Ethnic Identity and Engagement in Embodied Conversational Agents

Francisco Iacobelli and Justine Cassell

Northwestern University
Frances Searle Building 2-343. 2240 Campus Drive. Evanston, 60208-2952
[f-iacobelli, justine]@northwestern.edu

Abstract. In this paper we present the design, development and initial evaluation of a virtual peer that models ethnicity through culturally authentic verbal and non-verbal behaviors. The behaviors chosen for the implementation come from an ethnographic study with African-American and Caucasian children and the evaluation of the virtual peer consists of a study in which children interacted with an African American or a Caucasian virtual peer and then assessed its ethnicity. Results suggest that it may be possible to tip the ethnicity of a embodied conversational agent by changing verbal and non-verbal behaviors instead of surface attributes, and that children engage with those virtual peers in ways that have promise for educational applications .

Keywords: Virtual Peers, Embodied Conversational Agents, Culture, Ethnicity

1 Introduction and Background

Although a number of researchers have addressed the issue of embodied conversational agents that exhibit culturally authentic traits [for example 1, 10, 11], the majority of those studies have modeled culture only through surface traits such as skin color, hair style, clothing, etc. Additionally, the majority of these studies have targeted the issue of whether adults prefer to interact with a same or different racially-identified agent. However, research on Embodied Conversational Agents (ECAs) shows that users prefer ECAs that display verbal behaviors that are consistent with their non verbal behaviors. This suggests that in order to build a ethnically authentic ECA one needs to model verbal and non verbal behaviors that are consistent with one another and with the ethnicity that one wishes to emulate.

A number of researchers have studied user reactions to culturally diverse embodied conversational agents (ECAs). The design of these agents has primarily consisted of external features such as skin color or appearance [for example 1, 9, 11]. Baylor [1], for example, implemented ECAs that varied in skin color and hairstyle to represent Caucasian and African American agents. She found that African-American students may affiliate more strongly with an African-American agent, perceive it as more engaging and as more facilitating of learning; however, no evidence of increased learning outcomes was found. Nass and colleagues [11] evaluated a Korean and a Caucasian agent and found that participants judged the agent of their same ethnicity to

be more trustworthy, attractive and competent. Nass also found that participants conformed more to the decision of agents of the same ethnicity. In Nass' study, only physical appearance was altered to represent ethnicity. Prior research on Embodied Conversational Agents (ECAs) shows, however, that users prefer ECAs that displays verbal behaviors that are consistent with their non verbal behaviors [11]. Along these lines, Maldonado and colleagues [10] implemented three pedagogical agents that represented adolescent girls from Venezuela, Brazil and the United States. Nevertheless no empirical evaluation was carried out with these agents, and it is unclear where the authors derived the traits that they attribute to each ethnicity.

In general, the term "race" is associated with biology while "ethnicity" is associated with culture. Since recent scholarship questioned whether the human race can be divided into biological groups [8], in this paper we concern ourselves with ethnicity, which may be thought of as a set of traits that index to other members of the group, and to members of other groups, one's cultural community membership. Physical appearance is not the most reliable index of ethnicity. On the other hand, we claim that the combination of verbal and non-verbal behaviors can be a reliable index of ethnicity. For example, African Americans may speak a dialect of English known as African American Vernacular English, with its own syntax, morphology and lexicon, and they also may produce non-verbal behaviors that are different from those used by Caucasians in similar situations [6]. Identity is thus constructed on the cultural level and on the micro-interpersonal level and the two are brought together in moment-by-moment behavior in conversation [7]. To address the problem of cultural authenticity and the development of ethnic diversity in ECAs, then, we argue that the design process must be informed by careful research on actual communities of people in particular contexts, and must depend on a model that is derived from the verbal and non verbal behaviors that signal ethnic identity in those contexts.

