results from the above mentioned calculations constitute the evidence which is entered on the lowest layer of the Bayesian network. Updating the network then propagates the evidence through the network resulting in a probability distribution over the top-level culture node. This information then represents the inferred user's cultural background. The next section gives an extended example starting with the user input and resulting in the display of appropriate behavior for the target culture by the agents.

4.3 Putting it all together

To demonstrate or model, we apply it to the Virtual Beergarden which represents a meeting place, where embodied conversational agents and users can freely move around and interact. The Beergarden serves as a multiagent system to test models of social group dynamics or innovative interaction techniques for the user (e.g. [20], [21]). For the current purpose, the user is in the role of an observer. Agents are standing in groups in the Beergarden and interact with one another. By analyzing the expressive behavior of the user, the agents' behavior is adapted to display culture specific behavior.

An interaction cycle is demonstrated in this setting. Here we focus on the Wii remote as the sole input device for the user. Consider a user, who exhibits slow and modest gestural expressivity (Fig. 5 upper left). The system analyzes the user's behavior and calculates his cultural background based on the available input data.

In our example, the Wii remote's accelerometer data allows to calculate that the user's gestures are slow, not powerful and not extended in space. Thus the input nodes *SpatialExtent*, *Speed* and *Power* in our Bayesian net are set to the value low. As only one gesture was detected so far, nothing can be said about the *Overall Activation*. With the evidences for the three nodes set, the Bayesian network is updated to allow for inferring in a diagnostic way the user's cultural background. Figure 5 (upper right) gives an impression of the state of the network after the update process. The probability is propagated via the dimensional layer to the culture node. In our example, the system estimates the user to belong to the Swedish culture with a probability of 92%.

This examplifies how even with incomplete information, the network is able to estimate the user's cultural background. Let's assume we have more input data available for instance the user's distance to the agents. If the user is standing far away from the agents, this would be another clue in support of the Swedish culture and thus the probability for this inference would increase to 98%. If instead the user is standing very close to the agents, this would be information contradicting the inference of Sweden and thus the probability for this inference will drop to 56% with Thailand increasing to a probability of 34%. Nevertheless, Sweden remains the one with the highest probability because the other clues still support this inference.

After appraising the user's cultural background, the agents' culture is set accordingly. At the moment, the culture with the highest probability is chosen as the target culture of the user. More sophisticated decision mechanisms could be thought of that take the probability distribution over the different cultures into account. Whereas the user's behavior is not solely attributable to his cultural background and thus will contain also some idiosyncracies, the system's behavior is prototypical for a given culture. To this end, the Bayesian network is used for a causal inference. Evidence is brought in by setting the culture value to 100% for the inferred background of the user, in our case for Sweden. The result is given in Figure 5 (lower right) which depicts the Bayesian net after updating for a prototypic Swedish agent. During the causal inference, the evidence is propagated from the culture node through the dimensional layer to the expressivity nodes which now represent the most probable behavior for the given culture. In case of Sweden, the agents stand far away from each other and speak in a mid voice. They do not gesture much and if they gesture, they do it slowly and with little spatial extent (see Fig. 5 lower left).

Figure 6 gives another example. This time, the user's gestures are slower and wider. Thus, the infered cultural background is Chinese and results in a changed system behavior. The agents are now moving closer, and use more, wider, and more powerful gestures.

Comparing the probability distributions for the expressivity nodes in the Bayesian networks for the diagnostic and the causal inference shows that the agents are not just reflecting the user's behavior directly. Some features might differ as the agents exhibit behavior that is prototypical for the illustrated culture (according to the literature), whereas the user's behavior is also influenced by other factors like his personality or his current emotional state. Due to this many-to-many mapping between culture and behavior, Bayesian

