

Cultural Frame-Switching using Accented Spoken Language by a Virtual Character

Morteza Dehghani, Peter Khooshabeh,
Lixing Huang, Lia Oganessian, and Jonathan Gratch

{morteza,khooshabeh,lhuang,hovanessian,gratch}@ict.usc.edu
Institute for Creative Technologies,
University of Southern California
12015 Waterfront Drive
Playa Vista, CA 90094

Abstract. In this paper, we examine whether embodied conversational agents can be used to implement socio-cultural markers. We investigate whether the accent of a virtual character, as a marker for culture, can cause cultural frame-shifts in individuals. We report an experiment, performed among bicultural and monocultural individuals, in which we test the above hypothesis. Our results show that a virtual agent can have a socio-cultural effect on people’s cognition. This work makes unique contributions to the design and evaluation of intelligent virtual agents as well as the theoretical psychological literature.

1 Introduction

Virtual agents, particularly when rendered as embodied conversational characters, are capable of providing a compelling multimedia platform that serves as an effective interface for research purposes, educational applications or entertainment. Embodied conversational agents (ECAs) make it possible to manipulate external features such as visual appearance, speech type, and contextual graphical environments. This ability makes ECAs a convenient platform to isolate unique cultural characteristics and realize them through simulation. Along with enhanced experimental control, ease of manipulations, consistency and controlled measurements [10], these features make ECAs useful and reliable tools for studying cultural cognitions. For example, we can objectively pinpoint certain social and behavioral characteristics that are relevant to specific cultures and implement them using virtual agent technology.

In this paper, we examine whether *accent* of an ECA can be used as a marker for culture by evoking cultural *frame-switching* [6] in bicultural individuals. Cultural frame-switching refers to the idea that interpretive frames, in individuals who have internalized two cultural identities, can shift due to situational cues [1]. For example, Hong et al. [6] prime Chinese-Americans with American, Chinese or neutral iconic images and demonstrate that participants in the Chinese

primed condition interpreted the next task with more of a Chinese interpretive lens (focused more on external attributions) than those in the American or control conditions. On the other hand, participants in the American primed condition projected more American cultural values by focusing on individual attributions for the same task. A plausible explanation for frame-switching is that multi-cultural individuals have different norms and culturally specific systems of meaning [2], which can get activated depending on the context and the social environment. Hence, activation of different cultural interpretive frames can result in varied constraints on the individual’s psychophysical behaviors.

The shift in the interpretive frames can be especially notable if the differences in normative behavior between the bicultural individuals’ cultures are significant. There is growing evidence in the social sciences that culturally normative behaviors vary across cultures (see [5] for a review). In other words, behaviors that are considered typical in one culture may be considered abnormal in another. The variability of culturally normative behaviors and cognitions have been noted in different aspects of human behavior. Related to this work, researchers have shown cultural differences in morally motivated decision-making by identifying moral domains that are present (or salient) in some cultures and not in others [13,3]. Domains such as respect for authority and the saliency of the distinction between purity and impurity are some that have been identified in helping people to characterize certain situations as morally tinged within one cultural group but not another.

In this work, we carefully control for non-verbal behavior and manipulate solely the virtual agent’s accent of spoken English and thus measure the effect on the participants’ perception and cognition. Based on the previous literature, we predict individuals interacting with a culturally congruent agent will use moral frames specific to that culture.

2 Related Work

There have been a small number of studies investigating how culturally congruent virtual agent characteristics can affect users’ cognition. In an effort to examine the interaction between culture and ECA design in the domain of education, Rader, Echelbarger, and Cassell [12] developed virtual peers that matched the dialect of children speaking African-American English and asked the children to complete a bridge building exercise. The children alternated playing the role of student and teacher as they explained the building process. Rader et al. found that students who tend to speak more dialected English did so less when they played the role of a teacher. This work suggests that the virtual peer and culturally congruent context, coupled with the role switch, influenced students to speak mainstream English, which is shown to be related to higher student achievement.

