To appear in HCI International 2009 track on Online Communities and Social Computing, July 2009

Developing Believable Interactive Cultural Characters for Cross-Cultural Training

Glenn Taylor[1], Ed Sims, Ph.D.[2]

[1] Soar Technology, Inc., 3600 Green Court Suite 600,
Ann Arbor, MI 48105
glenn@soartech.com
[2] Vcom3D, Inc., 3452 Lake Lynda Dr., Suite 260,
Orlando, FL 32817
eds@vcom3d.com

Abstract. One path to training for cross-cultural competency is through immersion in a target culture, but real immersion can be expensive. Virtual immersion may be a viable alternative, but only if the experience is realistic and compelling enough. The characters in the virtual environment must embody the behaviors of the people in that culture in a way that is realistic and believable to facilitate training. In this paper, we describe a theory-based framework for building interactive cultural characters for cross-cultural training. The framework combines physical and cognitive aspects of behavior to enable more realistic generation of cultural behavior. We describe the theoretical basis for the framework, how it is being used to build interactive characters for cross-cultural training, and reflect on the challenges we have faced and lessons we have learned in doing this work.

Keywords: Cross-Cultural Training, Online Communities and Social Computing

1 Introduction

In this work, our aim is to develop interactive 3D avatars for the purpose of training people in cross-cultural, human-human interaction. While computer-based language training is widely available, effective cross-cultural communication has typically required interactions with natives of that culture. Our aim is a kind of cultural immersion typically available only in real-world situations. The key requirement here is the development of believable interactive characters that act in ways appropriate to the culture of interest, at multiple levels of detail. Not only must the characters have the right physical appearance and speak the right language, they must also capture the richness of the interactions, the expectations that people from that culture would have about those interactions, and the responses to those interactions when the expectations are met or broken.

Our work combines two distinct approaches – one modeling the physical behaviors of cultural characters, the other modeling the cognitive behaviors. We blend these

approaches to provide rich, interactive cultural avatars for cross-cultural training. In both cases, our approach derives from cultural studies, ethnographies, and theories about culture and cultural behavior, which motivates a modeling framework approach. The framework is general enough to be applied to building models of different cultures by providing different content into a common execution engine. In this paper, we describe our efforts in developing cultural characters for training purposes, how they can fit into training systems and training lessons, and our own lessons learned along the way.

2 Needs in Cross-Cultural Training

For our work here, we have focused on a few key aspects of cross-cultural training, especially with regard to communication: interaction protocols and non-verbal communication. Note that this is not language training per se, though verbal language is a part of it. Instead, we focus on what kinds of language are appropriate in different situations, and how those native to a culture might behave and expect others to behave in these situations.

The goal of a trainee might be to establish rapport in anticipation of a long-term business relationship or to establish trust such that the native person might share information important to an investigation. Acting in a culturally inappropriate way could draw offense and make cooperation difficult.

2.1 Interaction Protocols

Greeting someone of high status is typically different than greeting strangers on the street. For example, greeting an important community leader with "What's up?" might be deemed inappropriate and immediately offend the addressee. Business meetings are often conducted differently than informal meetings among friends and family. For example, business meetings in Iraq typically begin with tea and conversation as a way to develop rapport and only later turn to key business exchanges. To hurried Americans, tea and conversation might seem like a pointless diversion from the "real" purpose of the meeting.

These kinds of interactions, and the protocols involved in them, tend to be culture-specific. Those within the culture learn how to behave appropriately in these situations through observation and practice while interacting with others in that culture. To be effective as an outsider, learning these protocols and how to act within them is critical.

2.2 Non-verbal Communication

It is also important as a trainee to learn to recognize the non-verbal aspects of communication that can also be culture-specific. Although there is strong evidence that the interpretation of facial expressions of emotion is universal throughout all cultures [1], most gestures are learned from one's culture. Furthermore, even though

the meaning of facial expressions is universal, the "display rules" for when it is appropriate to show or suppress these emotions are quite variable [2]. Learning when to use particular gestures or when to exhibit emotion, and learning to recognize these aspects of communication, is a key element of learning effective cross-cultural communication. (For example, the "peace" sign common in the US is offensive in Australia and England. In some cultures, nodding while listening might not mean agreement or assent but is simply a way to say "I understand.")