The study reported here is a part of a larger research program into the development of virtual peers (lifesize child-like ECAs) to scaffold the development of literacy skills in children from different populations. In the study reported here we assess whether children are able to attribute ethnic identity to an ECA based on behavior and not skin color. We do this by having children interact with two virtual peers which have an identical appearance but of which one speaks African American Vernacular English, and uses nonverbal behaviors based on a study of African-American children, and the other speaks and uses non-verbal behavior from Caucasian children who speak Standard American English. We believe that children will be able to assess ethnicity on the basis of these kinds of identity features, and that they will be more engaged with ECAs of their own ethnicity.

4 Methodology

In order to test our hypotheses we implemented two virtual peers by applying two models of behaviors to an existing virtual child with a racially-ambiguous appearance [5]. The Caucasian peer was implemented on the basis of the narrative model described by Wang and Cassell [12]. Several dyads of Caucasian children who spoke Standard American English (SAE) were videotaped playing with a doll house and

figurines. Wang and Cassell coded narrative roles, speech acts, eye gaze patterns and other non-verbal behaviors. In order to construct the African-American model of behavior we observed and coded the verbal and non-verbal behaviors of seven African American children who were monolingual speakers of African American Vernacular English (AAVE), during spontaneous play. We coded narrative roles, speech acts and eye gaze patterns, and identified those verbal and nonverbal behaviors that were maximally distinctive from the Caucasian children. An excerpt from this model is given in Table 1. Under the "*Non-Verbal*" column, asterisks (*) indicate those behaviors maximally distinctive from Caucasian children who speak SAE [6]. For example, the African-American children were more likely to gaze at the toys than one another, which was not the case for Caucasian children who spoke SAE.

The African-American and Caucasian models of behavior were then implemented in virtual peers. The virtual peer was displayed on a back projection screen with a wooden doll house placed against the screen. Overall, the environment looked as if the virtual child and the real child were playing, facing each other (see Figure 1). The doll house contained figurines and play furniture, and children were invited to tell stories with the virtual peer.

Table 1. Excerpt from model of African American children's collaborative storytelling.

Roles	Speech Act	Speaker	Function	Non-Verbal
Critics and Authors	Suggest	Critic	To suggest an event or idea to the story	Eye gaze towards toy (66.3% of the tim)e. Towards author (33.3%)
	Correct	Critic	To correct what has been said	*Eye gaze on toy and remains there (50%). Eye gaze towards author and then turns towards toy (25%).
	Question	Critic	Seek clarification or missing information	Eye gaze towards toy 100% of the time.
	Answer	Both	To clarify or supply missing information	*Eye gaze to doll house or toy (66%). Gaze towards camera (33%)
	Elaborate	Author	Additions to story	Eye gaze on toy or house (79%)
Co-Authors	Role-Play	Both	Play role of characters in story	Eye gaze towards own toy (78%).
	Simultaneous Turns	Both	Compete for turn	Eye gaze towardes own toy (71%).
Nonstory	Direct	Both	Suggest position of toys or behavior	Eye gaze towards other and then quickly turns to house or toys (56%)
	Comment	Both	Comment unrelated to story	Eye gaze to toy or house (80%). Gaze towards other speaker (10%).
	Idiosyncratic	Both	Uutterance unrelated to story	Eye gaze to the toy or house (68%). Gaze to other speaker (10%).

Below we report the results of a pilot study to determine if we were successful in changing the perceived ethnicity of the virtual peer by changing its verbal and non-verbal behaviors, and how African-American children engaged with the two models.

29 third graders (ages 9-10) from a public school in Chicago participated. 100% of the students of the school are African American and 95.3% are low income. All participants are mono-lingual speakers of AAVE. 20 of the participants are taught by an African American speaker of AAVE while 9 belonged to a class with a Caucasian SAE-speaking teacher. 15 participants were girls. We used the scores of the Stanford

Learning First standardized test to make sure that the distribution of children to each condition was diverse with respect to academic performance.