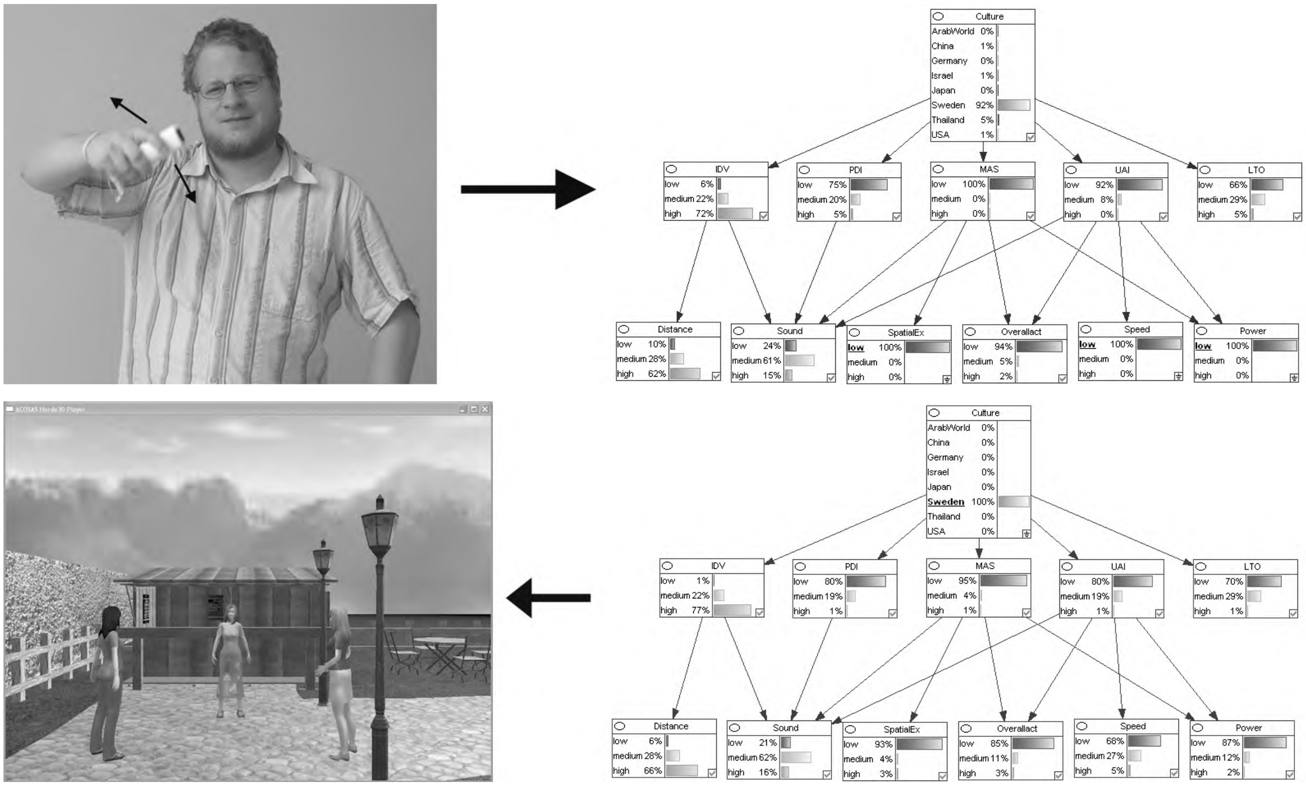

Figure 5: From user's cultural background (diagnostic inference) to culture specific agent behavior (causal inference). Example Swedish culture.

