Dialect Study Applications and Kinect Motion Capture

Shannon Gray, Samantha Finkelstein and Justine Cassell
Human-Computer Interaction Institute
Carnegie Mellon University

ABSTRACT
This paper discuses my experience with the Distributed Research Experiences for Undergraduates (DREU) program this summer. I spent this past summer working in the ArticuLab at Carnegie Mellon University, under Professor Justine Cassell. I ended up working on two different projects this summer. For the first seven weeks of the summer, I worked on creating applications that were used in several user-studies the ArticuLab was conducting. Similar to my DREU experience last year, I was also able to participate in several pilot studies to get user feedback on these applications.

The second project I worked on was implementing a Motion Capture system using a Microsoft Kinect and the Flexible Action and Articulated Skeleton Toolkit (FAAST) that allows users to create Motion Capture animations for the ArticuLab's Virtual Peer Alex. I also designed programs to walk users through how to create these animations and then automatically load them into the Alex software.

Keywords
Computer science education, programming, user study, dialects, Microsoft Kinect.

1. INTRODUCTION
The ArticuLab is part of Carnegie Mellon University's Human Computer Interaction Institute in the School of Computer Science. The ArticuLab's mission is to study human interaction in social and cultural contexts as the input into computational systems that in turn to help better understand human interaction, and to improve and support human capabilities in areas that really matter. The ArticuLab studies how people communicate with and through technology. Some of this research includes the interaction between humans and virtual peers, called Embodied Conversational Agents (ECA), and how ethnicity mediates technology use. The ArticuLab studies how technology can be used for positive educational and developmental initiatives, such as improving literacy skills for children who do not grow up speaking Standard American English (SAE).

Some of the ArtcicuLab's results show that African American children demonstrate awareness of how to use different language styles in different situations (such as using more African American Vernacular English (AAVE) during collaborative play than they use when practicing a formal presentation by role-playing as a teacher and student), and that they make this language transition whether they are collaborating with a human peer or a virtual peer partner [Cassell et al., 2009]. This work also demonstrated that

children are more fluent when speaking socially with a virtual partner which first introduces itself in the vernacular dialect than one which introduces itself with a standard English dialect. Regardless of partner, children also demonstrate increased fluency when they are speaking in the vernacular dialect rather than the standard dialect themselves [Finkelstein et al., 2012]. These results call for a re-examination of the cultural assumptions followed in the design of educational technologies, with a specific emphasis on the way in which we index culture and identity, and the ways in which we ask culturally-underrepresented groups to participate in learning activities.

Figure 1. "Chooser" Application that would load the other four applications at the discretion of the experimenter. When one of the loaded apps finished, the program would return to the "Chooser" screen so the experimenter could load another if she chose. The box in the lower right corner was for the password input.

2. STUDY APPLICATIONS

I created four different applications that were used in the study, and then I also added a login screen, and a "chooser" program that would open whichever program you selected [Figure 1]. All the screens in the programs were password protected, so that the participants couldn't move past certain designated parts of the programs without the experimenter putting in the password. This was introduced early on in the pilot testing. We found out quite quickly that children love to click buttons, and would often click past where they were supposed to be.

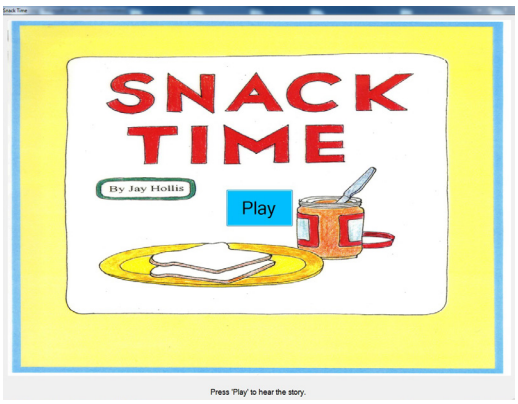

Figure 2. Start of the Story Time application.

