A Virtual Space for Children to Meet and Practice Chinese

Mei Si

Rensselaer Polytechnic Institute
sim@rpi.edu

Abstract. Second language acquisition after the students have learned their first language is a unique process. One key difference between learning a foreign language and one's mother tongue is that second language learning is often heavily facilitated with digital media, and in particular, through interacting with computers. This project is aimed at leveraging computer game technologies and Microsoft Kinect camera to create engaging and affordable virtual environments for children to "virtually" meet and practice their language and culture skills. We present a uniquely immersive and narrative-based environment for children to meet online and practice Mandarin Chinese with each other, providing a platform that is at once affordable, engaging to students and attractive to teachers.

Keywords: Language and Culture Learning; Virtual Environment; Kinect

1 Introduction

Language acquisition plays an important role in children's cognitive and social development processes. Being able to understand another language and culture is not only fun and useful, but also may facilitate the children's cognitive development.

Second language education is traditionally accomplished through classroom instruction and human tutors. While this is usually an effective approach, it has several drawbacks which limit its application, particularly with a younger audience. Learning a language requires tremendous amount of practice and repetition. Therefore, students are strongly encouraged to practice outside normal classroom hours. However, being able to find appropriate study partners, coordinate a convenient time and location is often difficult, and this is especially true for children. Even when the students can find someone to practice with, due to the students' insufficient language abilities, the practice session can be very frustrating for both the student and the partner, who may be a native speaker. This often causes the students to abort the practice. The students also may not be able to retain the lessons during these practices, because of information overflow.

Computer and network based learning systems on the other hand can keep records of the practice history, and allow people to practice with partners online. Such sys-

tems therefore can potentially make the practice process more accessible and effective.

In this project, we are interested in investigating how to create effective second language practice environments. We have created multiple virtual environments for children to "virtually" meet and practice their language and culture skills through leveraging computer game technologies and Microsoft Kinect camera. Among all the foreign languages for children in the US, Mandarin Chinese is by far receiving the most growth of interest over the past years, and therefore is chosen as the first language to be practiced in our games. We present the details of this system in this paper.

Furthermore, we are interested in studying the impact of various factors on students' language practices, and in particular the use of narrative and body movements. Narrative, of course is an integral part of people's lives. It is an important way for people to entertain, as well as to learn about a new society or culture. Its engaging power has been well observed in various media forms, e.g. novels, movies, and dramas. Simulating real-world or fictional scenarios offers a way to practice language in context, and provides the users with motivations and focus.

Moreover, we hypothesize that body movements and gestures can contribute to language learning. Movement, in essence, is a form of thinking. The theories of embodied cognition argue that our body, mind and the environment are tightly integrated, and our decision-making processes, perception and even memory are deeply rooted in our body and bodily movements (Clark, 2008). Gesturing is a perfect example. Expressive gestures are an important aspect of language use and communication. On the other hand, spontaneous gesturing, which do not directly relate to language use, has been shown for facilitating learning and recall of abstract concepts (Goldin-Meadow, 2003). This is because memory can be off loaded to body-environment relationships that are "artificially" created by us. In the future work section, we lay out the experimental studies we plan to conduct for evaluating the effects of these two factors.

2 Related Work

With the rapid development of computer and game console hardware, graphics, artificial intelligence and network technologies in recent years, computer aided pedagogical systems and intelligent pedagogical agents have been widely used for tutoring and training purposes, ranging from math (Beal, Walles, Arroyo and Woolf, 2007) and physics tutoring (Ventura, Franchescetti, Pennumatsa, Graesser, Jackson, Hu, Cai and the Tutoring Research Group, 2004) to language and social skill training (Johnson, Marsella, Mote, Si, Vilhjalmsson and Wu, 2004; Traum, Swartout, Marsella, Gratch, 2005), and from life style suggestions (Zhang, Banerjee and Luciano, 2010) to PTSD (Rizzo, Newman, Parsons, Reger, Difede, Rothbaum, Mclay, Holloway, Graap, Newman, Spitalnick, Bordnick, Johnston and Gahm, 2009) and Autism interventions (Boujarwah, Riedl, Abowd and Arriaga, 2011).