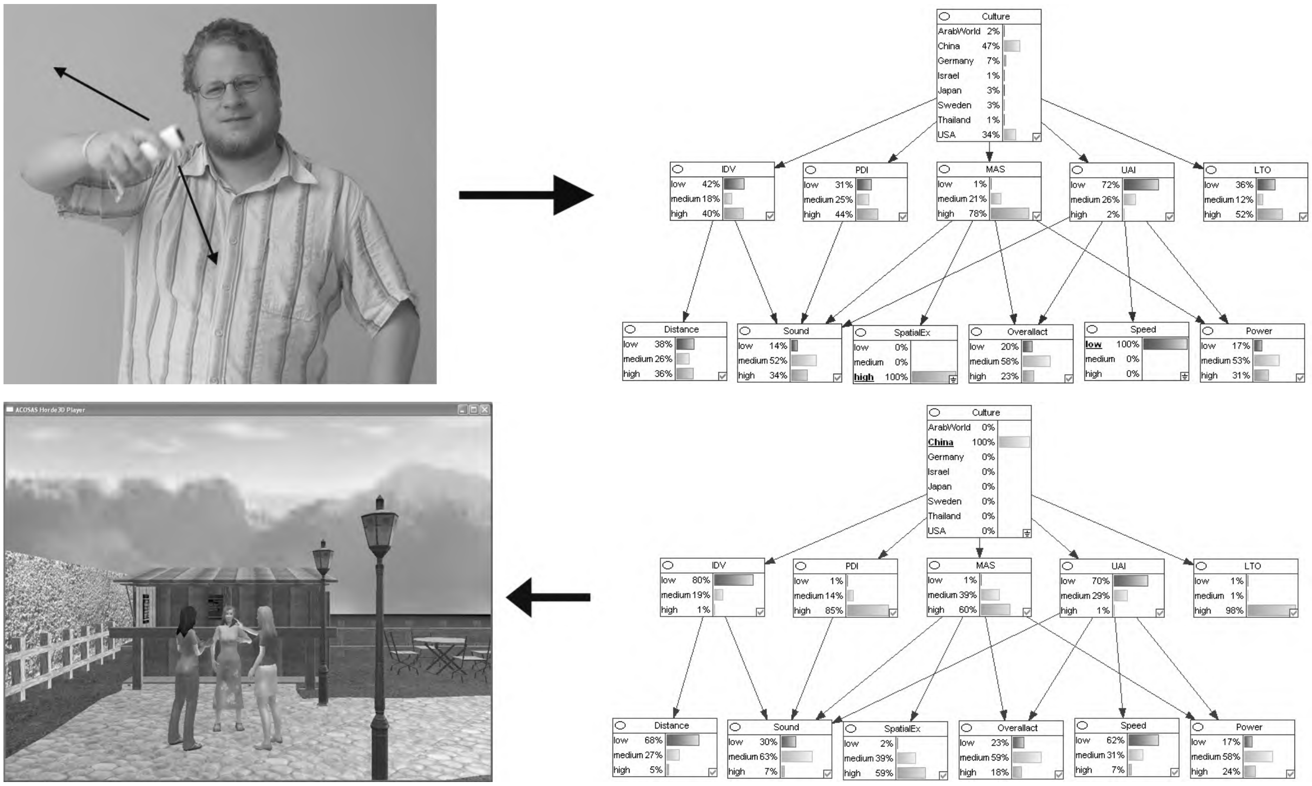

Figure 6: From user's cultural background (diagnostic inference) to culture specific agent behavior (causal inference). Example Chinese culture.

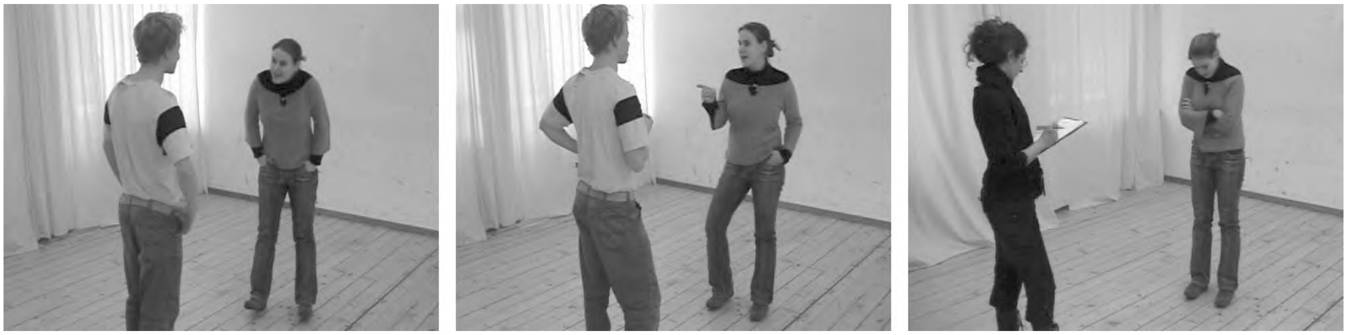

Figure 7: Actors rehearsing for the three prototypical scenarios: (i) First meeting, (ii) Negotiation, (iii) Status difference.

networks are a suitable modeling approach because their strength lies in coping with incomplete and unreliable knowledge.

4.4 Some caveats

The Bayesian network is based on what could be observed as the common behavior of a given culture. Of course there are idiosyncrasies in the behavior of every user. This is one reason why we are modeling the inference process by a Bayesian network. It is suitable to cope with unreliable and incomplete information. As long as the user does not deviate too far from the "prototypical" behavior of his culture, this will pose no problem. The result of the inference process is a probability dribution over the possible cultures. Thus, it might well happen that the user's cultural background has not the highest probability. At the moment, the culture with the highest probability is selected. But one can imagine different selection mechanisms that will take the probability distribution into account.