In this study, we examined one independent variable: ethnicity of the virtual peer: one virtual peer was modeled after Caucasian children that speak Standard American English (SAE VP) and one was modeled after African American children that speak African American Vernacular English (AAVE VP).

Instructions to the children were given as follows: First, the experimenter, who is Hispanic, and speaks with a slight Hispanic accent, introduced the virtual peer and invited children to tell stories with it; he then left the children alone. The virtual peer, who was operated by another experimenter in a Wizard of Oz mode (following the rules of either AAVE or SAE model), told one story and then elicited a story from the child. While the child was speaking, the VP gave feedback ("what happened next?", "for real?"). Once the child finished the virtual peer told a second story, and then elicited a second story from the child. The virtual peer then said goodbye and the experimenter entered the room. The participant was then asked to complete a sorting task that demonstrates ability to find similarity along more than one dimension in order to ensure that the child was capable of understanding multiple dimensions of similarity [2], and finally the child was asked which of two family strips (pictures of a family; in this case African-American and Caucasian appearance) represented the family that "came to take Alex, the virtual peer, home from school" [2]. Finally, in order to ensure that the family-matching task was picking up on ethnicity, we asked children which family strip would pick them up from school [2].

Due to the long-running nature of data collection, five experimenters ran the Wizard of Oz on different days. In order to ensure comparability, we looked at the mean number of feedback utterances by the VP, using pairwise t-tests. A significant effect pointed to one experimenter, a non-native speaker of English, in the US for 4 months. The data collected in those sessions was therefore not considered here. Additionally, in one case the child did not complete the tasks assigned and in another the experimenter did not explain a task properly. This left us with 17 participants of which 9 were assigned to the AAVE VP condition and 8 to the SAE VP condition.

5 Results

Results from the ethnic assessment task are presented in Figure 2. A power analysis reveals that significance would require 28 participants in each condition. Our sample size was significantly smaller; therefore, most of the results presented simply represent interesting trends. As seen in Figure 2, 56% of the AAVE VP children perceived the virtual peer as having African American parents versus 12% in the SAE condition. 83% of children in the SAE condition judged the VP as having Caucasian parents. This result did not reach significance ($p=0.15$, Fisher's Exact; hybrid version), but it should be noted that it might have had we included the children who quickly changed their minds from saying that Alex would be picked up by the Caucasian family to saying that Alex would be picked up by the African American family (no child changed his or her mind in the other direction). This suggests that our model is not complete, and that while the SAE VP was accurately judged as ethnically

Caucasian, the AAVE VP was not as convincingly African American. This is not particularly surprising since the current AAVE model contains fewer observations of nonverbal behaviors than the SAE model. Future work includes running a bigger study with a VP that incorporates a more complete model of African American interactions and a wider range of AAVE utterances and stories.

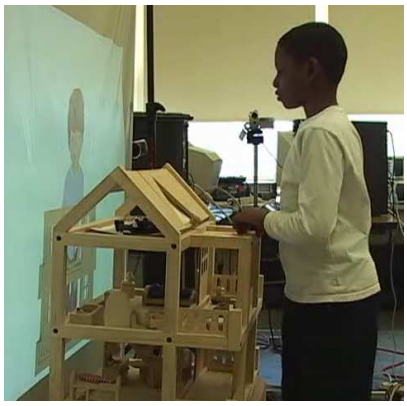

Figure 1. Layout of experiment.

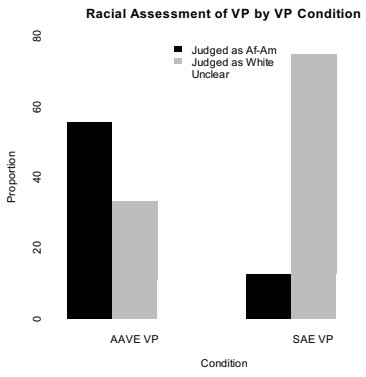

Figure 2. Perception of ethnicity by condition.