Non-verbal cues used in the Arab world are often quite different from those used in America or in Western Europe. Barakat [3] documented 247 gestures used in the Arab world. We found that many of these are understood by persons from a widespread area of the Middle East today; but few are understood by Americans. In order to build a more current database of nonverbal signals used in conversation today, we identified over 200 gestures from actual interviews with citizens in Baghdad. These included not only "emblem" gestures, but also gestures used for controlling dialog, such as the turn-taking gestures in Fig. 1 (borrowed from [4]).

Fig. 1: Gestures in conversational Iraqi Arabic.

An example we learned from working with soldiers involved in operations in Iraq is the use of "Insha'allah" in Arabic. Literally this means "God willing" and is used whenever speaking of events that might happen in the future. For example, "We will meet tomorrow at 9am, God willing." Taking the speech by itself, some listeners might think that the speaker won't show up for the proposed meeting. However, this is an incorrect assumption: this is simply how native speakers use Arabic to speak in the future tense. Clues of the real intent of the speaker must be found in other elements of the communication, such as the tone of the utterance or in the accompanying gestures.

3 Physical and Communicative Cultural Behavior in 3D Avatars

The most apparent aspects of a culture are the "surface features," including the style of dress, the way in which people speak, etc. These are also the most apparent in developing a physical avatar representing someone from another culture and have to reach a threshold of believability before a training system will be taken seriously. To our advantage, technology advances in graphics for 3D characters, as evidenced by modern video games and movies, have reached a point of naturalness and photo-realism that the appearance of characters is sometimes taken for granted. Even movement patterns such as in walking or lip synching to voice can seem natural.

Challenges remain, especially in the generation of appropriate content for these avatars.

Of course, there are many subtleties of communication that are not immediately apparent to someone unfamiliar with that culture: gestures, eye contact behaviors, and the like. The focus of our physical 3D models is in the generation of these subtleties in a way that increases the believability of the model for a particular culture. It is only in building cultural avatars that include these non-verbal aspects that we can hope to teach effective lessons.

The system underlying the generation of physical behavior is called Vcommunicator [5], which includes both a 3D animation engine as well as tools for developing 3D avatars. Vcom3D has developed a library of culture-specific avatars, gestures and expressions that can be invoked on demand. These libraries consist of over 60 culturally diverse virtual human models as well as 40 facial expressions and 500 gestures, and can automatically lip-sync to over 22 mouth shapes that map to over 100 speech sounds of the International Phonetic Alphabet. Fig. 2 shows some screenshots of a cultural avatar using a few gestures. Primitive gestures and expressions can also be composed to larger animation sequences.

Fig. 2: 3D cultural character with a sample of gestures.

4 Cultural Cognitive Architecture

For cultural training to be immersive, the characters that inhabit that space must behave in ways consistent with the culture and, more importantly, interact with the trainee in ways consistent with the culture. Interactivity in any real sense is possible only with deep models of behavior. To this end, we have been developing a framework for building rich cultural models for driving interactive characters. This framework, called the Cultural Cognitive Architecture (CCA) [6], is based on theories of human cognition ([7, 8]) and culture, including Cultural Schema Theory ([9, 10]).

Cultural Schema Theory itself builds from theories of cognition, positing that much of culture as demonstrated through behavior is drawn from knowledge learned by living in that culture: knowledge about the correct ways to interact (e.g., norms), about what is important in the culture (e.g., values), etc. This knowledge is encoded as *schemas* that represent, among other things, the relationships between important

things in the environment, the expectations about how situations might play out, and goals different characters in the situation might have. With schemas come representations of these concepts, as well as a process of connecting elements in the environment to the schema. This process allows cultural situations to be recognized and understood. It also allows for appropriate goals to be generated within the context of the situation.

An example interaction schema for ways to engage in small talk is given in Fig. 3. The schema is hierarchical – there are multiple ways to accomplish engaging in small talk (talk about family or talk about football). The schema also contains a protocol for talking about family – a sequence of asking and telling (indicated by the horizontal arrow below), including a reciprocation of asking. Not reciprocating might be interpreted as rude. At this level of description, this schema could probably be applied to many cultures around the globe. However, highlighting again the hierarchical nature of schema, finer detail in each of the leaves given here might have culture-specific rules. For example, among two men talking in Arab cultures, it may be improper to speak or ask specifically about the females in either family. Likewise, an American showing a picture of his family at the beach might be seen as breaking rules of social propriety. The hierarchical organization enables us to capture the regularities of cultural behavior, allowing for re-use across situations and even cultures, while at the same time enabling the encapsulation of more specific cultural differences in the leaf nodes.