In another line of work, Yin, Bickmore, and Cortes [15] report that individuals who process information using peripheral cues were influenced by an agent tailored to their own culture. An examination of their cultural agents shows

that they might have been at a home-field disadvantage [11], where the cultural groups scholars select to study get ‘marked’, which implies that the home-field culture of the researchers goes ‘unmarked’ and its cultural nuances become unnoticed. Yin et al. had two female agents; one Latina and the other Anglo. While the Latina agent was visibly busty, in a colored room, and pictured with a Frida Kahlo painting, the Anglo agent was in a sterile, white room, with only a college banner. The researchers seemed to be biased in their own (American) cultural values and therefore over-marked the Latino culture.

3 Experiments

We experimentally model *frame-switching* among bicultural (and monocultural) individuals using ECAs and measure if their preference for certain moral acts are affected by this manipulation. Our hypothesis is that the accent of a virtual agent should affect people’s perception of the culture of the agent and, therefore, can be used as a marker for a culture. If this premise holds, then virtual humans that have an accent that is congruent with a participant’s culture will bias that participant to use the congruent cultural frame.

In order to test our hypothesis, we designed an experiment in which we control for non-verbal behavior of an ECA while manipulating only its accent. We recruited Iranian-American and American participants and had them read a story which included Iranian and American cultural values and customs. The participants were then asked to summarize the story and answer a few questions about the material.

3.1 Method

Participants Fifty-two Americans (mean age = 40) and Fourteen Iranian-Americans (Iranians living in Southern California for more than 5 years) (mean age = 34.61) participated in this study. The participants were recruited using craigslist.com and snowball sampling, which consisted of asking subjects to refer other subjects for the study. Each subject received \$25 at the end of the experiment for participating in the study. The participants were not aware that they were participating in a culture study.¹ Each participant completed the task in individual experimental sessions.

Materials

¹ We note that participants performed three other studies in which this experiment was one of the tasks. The order of the experiments was not fully counterbalanced, however this task was performed in a different lab compared to other studies and was run individually and not in groups.

Yesterday was an interesting day. I got out of class at around 4:30. The weather was great and spring was in the air. My classmate Shawn, who I'm not really good friends with and don't know too well, invited me to a bonfire at the beach. I told him that I was hungry and needed to get something to eat before going to the beach. He was going to his parent's house for dinner and invited me over. I don't know him well, so I initially refused his offer. But he kept on insisting that it's the beginning of Spring and I should go and have dinner with his family, he said then we can all go to the bonfire together. On the way to his house, Shawn asked why I was wearing almost all green. I thought it was a strange question as a lot of students were wearing green that day.

I went over to his house and met his dad at the living room. I thought to myself that the apple doesn't fall far from the tree. Upon seeing his father, Shawn introduced me to him, saying this is my good friend Anthony. He seemed surprised by my presence. Now I wasn't sure if they were expecting a guest for dinner. Then Shawn's mother came to the living room. I had met her before. She used to work in the registrar of the school. I said hi to her and she greeted me back saying that it looked like water had gone under my skin.

They had some apples on the table, they also had some coins right by the apple. He offered some salad to me, I asked if it was okay to chop up the garlic which was conveniently already on the table by the apples, which I think he had forgotten to put in the salad. He hesitated and took a few seconds before saying yes...

Earlier when I was at Jamba juice, I had seen the same green grass that Shawn had on his table. I asked his dad whether they drink the juice of the green grass. He looked at me as if he had not understood what I was talking about. So, I told him that grass juice is very nutritious and includes a broad spectrum of vitamins, minerals and antioxidants. As he was looking out the window, he said that water is past his head.