Similarly, the effective use of language training has been demonstrated in immersive virtual environments such as the Tactical Language Training System (Johnson,

Marsella, Mote, Si, Vilhjalmsson and Wu, 2004), and Rosetta Stone. O'Brien, Levy, and Orich describe a CAVE-based language learning environment targeted at more general L2 applications, in which students explore a virtual model of Vienna in search of the mayor's missing daughter (O'Brien, Levy, and Orich, 2009). Chang, Lee and Si have investigated using immersive narrative and mixed reality for teaching Mandarin Chinese to college students (Chang, Lee and Si, 2012).

Language and culture training involves more than teaching the students how to speak. In real life, people use non-verbal behaviors -- gazes, gestures, and body movements -- to accompany their speech. Not all of the non-verbal behaviors are straightforward to mimic for foreigners. Most existing virtual training environments require the learner to sit in front of a computer and use keyboard and mouse to interact. The learner therefore cannot practice their non-verbal behaviors and conversational skills at the same time.

Gesture based natural user interfaces has been explored in cultural training (Rehm, Leichtenstern, Plomer, and Wiedemann, 2010; Kistler, Endrass, Damian, Dang and André, 2012). In this project we combine gesture based user interface with narratives and puzzles to provide the users with a platform to practice their verbal and non-verbal skills together. More specifically, we created three types of Kinect enabled virtual environments. All of these virtual environments allow multiple users to log in from different locations. The users can control a character's body movements using a Kinect camera and simultaneously have voice chat with other users. In fact, they have to discuss and collaborate with each other to solve the problems presented in the virtual environments. We hope these virtual environments can thus engage the users and help the users to practice in a natural way.

3 Project Description

This project is aimed at leveraging computer game technologies and Microsoft Kinect camera to create engaging and affordable virtual environments for children to "virtually" meet and practice their language and culture skills in Mandarin Chinese. Our goal is to create affordable, engaging and realistic learning environments for children to meet and practice Chinese. This project is not aimed at replacing language classes or human tutors, but is meant to supplement classroom instruction.

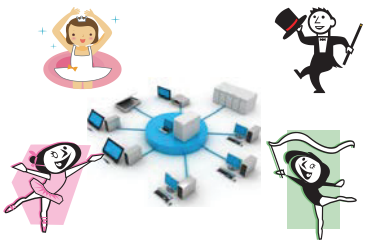

Fig. 1. System Architecture

3.1 Overview of the System

Figure 1 illustrates the overall architecture of the system. This system contains a central server and multiple clients. The server handles most of the computation, and therefore low-end machines can be used as clients. Each client needs to be equipped with a Kinect camera, microphone and speaker. The Kinect camera enables the system to map the user's body movements to a character's body movements, and thus allow the user to directly control the character's motions using his/her body. We choose to use the Kinect camera in this project, because it provides this important functionality, and is affordable, non-invasive and easy to install, which are important features when considering deploying the project outside of the lab.

3.2 Development Environment

The Unity game engine was used for developing this project. Unity supports the development of multi-player network games, and can easily produce the final executable for different platforms. Unity is also easily compatible with the OPENNI package for driving the Kinect camera and Teamspeak 3 package for providing real time multi-user online chat, which will be described in more details below. As a result, we expect minimal effort for deploying the final project at interested schools.

We want to seamlessly integrate the voice chat function into the rest of the application, so that the users do not need to perform any special operations to chat with each other. Considering the age group of our users, it is very important to make all the interactions with the system feel natural. Moreover, we want to be able to replay the whole game session for research purposes, and for the children's teachers and parents. In order to add a voice chat component to our system, we used Teamspeak 3, a popular voice chat service for gaming and other consumer uses that provides both off-the-shelf and SDK tools, to create a voice chat server that we can host in-house and a client that is integrated into the client side of the system. This means that upon entering the environment, the user is immediately connected to the voice chat server with an open microphone so that all parties can begin talking to each other right away. It also means that all the voice chat messages come through our server so we can keep record of them.

3.3 Environments and Characters

Three types of virtual environments were created, which are described below in the order of the amount of background stories involved.