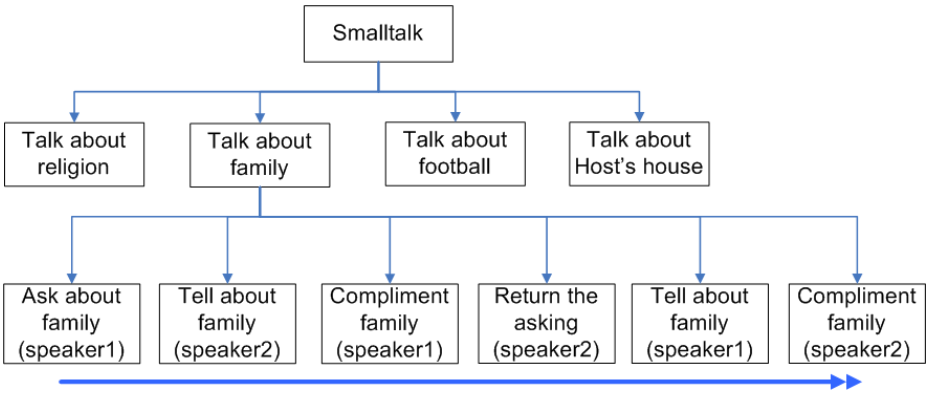

Fig. 3: An example cultural schema in CCA.

In addition to Cultural Schema Theory, CCA also incorporates theories of human emotion, specifically based on appraisal theories of emotion, per Scherer [11]. In CCA, perceived events are appraised along a number of dimensions such as how well the event fits with the avatar's goals (*goal conduciveness*), how well the event fits within established norms *(internal and external standards)*, and how surprising or novel the event is. We implement a subset of Scherer's appraisal dimensions and emotion dimensions. The outputs of the appraisal-based emotion system include the appraisal for a particular event, an immediate emotional state based on that appraisal, and an updated running average emotional state (updated with every event) that allows for a more coherent basis for generating behavior. It is this process that

accounts for the interpretation of actions as rude or improper. For example, if the next step in the "small talk schema" above is for the trainee to "return the asking," and the trainee does not do this, the act of breaking the schema generates a negative emotional state.

Building a cultural model consists of encoding schema drawn from the culture in question, as well as constructing the appraisal process to assign culture-specific appraisal of perceived events or situations. For the purposes of cultural training, a cultural character includes knowledge of the culturally correct interaction patterns relevant to a particular training situation. A trainee interacts with this cultural character by choosing from a set of actions (for example, things to say or do in the situation – asking about family, removing hat and glasses). The cultural avatar processes these trainee actions against its own expectations about the interaction, represented as schema and appraisals. In this process, the cultural character generates an emotional response to the trainee's actions. The avatar's action choice is selected based on a combination of what the current situation suggests and the computed emotional state of the agent as a result of observing the trainee's actions.

5 Connecting Physical and Cognitive Models for Training

The two systems we have described above, Vcommunicator and CCA, were developed independent of one another, so at one level we have a straightforward engineering task of connecting two pieces of software. This is guided by good software engineering practices of encapsulation, modularity, loose coupling, etc. However, in a real person with real physical-cognitive behaviors, these two systems are tightly coupled with many feedback loops. Integrating these systems becomes an exercise of balancing the needs of good software design with the requirement for theory-driven, believable 3D cultural avatars. Fig. 4 illustrates a schematic of the integration of these different components to support a training environment.

For our implementation, the cognitive system is in charge of high-level goal-setting and high-level behavior generation. As described earlier, cognitive behavior generation occurs through the matching of situation-dependent schema and the generation of emotional state, which itself may trigger schema. Where a schema indicates the 3D cultural avatar should take an action, the cognitive model exports to the physical model the selected action as well as the immediate appraisal and the emotional state of the character at that point in time. Actions can be selected based on responses to external events, emotional states of the agent, or both.