I finished my salad and asked Shawn when we can leave to see the bonfire. Suddenly Shawn and his family members all started saying that I should stay for dinner and go back to their place after the bonfire because my apartment was far from their house. I didn't understand why they were asking me to stay for dinner, because I think they didn't have any dinner prepared. I told Shawn that I need to go and watch the game later at night. My favorite college team was in the second round of a tournament. A bunch of the guys from the baseball team were going to watch the game at the local watering hole. I really needed to go to blow off some steam. But he said that his dad is already making food for us. Shawn then asked me how I've lost so much weight in the past couple of months. I told him you will not lose weight until you give up carbs, and I suggest you go cold turkey.

I saw a very small art piece on the wall of their hallway. It was a picture of some Chinese looking guys playing with a ball on horses. I told his dad that this is a lovely picture. He thanked me and told me that I could take it. I first thought he was kidding, but he seemed serious and told me that he wants me to take it. Shawn's mother also said that it will look better in my house and I should take it. Given that they were insisting so much, and the piece didn't look expensive I took it and thanked them for it. Shawn's parents told Shawn that they had to wait for another hour and a half to serve dinner because some family friends had just called and were coming to visit them for the first day of spring. Given that I had plans to go watch the game and couldn't wait that long, I got up thanked them for the picture and salad, and had my friend pick me up. Shawn and his parents insisted that I should wait and have dinner with them. But my friend was already there to pick me up. She had put green papers all over her car, and had beer too.

Fig. 1. Stimuli read by participants

Design The study employed a between subject 2 X 2 full factorial design. The first factor is the culture of the participants (American or Iranian-American). The second factor is the agent's accent, which was either a standard American English or Iranian English accent, spoken by the virtual agent.

Stimuli The participants were first asked to read a short story (Figure 1). The story included a balanced number of American and Iranian cultural products (e.g. proverbs), values, and events (Iranian celebration of New Year compared to Saint Patrick's day²). None of these idea units were explicitly labeled with their cultural referent (there was no explicit reference to St. Patrick's day as such) and the idea units were interleaved so as to minimize memory distortions due to recency or primacy. After reading the story, participants interacted with an ECA.

Rapport Agent The agent used in this experiment, Utah [4] (Figure 2), is designed to establish rapport with human participants by providing contingent feedback while the human speaker is speaking. To produce feedback, the agent first detects and analyzes in real-time the human speakers' audiovisual features, which are silence, head nod, eye-gaze (looking at the agent or not) and smile. The audio feature detector extracts intensity from the raw signal every 100ms using

² Saint Patrick's day is widely celebrated in north America



Fig. 2. The ECA used in our Experiments

Praat³. With the intensity information, it outputs a binary feature, speech or silence, every 100ms. The visual feature detector⁴ tracks the position of the face, the facial feature points, the direction of eye gaze and the smile level. With this information, it outputs visual features indicating whether the human is nodding or not, looking away or not, and smiling or not. Based on the perceived audiovisual features, the response model [7] decides, in real-time, the most appropriate responses, such as head nod and smile. These different styles of animations are first converted into Behavior Markup Language (BML) [9] and then sent to an action scheduler, which keeps track of the duration of each animation. If the current animation has not been completed, the new animation will be ignored. The BMLs are passed to Smartbody [14], a virtual human animation system designed to seamlessly blend animations and procedural behaviors. Finally, the byproducts of Smartbody are rendered by a commercial game engine, Gamebryo⁵, and displayed to users. For the experiment, the voice of the ECA was prerecorded using the voice of the second author, whom is familiar with both Iranian and American cultures. The second author did not participate in recruitment nor in running the experiments.