The behavior of agents is prototypical for a given culture and does not vary. This is a desired effect if the agents are meant to display the culture's behavior for demonstation purposes. In a real interaction it might get too boring if all agents show exactly the same behavior routines. If more idiosyncratic behavior of single agents is desirable, one could imagine a different selection process for the actual behavior. At the moment, the behavior with the highest probability is chosen. But one could also imagine a statistical decision process that takes the actual probability distribution of the output values into account which would prevent completely identical behavior modifications for every agent.

Although Hofstede did a comprehensive study to derive his dimensional theory, his theory is based on the analysis of questionnaires [10]. Thus, the proposed culture specific behavior might not be thoroughly grounded in reality. That is why we complement the model-driven approach with an empirical study of culture specific behavior which is shortly described in the next section.

5. GROUNDING THE MODEL IN EMPIRICAL DATA

If we talk about nonverbal behavior, we talk about observable communicative behavior. So far, there is hardly any principled study analyzing and comparing observational data from different cultures in a standardized way. To ground the model described in the last sections in hard empirical data we devised such a standardized observational study starting with two cultures that are located on different areas of the Hofstede dimensions, namely Germany and Japan.

Three prototypical interaction scenarios were defined that are found in every culture to allow for comparing the verbal and nonverbal behavior (see Fig. 7 for an impression).

1. Meeting someone for the first time: This is the standard first chapter of every language learning textbook and one of the most fundamental interactions in everyday communication.

2. Negotiating: Coming to an agreement with others can also be considered as a fundamental interaction esp. in intercultural communication. This scenario allows us to compare different negotiation styles and the accompanying verbal and nonverbal behavior.

3. Interacting with higher status individual: Cultures differ in how they interpret the unequal distribution of power and status among the members of the culture, resulting in quite different behaviors towards interaction partners with a higher status.

21 subjects (11 male, 10 female) participated in the German data collection, 26 subjects (13 male, 13 female) in the Japanese collection. For each subject, around 25 minutes of video material were collected, 5 minutes for the first meeting, 10-15 minutes for the negotiation, and 5 minutes for the status difference. To ensure a maximum amount of control over the recordings, subjects interacted with actors. The rationale for using actors as interaction partners was that we would be able to elicit sufficient interactions from the subjects.

Subjects were told that they would participate in an international study of negotiation styles which was conducted by a famous consulting company at the same time in different countries. One actor posed as the high status representative of this company. The other actor was introduced as another student who participated in the experiment.

The analysis of the data is pending at the moment and concentrates mainly on nonverbal behaviors like distance, position, speech volume, gestural expressivity, and body posture. The analysis of the verbal behavior, e.g. regarding different negotiation styles is scheduled as a second analysis step.

6. CONCLUSION

In this article we presented an approach how to make the notion of culture available for computational purposes. The model relies on a dimensional theory of culture and will be complemented in the future by empirical data that was collected for the German and Japanese culture. By defining the dimensional position of a culture, corresponding expressive behaviors can be derived. This behavior is prototypical according to the literature but it remains to be shown if it really resembles the target culture. To this end, a broad evaluation is necessary.

The model is tailored to the use in multimodal systems like embodied conversational agents. But other expressive behaviors are imaginable. For example, expressive system behavior could also be concerned with the presentation of information on a website. In this case different nodes have to be added to the Bayesian network to ensure that e.g. suitable colors are used or that there is a specific balance between text and images. Thus, we claim that our model is not restricted to the use with embodied conversational agents but tries to model the effects of the dimensional model of culture proposed by Hofstede and might thus be extended to comprise other types of expressive behavior.

7. ACKNOWLEDGMENTS

The work described in this article is funded by the German Research Foundation (DFG) under research grant RE 2619/2-1.

The authors would like to thank Prof. Toyoaki Nishida, Prof. Yukiko Nakano, and Hung-Hsuan Huang for their support collecting the Japanese corpus.

The Bayesian network model was created using the GeNIe modeling environment and integrated with SMILE developed by the Decision Systems Laboratory of the University of Pittsburgh (http://dsl.sis.pitt.edu).

8. REFERENCES

[1] J. M. Allbeck and N. I. Badler. Creating Embodied Agents With Cultural Context. In S. Payr and R. Trappl, editors, *Agent Culture: Human-Agent Interaction in a Multicultural World*, pages 107–126. Lawrence Erlbaum Associates, London, 2004.