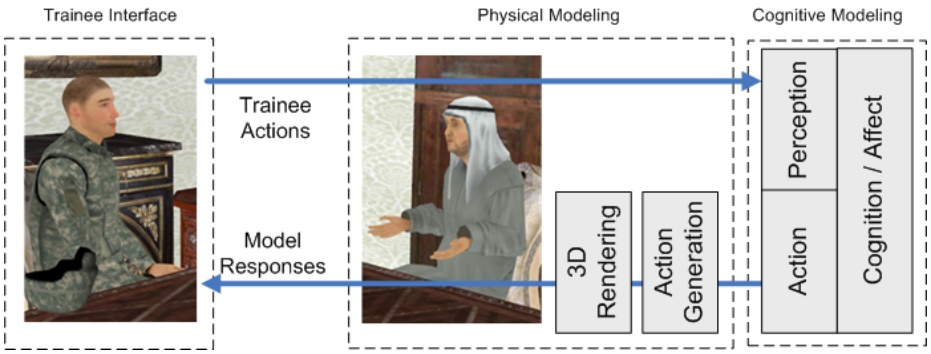

Fig. 4: Schematic of integrating a physical and cognitive model into a training environment.

The physical model receives the action to perform and the emotional state, and uses the emotional content to vary some of the parameters in the generated physical action, such as the emphasis that is placed on some gestures. The culture-specific gestures are drawn from a pre-developed library of behaviors, but some of their parameters can be modified on the fly during generation. For example, a gesture to gain control of the conversation with a low anger rating would be somewhat muted, whereas with a high anger rating would be much more punctuated physically. The emotional state might also influence facial expressions (e.g., surprise by raising eyebrows), or ambient behaviors, such as eye gaze, blinking rates, or shifting weight between feet. All of these contribute to the overall physical manifestation of the emotional state of the agent. It is through this physical appearance of the emotion that the trainee can learn to pick up on non-verbal cues.

We have not yet drawn connections from the physical back to the cognitive. A fully realistic model would obey the physical limits of perception – seeing only what the eyes are looking at, hearing only what is in range, etc. Similarly, the 3D avatar does not perceive emotional or other non-verbal content in the trainee's actions. In the training vignettes we have been developing, this level of fidelity has so far not been required.

Fig. 5: Scene from rapport-building training scenario.

Fig. 5 illustrates our immersive training environment that includes multiple 3D avatars, one of which is played by the trainee. The trainee can act on objects in the 3D environment using context-sensitive object selection menus (e.g., "remove helmet") or can choose from a list of context-sensitive speech acts (not shown).

6 What Is Believable Cultural Behavior?

What constitutes "believable" behavior in a cultural avatar? How do we test that the models we have developed have validity and utility? These are all connected concepts, though not the same. A model of human behavior may not be validated but can still have utility. A model can be believable but not validated. All of these measures must be taken in the context of the use of the system.

The idea of validating models of human behavior is a challenge in general, much different than validating the behavior of physical systems like models of an aircraft or a car chassis. Simply the breadth of human behavior is too great and too variable to determine if a model is completely valid. For our work here, we have clearly narrowed the scope of what we expect to model in 3D characters and their interactions, so any evaluation needs to be in this more limited scope. Methods for evaluating the *validity* of human behavior models often fall to "face validation": putting the models in front of a group of experts (e.g., in this case cultural anthropologists or even natives of the culture under study) and eliciting their subjective view of how well the model behaves as compared to how they would expect a human from that culture to behave in the same situations. This includes the overall interaction protocols (e.g., from the cognitive schema) that the 3D avatar appears to be following, as well as the finer-grained cognitive and physical manifestations of the avatar's behaviors in speech, gesture, and emotion. Likert Scale questionnaires can be used to gather observers' subjective ratings of the system's believability in the context of the situation (see, for example [12]). However, even a simple system can exhibit a range of outward behavior from which it may be difficult to tell if the behavior is being generated for valid reasons. Inspection at multiple levels of a model is often required to be able to assess its validity.

Evaluating the *utility* of the model in the context of training may entail putting trainees through a training course and assessing the how well the cultural models contributed to the trainees' learning, especially when transferred to a real-life, cross-cultural interaction setting. This is often a much more involved procedure, requiring a sufficient number of participants and tracking their performance over time, then comparing to control cases where traditional methods of cultural training were used.

In our work here, we have so far conducted informal feedback sessions with experts to evaluate the models' face value, mostly as part of improving the models during development. More formal evaluations of validity and utility will be performed as we mature the overall system.

6 Conclusions and Lessons Learned

Rather than develop a one-off model of a single culture, our approach has been to develop frameworks, architectures, and libraries that can be instantiated for multiple cultures, given appropriate data sets. These architectures have been guided by cross-cultural theories, including those related to expression of non-verbal communication and human cognition. We believe that deep models of cultural behavior present the best opportunity for building realistic models for effective cross-cultural training. However, this approach introduces a number of interesting challenges.