³ Praat, <http://www.fon.hum.uva.nl/praat/>

⁴ OKAO Technology, http://www.omron.com/r_d/coretech/vision/okao.html

⁵ <http://www.gamebryo.com>

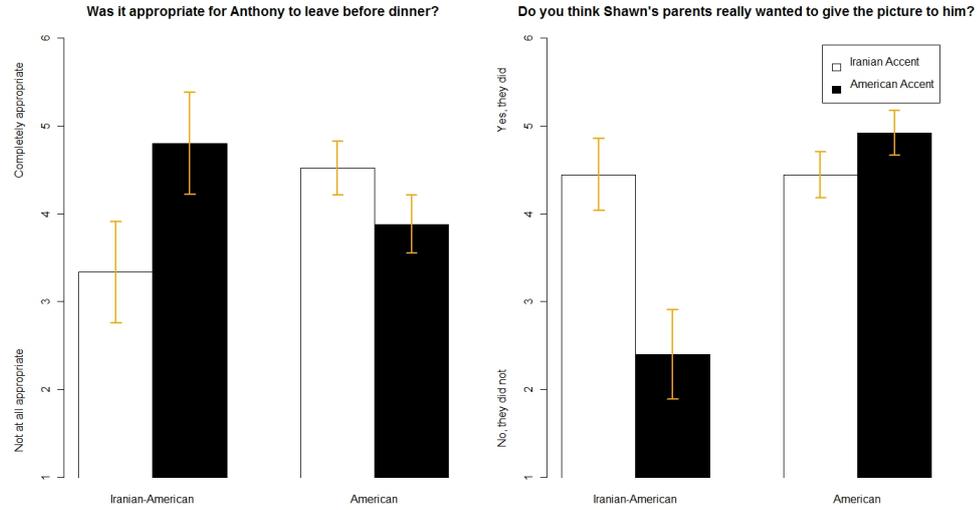


Fig. 3. Interaction between culture and accent of the agent

3.2 Procedure

After participants finished reading the story, the virtual agent greeted them, explained an overview of the research at our research center and asked them to summarize the story they had just read. Next, they filled out a questionnaire about the appropriateness of certain actions and intentions of the characters within the story. They were specifically asked the following two questions: 1. Was it appropriate for Anthony to leave before dinner? 2. Do you think Shawn's parents really wanted to give the picture to him?

Our hypothesis predicts that participants should use culturally congruent frames to interpret and answer these moral questions. The Iranian cultural frame suggests that it is not appropriate to refuse someone's generosity and hospitality. If cultural frame-shifting does indeed take place for Iranians-Americans when interacting with the culturally congruent agent, then they should say it is inappropriate for Anthony to leave dinner early. For the second question, the Iranian frame could interpret the event as an instance of Iranian hospitality, especially when it comes to sharing their cultural artifacts (in this case Persian miniature).

Next, participants were asked several questions about different emotions of the characters in the story. Lastly, to check the effectiveness of our manipulation, participants answered the following two questions in a random order: 3. Did the agent have more of an American accent or Middle-Eastern accent? 4. Did the agent appear more Western or more Middle-Eastern? Each question was answered on a 6-point scale (1 = No they did not, 6 = Yes they did; 1 = Not at all appropriate, 6 = Completely appropriate; 1 = Very much American, 6 = Very much Middle-Eastern; 1 = Very much Western, 6 = Very much Middle-Eastern).

3.3 Results

For both manipulation check questions, we used the responses to questions 3 and 4 as dependent variables in a 2 X 2 ANOVA, where the first factor was the culture of the participants (American or Iranian-American) and the second factor was the accent of the agent (American or Iranian). There was a main effect of agents' accent for both questions (appearance: $F(1, 62) = 11.038, p = 0.0015$; accent: $F(1, 62) = 68.1434, p < 0.001$). The agent with an Iranian accent was viewed to not only have a more Middle-Eastern accent but also appeared more Middle-Eastern. Also, there was a main effect of culture for the appearance question ($F(1, 62) = 4.276, p = 0.0428$) where Americans ranked the agent as more Middle-Eastern looking than did the Iranian-Americans.