2.1 Story Time

The first application I made was called the Story Time application. The purpose of the Story Time app was to determine what dialect the children naturally had. It was originally a story-book kind of interface, I believe. Participants were shown images depicting two children making snacks, and an experimenter would read out the lines of the story, describing the image on that page. An example page is shown in *Figure 2*. The children would be asked to press play, and then an audio clip would play, reading the

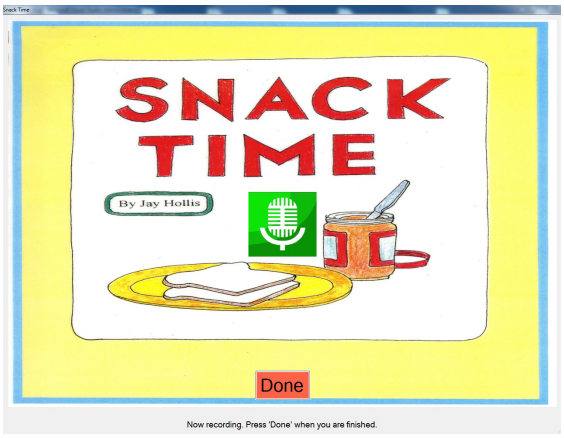

Figure 3. Recording screen in the Story Time app.

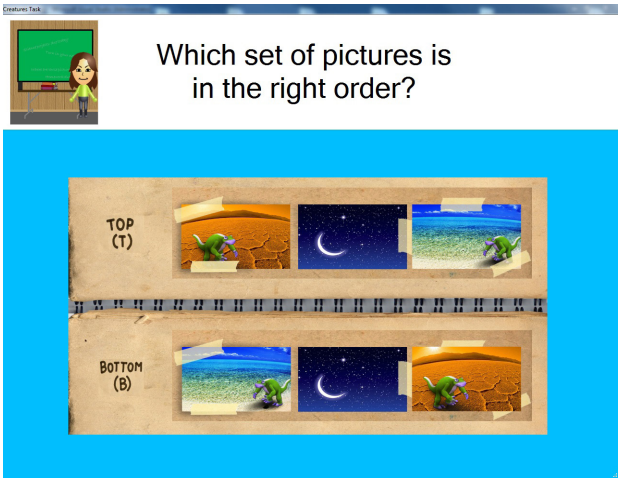

Figure 4. The Creatures application.

line of the story. The participants would then be shown a screen that indicated the program was now recording, and were supposed to repeat back what they had just heard [Figure 3]. They were instructed to hit the "Done" button when they had finished recording. The rest of the story continued on in the same manner. I was informed by the members of the ArticuLab that this was the first time the Story Time process had ever been automated in this way, and that this tool will be very helpful to them in the future as well.

2.2 Creatures

The next application I created was called the Creatures application. The purpose of this application was to see if participants could better understand the use of past or present tense verb usage in their own dialect, versus a dialect that is different from their own. This application also studied whether a perceived "mismatch" between appearance and dialect had an effect on performance as well. Participants were introduced to four different avatars, in sudo-random order. Two were designed to appear Caucasian and two were designed to appear African American. We had four different speakers record the lines of dialogue for these "avatars"; two of which spoke Standard American English, and two that spoke African American Vernacular English. These dialects were paired sudo-randomly with each avatar. One Caucasian avatar was matched with an SAE speaker and one with an AAVE, and the same for the African American avatars.

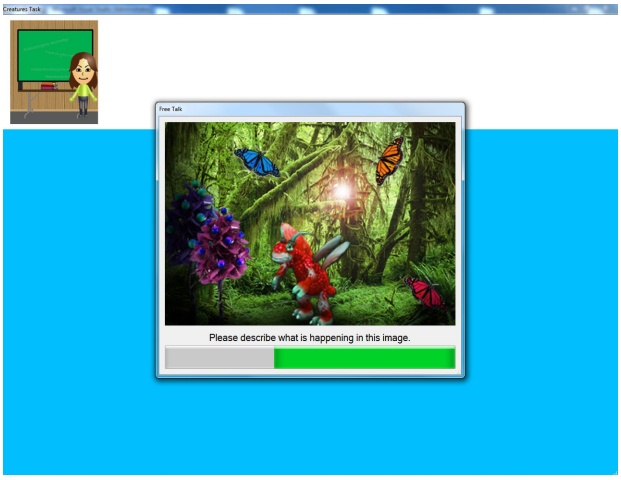

Figure 5. Free Talk section.