In order to test whether children are more engaged when interacting with agents of their own ethnicity, it is valid to use the modeled ethnicity of the VP as the independent variable. However, because we were not able to emulate ethnicity in a convincing way, this independent variable does not fully capture the concept of an agent of the same ethnicity as the child. For this reason, below we present results both as a function of the condition "ethnicity of VP" *and* as a function of the ethnicity that the child attributed to the agent.

Our first analysis looks at the duration of the children's stories in number of seconds; a trend suggests that the *first* stories told with the AAVE VP were longer $t(16)=1.52$, $p<0.08$). The duration of the second story, however, was not significantly different. In addition, a re-analysis of these results in terms of *perceived* ethnicity is also not significant. However, in children's play, engagement is signaled by duration of play, complexity of stories, and also the amount of thematic overlap between the two children. That is, when children are very engaged, they pick up on one another's language [3]. In order to test for this, we coded instances in which the child mimics Alex's lexical or thematic choices [4]. Looking at the proportion of total utterances that mimic Alex, we found a trend that suggests that children in the SAE VP condition produce a higher proportion of mimicking utterances in Story1. $t(15)=-1.5$,$p<0.07$ (see Figure 3). This trend remains, albeit to a lesser degree, in the second story. Re-analyzing these results as a function of *perceived* ethnicity demonstrates that children do mimic the virtual peer significantly more when they perceive it as Caucasian. $t(14)=-2.51$; $p<0.02$ (see Figure 4).

In order to assess the effect of ethnicity of the virtual peer on the children's language, we also coded instances where the children produced SAE linguistic features in their talk. We used the two most common AAVE-SAE contrasts for third

grade children: (a) zero copula (e.g. she drinking instead of she is drinking) and (b) subject-verb agreement (e.g. he sleep every night instead of he sleeps every night) [6]. We found that children who interacted with an SAE-speaking peer produced a mean of .04 SAE features as opposed to .02 for children who interacted with the AAVE-speaking peer. This difference was not, however, significant. Re-analysis of these results in terms of the perceived ethnicity of the virtual peer was still not significant. However, in the second story, we observed that children who perceived the VP as African American produced an average of 0.015 SAE utterances, whereas children who perceived the VP as Caucasian produced a mean of 0.042 SAE utterances.

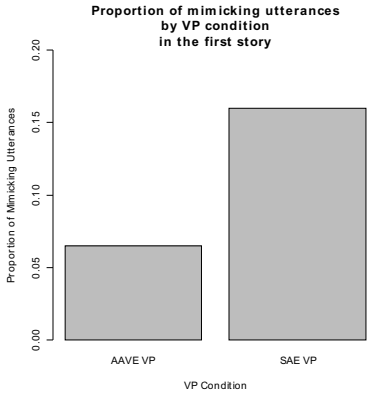

Figure 3. Proportion of child's utterances in Story1 that mimicked Virtual Peer, by VP condition ($p<0.07$)

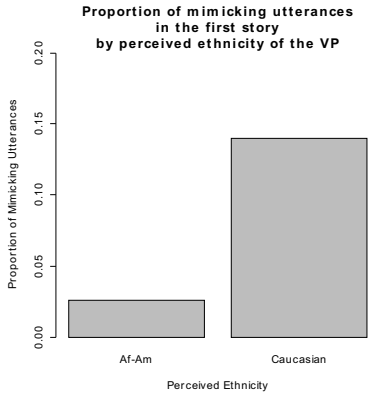

Figure 4. Proportion of child's utterances in Story1 that mimicked Virtual Peer, by perceived ethnicity of the VP ($p<0.02$)

6 Conclusions

In this study we set out to assess whether we could tip the perceived ethnicity of an embodied conversational agent, specifically a virtual peer (VP), by changing the verbal and non verbal behaviors and leaving the appearance constant. The ethnicities we chose to model were an African American speaker of African American Vernacular English (AAVE) and a Caucasian American speaker of Standard American English (SAE). Our results were very partial, but do illustrate differences in kinds of engagement between versions of the VP. This exploratory study is a part of an iterative design process, and the next step is to go back to the original data of child-child African-American interaction in order to better understand the nature of linguistic and nonverbal features that signify African-American identity.