The responses to the first question were used as the dependent variable in a 2 X 2 ANOVA, with similar factors as above. There was a significant interaction between the two factors ($F(1,62) = 4.3649, p = 0.0408$). A planned comparison revealed that Iranian-American participants who interacted with the agent with an American accent viewed Anthony leaving before dinner as more appropriate than the Iranian-Americans who interacted with the agent with an Iranian accent (Welch Two Sample one-tailed t-test: $t(10.596) = 10.596, p = 0.0512$)⁶ (Figure 3). However, this difference did not reach significance for Americans ($t(49.48) = 1.4179, p = 0.1625$)

Next, using the responses to the second question as the dependent variable, we ran the same 2 X 2 ANOVA. There was a main effect of culture ($F(1, 62) = 9.5759, p = 0.0029$), where Iranian-Americans ranked this question as lower than Americans. There was an interaction between culture and the accent of the agent ($F(1, 62) = 9.5759, p = 0.0029$), where Iranians who interacted with the American accent agent indicated that Shawn's parents didn't want to give the picture to Anthony compared to those in the Iranian accent condition (Welch Two Sample t-test: $t(9.008) = 3.1186, p = 0.0123$). However this difference did not reach significance for Americans ($t(49.983) = -1.2908, p = 0.2027$).

There were also interactions between culture and agent's accent for Anthony's feeling of happiness ($F(1, 62) = 5.0474, p = 0.0282$) and satisfaction ($F(1, 62) = 12.7468, p < 0.001$). For happiness, Iranian-Americans interacting with the agent with an Iranian accent rated Anthony's happiness higher than those Iranian-Americans who interacted with an American accented agent ($t(11.725) = 2.4057, p = 0.0336$), and vice versa for American participants ($t(49.408) = -2.2325, p = 0.0301$). The same significant trend held for Anthony's satisfaction (Iranian-Americans: $t(11.855) = 3.2666, p = 0.0068$; Americans: $t(48.815) = -3.543, p < 0.001$). Similarly, there was a two-way interaction between culture and accent for Shawn's parents' satisfaction ($F(1, 62) = 7.0429, p = 0.0101$) and the effect also approaches significance for Shawn's satisfaction ($F(1, 62) = 3.0044, p = 0.0880$). Similar to the emotions reported above, participants ranked the satisfaction of Shawn's parents and Shawn higher when the accent of the agent matched

⁶ A power test revealed that if we had the same number of Iranian-American participants as American participants, with probability of 99.15% a two-tailed test with the means of the above sample would have reached significance.

their own accent (Shawn’s parents’: Iranian-Americans: $t(10.121) = 3.6028$, $p = 0.0047$; Americans: $t(49.362) = -1.4053$, $p = 0.1662$, Shawn’s: Iranian-Americans: $t(5.953) = 0.7476$, $p = \text{n.s.}$; Americans: $t(49.054) = -2.2229$, $p = 0.0309$).

4 Discussion

This study provides evidence of how a virtual agent can successfully make a socio-cultural effect on people’s cognition. In a fully factorial design, Iranian-Americans who interacted with a virtual agent that spoke Middle Eastern accented English were more likely to use a congruent cultural frame to interpret a morally charged scenario. The compelling aspect of this effect is that the accented virtual agent’s visual appearance was identical across experimental conditions and the only manipulation was the agent’s accent. Although simple effects were not significant for the Americans, the trends were in the correct direction. The trends could be due to the fact that most of the American participants recruited in this study have a multicultural background (33% African-American, 14% Latino). To our surprise, our manipulation also affected people’s evaluations of the emotions of the characters in the story. We speculate that this effect might be due to the fact that when the Iranian-Americans interacted with the ECA that had an Iranian accent, their overall mood became more positive. Conversely, the ECA with the foreign accent might have had a negative effect on the mood of the Americans. Therefore, participants who interacted with an agent whose accent matched their own, interpreted the characters’ emotions more positively.

We hope to code and analyze participants’ verbal responses to see whether the cognitive cultural frame-shift also affects participants’ recall of different cultural values in the story. We acknowledge the low number of Iranian-Americans participants in our study. However, we would like to note that the probability of replication of a result is dependent on p-levels but not affected by sample size (e.g. [8]).