Cao Chung Weighing an Elephant

We have created a 2.5D virtual environment (Figure2) which is based on a traditional Chinese children's story – Cao Chong Weighing an Elephant:

> *This happened about one thousand and seven hundred years ago. One day somebody sent Cao Cao, the king of WEI, an elephant. Cao Cao wanted to know its weight. "Who can think of a way to weigh it?" He asked. But*

nobody knew what to do, because there was nothing big enough to weigh it. Then Cao Chung, one of the king's young sons, came up and said, "Father, I've got an idea. Let me have a big boat and a lot of heavy stones, and I'll be able to find out the weight of the elephant." Cao Cao was surprised, but he told his men to do as the boy asked.

When the boat was ready, the boy told a man to lead the elephant down into it. The elephant was very heavy, and the water came up very high along the boat's sides. Cao Chung made a mark along the water line. After that the man drove the elephant onto the bank. Cao Chong then told the men to put heavy stones into the boat until the water again came up to the line. Cao Chung then told the men to take the stones off the boat and weigh them one by one. He wrote down the weight of each stone and then added up all the weights. In this way he got the weight of the elephant.

Fig. 2. Cao Chung Weighing an Elephant

In this virtual world, the users are asked to play the kids in the story and to find the right way to weigh an elephant without hurting it. The users are provided with multiple tools, such as a knife, which is not big enough to chop the elephant into pieces but can wound the elephant, a scale that is not large enough to weigh the whole elephant and multiple stones as in the original story. When each user enters the game, they are provided with different sets of information regarding where the tools are. The users can find and try out the tools. They are encouraged to discuss how to use the tools and how to solve the problem with each other.

The characters were modeled in 2D with movable body parts. Using a Kinect camera, the user can control the characters' movements through their own body movements. The characters mimic the user's actions, e.g. the user can move around, wave his/her hand, and bend down to pick up or drop an object.

The Elephant and the Blind Men

The second practice scenario is created based on another traditional Chinese children's story – The Elephant and the Blind Men:

Once upon a time, an elephant came to a small town. People had read and heard of elephants but no one in the town had ever seen one. Thus, a huge crowd gathered around the elephant, and it was an occasion for great fun, especially for the children. Five blind men also lived in that town, and consequently, they also heard about the elephant. They had never seen an elephant before, and were eager to find out about elephant. Then, someone suggested that they could go and feel the elephant with their hands. They could then get an idea of what an elephant looked like. The five blind men went to the center of the town where all the people made room for them to touch the elephant. Later on, they sat down and began to discuss their experiences. One blind man, who had touched the trunk of the elephant, said that the elephant must be like a thick tree branch. Another who touched the tail said the elephant probably looked like a snake or rope. ... Finally, they decided to go to the wise man of the village and ask him who was correct. The wise man said, "Each one of you is correct; and each one of you is wrong. Because each one of you had only touched a part of the elephant's body. Thus you only have a partial view of the animal. If you put your partial views together, you will get an idea of what an elephant looks like."

Just like in the original story, in this practice scenario, each user can only see a portion of a large object, and they have to discuss with each other to figure out what the object is. In addition to the elephant in the original story, we are also showing other 2D and 3D objects with different levels of difficulties for this practice. Figure 3 shows an example. Using the Kinect camera, the users can use their hands to move their camera view of the object a little bit to see more of the object. The user will never be able to see the whole object. We designed this function to allow us to later evaluate whether encouraging body movements will engage the users more in a learning environment like this.

Fig. 3. The Elephant and the Blind Men

Chat Rooms

Finally, we have created two 3D environments with 3D characters in them, as shown in Figures 4 and 5 below. One is a student café with realistic models of human characters. The other is a living room with children characters modeled in cartoon style. They were designed to attract users of different ages. In both environments, the users can control a character, "walk" around in the environment, talk with other characters, and interact with objects in the room. For example, in the café, the user can pick up a cup and hand it to a virtual characters or to another user. The user can also collaborate with another user and push the tables around.

Fig. 4. School Café

Fig. 5. Living Room

4 Current Status and Planned Evaluation

We have finished implementing the system and are currently conducting informal usability testing. We are planning on two formal evaluations in the future.