[2] M. Argyle. *Bodily Communication*. Methuen & Co. Ltd., London, 1975.

[3] E. Ball. A bayasian heart: Computer recognition and simulation of emotion. In R. Trappl, P. Petta, and S. Payr, editors, *Emotions in Humans and Artifacts*, pages 303–332. MIT Press, 2002.

[4] N. Bee, H. Prendinger, A. Nakasone, E. André, and M. Ishizuka. AutoSelect: What You Want Is What You Get: Real-Time Processing of Visual Attention and Affect. In E. André, L. D. W. Minker, H. Neumann, and M. Weber, editors, *Perception and Interactive Technologies (PIT 2006)*, pages 40–52, Berlin, Heidelberg, 2006. Springer.

[5] E. Bevacqua, A. Raouzaiou, C. Peters, G. Caridakis, K. Karpouzis, C. Pelachaud, and M. Mancini. Multimodal sensing, interpretation and copying of movements by a virtual agent. In *PIT*, pages 164–174, 2006.

[6] 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.

[7] F. de Rosis, C. Pelachaud, and I. Poggi. Transcultural Believability in Embodied Agents: A Matter of Consistent Adaptation. In S. Payr and R. Trappl, editors, *Agent Culture: Human-Agent Interaction in a Multicultural World*, pages 75–106. Lawrence Erlbaum Associates, London, 2004.

[8] GeNIe and SMILE. http://genie.sis.pitt.edu/.

[9] Hofstede. http://www.geert-hofstede.com/hofstede_dimensions.php.

[10] G. Hofstede. *Cultures Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Sage Publications, Thousand Oaks, London, 2001.

[11] G. J. Hofstede, P. B. Pedersen, and G. Hofsted. *Exploring Culture - Exercises, Stories and Synthetic Cultures*. Intercultural Press, 2002.

[12] G. J. Hofstede, P. B. Pedersen, and G. Hofstede. *Exploring Culture: Exercises, Stories, and Synthetic Cultures*. Intercultural Press, Yarmouth, 2002.

[13] F. V. Jensen. *Bayesian Networks and Decicion Graphs*. Springer, 2001.

[14] W. Johnson, S. Choi, S. Marsella, N. Mote, S. Narayanan, and H. Vilhjálmsson. 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.

[15] F. Kluckhohn and F. Strodtbeck. *Variations in value orientations*. Row, Peterson, New York, 1961.

[16] A. Marcus and E. W. Gould. Crosscurrents: Cultural Dimensions and Global Web-User Interface Design. *ACM Interactions*, 7(4):32–46, 2000.

[17] J.-C. Martin, S. Abrilian, L. Devillers, M. Lamolle, M. Mancini, and C. Pelachaud. Levels of Representation in the Annotation of Emotion for the Specification of Expressivity in ECAs. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist, editors, *Intelligent Virtual Agents (IVA)*, pages 405–417, Berlin, Heidelberg, 2005. Springer.

[18] H. Noot and Z. Ruttkay. Variations in Gesturing and Speech by GESTYLE. *International Journal of Human-Computer Studies, Special Issue on Subtle Expressivity for Characters and Robots*, 2005.

[19] B. Reeves and C. Nass. *The Media Equation — How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, Cambridge, 1996.

[20] M. Rehm, E. André, and M. Nischt. Let's Come Together — Social Navigation Behaviors of Virtual and Real Humans. In M. Maybury, O. Stock, and W. Wahlster, editors, *Intelligent Technologies for Interactive Entertainment*, pages 124–133, Berlin, Heidelberg, 2005. Springer.

[21] M. Rehm, B. Endrass, and M. Wissner. Integrating the user in the social group dynamics of agents. In *Proceedings of Social Intelligence Design (SID)*, in press.

[22] P. Sengers. The Agents of McDonaldization. In S. Payr and R. Trappl, editors, *Agent Culture: Human-Agent Interaction in a Multicultural World*, pages 3–20. Lawrence Erlbaum Associates, London, 2004.

[23] Y. Takeuchi, Y. Katagiri, C. I. Nass, and B. J. Fogg. Social Response and Cultural Dependency in Human-Computer Interaction. In *Proceedings of PRICAI*, pages 114–123, 1998.