As well as investigating how children assessed their ethnicity, we also examined the effects of interaction with VPs on language use. We found that the African American children tend to spend more time telling stories with the African American VP, although since this is only true for Story1, this effect may wear off. However, African American children mimicked the Caucasian peer more often. Our ultimate goal is to implement an African-American virtual peer capable of sustaining

engagement with a child such that the VP can model and scaffold the use of code-switching between AAVE and SAE, to increase school-based literacy. The current results suggest that an AAVE-speaking VP will indeed be able to sustain engagement, and that an SAE VP will encourage children to mimic its language and trigger code-switching behaviors. Both of these results are extremely promising for future research.

Acknowledgments. We thank Andrea Tartaro, Joris Janssen, Nathan Cantelmo, Deb Zutty and the other members of the Articulab for their help. This study was made possible by grants from the National Science Foundation.

References

[1] Baylor, A. and Kim, Y., (2003) "The Role of Gender and Ethnicity in Pedagogical Agent Perception," in *Proceedings of ELearn (World Conference on ELearning in Corporate, Government, Healthcare, & Higher Education*, Phoenix, AZ.

[2] Bigler, R. S. and Liben, L. S., (1993) "A Cognitive-Developmental Approach to Racial Stereotyping and Reconstructive Memory in Euro-American Children," *Child development,* vol. 64, p. 1507.

[3] Cassell, J., (2004) "Towards a Model of Technology and Literacy Development: Story Listening Systems," *Journal of Applied Developmental Psychology,* vol. 25, pp. 75-105.

[4] Cassell, J. and Ryokai, K., (2001) "Making Space for Voice: Te chnologies to Support Children's Fantasy and Storytelling.," *Personal Technologies,* vol. 5, pp. 203-224.

[5] Cassell, J., Tartaro, A., Oza, V., Rankin, Y., and Tse, C., (2007) "Virtual Peers for Literacy Learning," *Educational Technology, Special Issue on Pedagogical Agents,* vol. XLVII, pp. 39-43.

[6] Craig, H. K. and Washington, J. A., (2005) *Malik goes to school: Examining the language skills of African American students from preschool-5th grade*. New York: Lawrence Erlbaum.

[7] Eckert, P., (2005) "Variations, conventions and social meanings," in *Proceedings of Linguistic Society of America*, Oakland, CA.

[8] Gould, S. J., (1981) *The mismeasure of man*, 1st ed. New York: Norton.

[9] Hayes-Roth, B., Maldonado, H., and Moraes, m., (2002) "Designing for diversi ty: Multi-cultural characters for a multi-cultural world," in *Proceedings of IMAGINA 2002* Monte Carlo, Monaco

[10] Maldonado, H. and Hayes-Roth, B., (2004) "Toward Cross-Cultural Believability in Character Design," in *Agent Culture: Human-Agent Interaction in a Multicultural World*, S. Payr and R. Trappl, Eds. New Jersey: Lawrence Erlbaum Associates, pp. 143 - 176.

[11] Nass, C., Isbister, K., and Lee, E. -J., (2000) "Truth is Beauty: Researching Embodied Conversational Agents," in *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, Eds. Cambridge, MA: MIT Press, pp. 374 -402.

[12] Wang, A. and Cassell, J., (2003) "Co-authoring, Corroborating, Criticizing:Collaborative Storytelling for Literacy Learning," in *Vienna Workshop '03: Educational Agents - More than Virtual Tutors*, S. Payr and R. Trappl, Eds. Vienna, Austria