In summary, contributions of this work are two-fold. First, the study adds to the literature in cross-cultural psychology by showing that just spoken language accent can induce cultural frame shifts. Second, this work makes a methodological contribution to the field of human-computer interaction and experimental psychology. Our results also have implications for teaching cross-cultural fluency and competency.

Acknowledgments. This research was supported by NSF IIS-0916858, a post-doctoral fellowship funding for the second author from the Army Research Lab and U.S. Army Research, Development, and Engineering Command (RDE-COM). The content of this paper does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

MD, PK and LH were involved in designing and analyzing the experiment and also in writing the paper. LO was involved in running the experiment and writing the paper. JG was primarily involved in the writing and analysis stages of this work.

References

1. V. Benet-Martinez, J. Leu, F. Lee, and M. W. Morris. Negotiating biculturalism: Cultural frame switching in biculturals with oppositional versus compatible cultural identities. *Journal of Cross-Cultural Psychology*, 33 (5):492–516, 2002.
2. R.G. D’Andrade. *Cultural meaning systems*, pages 88–119. Cambridge University Press, Cambridge, 1984.
3. J. Haidt, S. Koller, and M. Dias. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65:613–628, 1993.
4. A. Hartholt, J. Gratch, L. Weiss, and The Gunslinger Team. At the virtual frontier: Introducing gunslinger, a multi-character, mixed-reality, story-driven experience. In Z. Ruttkay, M. Kipp, A. Nijholt, and H. Vilhjálmsón, editors, *Intelligent Virtual Agents, Lecture Notes in Computer Science*, pages 500–501. Springer, 2009.
5. J. Henrich, S. J. Heine, and A. Norenzayan. The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3):61–83, 2010.
6. Y. Hong, M. W. Morris, C. Chiu, and V. Benet-Martinez. Multicultural minds. a dynamic constructivist approach to culture and cognition. *American Psychologist*, 55 (7):709–20, 2000.
7. L. Huang, L. P. Morency, and J. Gratch. Virtual rapport 2.0. In *International Conference on Intelligent Virtual Agents*, 2011.
8. P. R. Killeen. An alternative to null-hypothesis significance testing. *Psychological Science*, 16:345–353, 2005.
9. S. Kopp, B. Krenn, S. Marsella, A. N. Marshall, C. Pelachaud, H. Pirker, K. R. Thirsson, and H. Vilhjálmsón. Towards a common framework for multimodal generation: The behavior markup language. In *International Conference on Intelligent Virtual Agents*, pages 21–23, 2006.
10. J. M. Loomis, J. J. Blascovich, and A. C. Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behavior research methods instruments computers a journal of the Psychonomic Society Inc*, 31(4):557–64, 1999.
11. D. Medin, W. Bennis, and M. Chandler. Culture and the Home-Field disadvantage. *Perspectives on Psychological Science*, 5(6):708–713, December 2010.
12. E. Rader, M. Echelbarger, and J. Cassell. Brick by brick. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI ’11*, page 2971, Vancouver, BC, Canada, 2011.
13. R. A. Shweder, N. C. Much, M. Mahapatra, and L. Park. *Morality and health*, chapter The ”Big Three” of morality (autonomy, community, divinity) and the ”Big Three” explanations of suffering., pages 119–169. Routledge, New York, 1997.
14. M. Thiebaut and S. Marsella. Smartbody: Behavior realization for embodied conversational agents. In *In 7th International Conference on Intelligent Virtual Agents (IVA)*, 2007.
15. L. Yin, T. W. Bickmore, and D. E. Cortes. The impact of linguistic and cultural congruity on persuasion by conversational agents. In *Proceedings of IVA, Lecture Notes in Computer Science*, pages 343–349, Philadelphia, PA, 2010. Springer.