One challenge has been in finding the right divisions and connections between the physical and cognitive systems that lend themselves both to straightforward engineering and theoretical validity. One example of this is in choosing an appropriate level of granularity for model actions, cognitive as well as physical. Specifying actions at a very fine level might result in high-fidelity models and interactions, but this places a higher burden on model development. Specifying actions at too gross a level might make engineering easier, but reduces the realism of the characters and therefore the overall training experience. Finding the appropriate level remains an art.

Data availability also presents issues. Rich models require rich data. Particular training vignettes require data collection related to those vignettes, which can be costly. Online virtual worlds have been suggested as potential sources for cultural data, but the fidelity of the data would be only as good as the fidelity of the interactions available in the virtual world – most virtual worlds have limited body articulation so gestures are typically not culturally realistic; even linguistic exchanges using chat is corrupted by the medium in many ways. The best sources of data remain the most expensive to gather: ethnographies of the cultures of interest. We have here used audio-visual resources to gather data on gestures. For example, we have used existing literature and culture guides to understand other aspects of a culture relevant to training. We have supplemented these resources with experts in the culture and have also relied on training objectives to help narrow the scope of what to include in the models. We have tried where possible to get multiple perspectives to build more robust models. It is not uncommon for different sources to offer contradictory positions on, for instance, the meaning of a phrase or the expected behavior in a given situation.

Another challenge is in natural human interfaces for interacting with 3D cultural avatars. Natural language technology such as speech interfaces or language understanding systems are still not generally robust enough to conduct open-ended conversations with computer-based characters. Typical applications instead settle on small subsets of spoken utterances (e.g., [13]) or menu-driven interfaces that can reduce the overall realism of the interaction (e.g., [14]). Likewise, we have focused on the frameworks and model building rather than natural human interfaces, though advances here are clearly needed for more effective training.

Acknowledgments. This work was performed under US Government contract W91WAW-08-C-0052, funded by the Department of Defense, Director of Defense Research and Engineering and monitored by the US Army Research Institute for the Behavioral and Social Sciences. The views, opinions, and/or findings contained in

this paper are those of the authors and should not be construed as an official Department of the Army position, policy, or decision.

References

1. Ekman, P.: Universals and cultural differences in facial expressions of emotions. In Nebraska symposium on motivation. Lincoln, NB: University of Nebraska Press (1972)
2. Matsumoto, D., et al.: The contribution of individualism-collectivism to cross-national differences in display rules. Asian Journal of Social Psychology. **1**: p. 147-165. (1998)
3. Barakat, R.: Arabic Gestures. Journal of Popular Culture. p. 750-794. (1973)
4. Antoon, S., et al.: DVD: About Baghdad. AFD Studio. p. 90 minutes (2005)
5. Sims, E.: Simulating Believable, Context-aware and Culture-specific Human Behaviors. In IITSEC. Orlando, FL (2005)
6. Taylor, G., et al.: Toward a Hybrid Cultural Cognitive Architecture. In CogSci Workshop on Culture and Cognition. Nashville, TN: Cognitive Science Society (2007)
7. Newell, A.: Unified Theories of Cognition. Cambridge, MA: Harvard University Press (1990)
8. Schank, R. and R. Abelson: Scripts, Plans, Goals, and Understanding: An enquiry into human knowledge structures. Hillsdale, NJ: Earlbaum Assoc. (1977)
9. D'Andrade, R. and C. Strauss, eds: Human Motives and Cultural Models. Cambridge University Press: Cambridge, UK (1992)
10. Shore, B.: Culture in Mind. Oxford, UK: Oxford University Press (1996)
11. Scherer, K.R.: The Role of Culture in Emotion-Antecedent Appraisal. Journal of Personality and Social Psychology. **73**(5): p. 902-922. (1997)
12. Reilly, W.S.N.: Believable Social and Emotional Agents, in School of Computer Science. Carnegie Mellon University: Pittsburgh, PA (1996)
13. Johnson, W.L. and A. Valente: Tactical Language and Culture Training Systems: Using Artificial Intelligence to Teach Foreign Languages and Cultures. In Innovative Applications of Artificial Intelligence. Chicago, IL (2008)
14. Rosenberg, M., et al.: A Language for Modeling Cultural Norms, Biases and Stereotypes. In Behavior Representation in Modeling and Simulation (BRIMS). Providence, RI: SISO (2008)