These avatars introduced themselves to the participants and introduced them to the task they were going to perform. They would then hear a sentence about a creature doing something. This sentence would contain either a past or present tense verb. Then they were shown two images; one indicating past tense and one present. They would then choose which image matched the tense they had previously heard [Figure 4]. For each avatar there were ten trials, so the participants did forty total trials. In between each avatar there was a Free Talk section where the participants were shown an image and asked to describe what was happening in the image [Figure 5]. This section was also recorded.

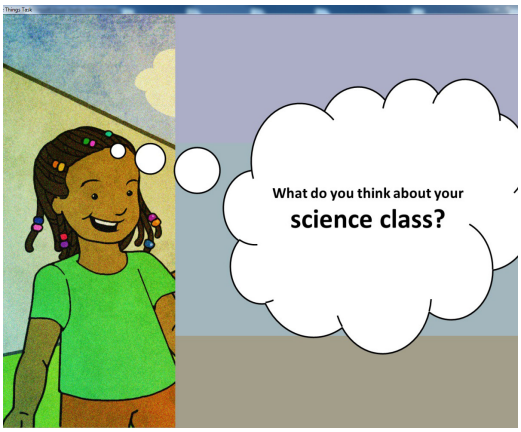

Figure 6. Example screen of the Favorite Things application, showing one of the peer avatars.

2.3 Favorite Things

This application was designed to test whether a participant would modify his or her dialect according to whom they were speaking. This was tested in three different contexts: perceived race/dialect, age/authority and specific school subjects. This task had four avatars as well; two were Caucasian and two African American, with SAE and AAVE dialects respectively. The avatars were also introduced as either peers or teachers. Each of the avatars introduced themselves to the participants and then asked several questions for the participants to answer, about one of four classes: math, science, social studies and English (Language Arts) [Figures 6 & 7]. These questions included things like: "What is your class like?", "What do you like or dislike about your teacher?", "What are your favorite and least favorite parts of this class?", etc. The participants then had a few minutes to talk to the avatar, and their responses were recorded. The participants repeated this process for all four avatars. The order of the avatars and the subjects they were paired with were sudo-random as well.

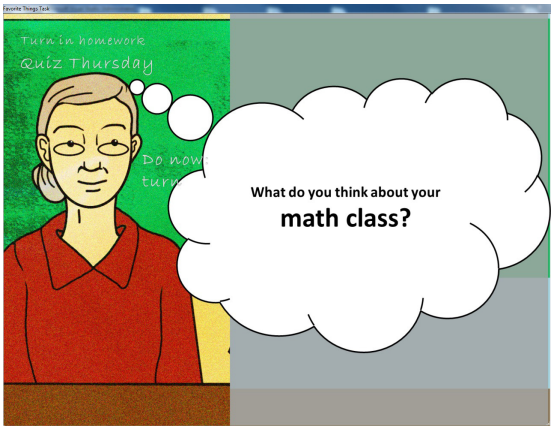

Figure 7. A teacher avatar from the Favorite Things app.

2.4 Sticker Task

The last application was designed to measure whether participants attributed certain dialects to certain scenarios. The participants were given a screen that showed two images, with a slider in the middle. The images showed a child at a park with other children, or in a classroom with other children [Figure 8]. The participants would hear an audio clip of a child speaking, and were instructed to place a "sticker" of the child on the slider, near the image they thought best matched the child's location. If they were not sure, they were instructed to place the sticker somewhere in the middle. There were three introductory trials and then ten normal trials, all with varying degrees of SAE and AAVE dialects. The ten normal trials were randomized in order.

Figure 8. Sticker application.

3. MOTION-CAPTURE PROCESS

Motion capture is the process of recording the movement of objects or people. In this case, the ArticuLab wanted to use motion capture to create new animations for their virtual peer Alex. Alex is the Articulab's embodied conversational agent, who was designed to strategically employ either an AAVE or SAE dialect during interactions with students. Hand-animating gestures is a very time-consuming process, and being able to create new animation using actual body movements could significantly decrease the time requirements of this process. Unfortunately, I do not have any images of the motion capture process or the tutorial applications.

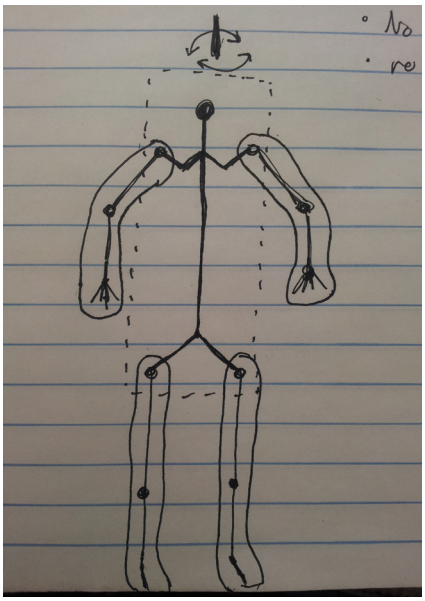

Figure 9. Motion capture capabilities model.

3.1 FAAST and the Kinect

I went about the motion capture process by using a Microsoft Kinect, and the Flexible Action and Articulated Skeleton Toolkit (FAAST). FAAST is middleware used to facilitate integration of full-body control with games and virtual reality applications. The newer version of FAAST also allows you to record body gestures using the Kinect, with some restrictions. Generally, FAAST can code full body movement in the X and Z directions, but not Y, meaning it cannot code a jump, for instance. It can, however, record limb movement in all directions relatively reliably.

Figure 9 breaks the skeleton into independent moveable blocks. For instance, the spine, hips, shoulders and head are all one moveable piece, and thus marked in the dotted box. The limbs however, have free movement, so they are in their own bubbles. The round dots on the skeleton represent joints (besides the "head" dot I drew), and indicate the ability for movement as well. The creators of FAAST were nice enough to give us access to this version of the program, which is still being developed. With FAAST, you can access the data from the Kinect, and create SKM animations from real body movements.

3.2 Tutorial Applications

Once I got the motion capture working, I created a tutorial application that would run a user through the entire process of creating an animation, editing it in Autodesk Maya 3D, and adding it to the Alex software. It is designed to be run along with the other applications a user will need to complete this process, like FAAST and Maya. It provides users with in-depth instructions and step-by-step images for each process involved. These tutorials were tested by other lab members during several "Talk Alouds". Basically a user would run though the tutorial application and narrate any thoughts, actions, confusions and general ideas about the process. I could then make changes based on the input I received. This was also a very helpful and useful process for me.

4. CONCLUSION

This was a very productive summer for me. I not only got to work with professionals in the field, but also got a first-hand account of what the research process is really like. I feel like I got a real crash course of what I could expect from Graduate school, and that was a great experience. I am pleased with the work I was able to get done on the study applications and the motion capture process.

It was absolutely wonderful to be able to take part in the studies this summer. It was very useful to be able to see the reactions that kids had to the programs during the pilot studies, and be able to make changes based on these reactions. The experience I was able to have this summer was extremely valuable. I really enjoyed the coding work I was able to do, and the other members of the lab were joy to work with. It gave me a feel for a real workplace atmosphere. I also go to try my hand at creating my first self-contained, programs with interfaces. That was really great to be able to do. I will be forever grateful that I was able to have such a wonderful summer.

5. ACKNOWLEDGEMENTS

I would like to sincerely thank the DREU program, Carnegie Mellon University, Justine Cassel, and Samantha Finkelstein, and the ArticuLab for providing me with the wonderful opportunity I had this summer.

All of this information can be found in more detail at shannongraydreu.webs.com, which documents my DREU experience at Carnegie Mellon.

6. REFERENCES

[1] "The ArticuLab." ArticuLab. N.p., n.d. Web. 30 May 2013. <http://www.articulab.justinecassell.com/index.html>.

[2] Cassell, Justine, Kathleen Geraghty, Berto Gonzalez, and John Borland. "Modeling Culturally Authentic Style Shifting with Virtual Peers." ICMI-MLMI '09 Proceedings of the 2009 International Conference on Multimodal Interfaces (2009): 135-42. Web.

[3] Finkelstein, S., Scherer, S., Ogan, A., Morency, L.P., & Cassell, J. "Investigating the Influence of Virtual Peers as Dialect Models on Students' Prosodic Inventory." in Proceedings of WOCCI (Workshop on Child-Computer Interfaces) at INTERSPEECH (2012):1-8. Web.