First, the goal of this project is to supplement classroom instruction and exercise, raise and sustain children's interest in learning Chinese language and culture, and help them practice their language skills outside of classroom. The overall effectiveness of the project will be evaluated by comparing the learning outcomes from children who regularly use this project and those who do not use this project. We will be using the school's standard assessment for evaluating the students' performances.

Secondly, we want to evaluate the effects of using narrative and body movements in language learning. For this purpose, we have also created a mouse and keyboard version of the system. The same menu interface is used. However, the user control his/her character's body movements using mouse and keyboard instead of using their

own body movements. The three types of virtual environments we developed for this project involve different amount of narrative components. We hypothesize that the environments with more narrative components can engage the users more, and using body movements with a Kinect camera is a more engaging/effective learning approach compared to using mouse and keyboard.

We will use a 3*2 between-group design with the amount of narrative involved and the type of user interface as two independent variables. Each subject will attend a 3-session study. They will spend one hour in the lab interacting with our system on three consequent days. The course of their interactions will be recorded. On the last day, their learning results will be evaluated. The specific measurements for evaluation will be determined together with our teaching consultants when we develop our learning materials. Based on the results from this experiment, we will conduct a second study exploring ways to encourage the desirable behavior patterns in the students.

5 Conclusion

Learning a language requires a tremendous amount of practice both inside and outside of the classroom. One common problem faced by language learners is where/how to find people to practice with and what to talk about with strangers. This is especially true for children because they have to rely on their parents or other adults for transportation.

In this project we propose to attack these problems by creating virtual spaces with narratives and puzzles embedded in them. We created three types of Kinect enabled virtual environments. The success of this project will make finding and meeting practice partners a lot easier, in addition to the numerous benefits a computer based pedagogical system can provide, such as automatically keeping a record of one's practice history.

The rapid development of computer technologies in recent years enabled a variety of user interfaces to be continently accessible in people's everyday life, ranging from the traditional mouse and keyboards interface to touch screens and touch free camera based technologies for interactions. This work provide a platform for studying how the form of interaction affects children and young adults' language learning processes, and how the design of the learning systems can leverage this effect and make the students' learning process more effective.

Reference

1. Beal, C. R., Walles, R., Arroyo, I., & Woolf, B. P.: On-line tutoring for math achievement testing: A controlled evaluation. *Journal of Interactive Online Learning*, 6 (1), 43-55, 2007.
2. Boujarwah, F., Riedl, M. O., Abowd, G. and Arriaga, R.: REACT: Intelligent Authoring of Social Skills Instructional Modules for Adolescents with High-Functioning Autism. *ACM SIGACCESS Newsletter*, vol. 99, 2011.

3. Chang, B., Sheldon, L. and Si, M.: Foreign language learning in immersive virtual environments. In Proceedings of *IS&T/SPIE Electronic Imaging*, Burlingame, CA, 2012.
4. Goldin-Meadow, S.: Thought before language: Do we think ergative? In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought*, pp. 493-522. Cambridge, MA: MIT Press, 2003.
5. Johnson, W. L., Marsella, S.C., Mote, N., Si, M., Vilhjalmsson, H., and Wu, S.: Balanced Perception and Action in the Tactical Language Training System. In proceedings of *Balanced Perception and Action in ECAs in conjunction with AAMAS*, July 19-20, New York, 2004.
6. Kistler, F., Endrass, B., Damian, I., Dang, C.T. and André, E.: Natural interaction with culturally adaptive virtual characters. *Journal on Multimodal User Interfaces,* 6 (1-2), pp. 39-47, 2012.
7. Miller, L.C., Appleby, R. P., Christensen, J.L., Godoy, C., Corsbie-Massay, C., Read, S. J., Marsella, S., and Si, M.: Virtual agents and virtual sexual decision-making: Interventions for on-line applications that change real-life risky sexual choices. In S. Noar & N. Harrington (Eds.), *Interactive Health Communication Technologies: Promising Strategies for Health Behavior Change*. Mahwah NJ: Lawrence Earlbaum Associates, 2011.
8. O'Brien, M., Levy R. & Orich, A.: Virtual Immersion: The Role of CAVE and PC Technology, *CALICO Journal,* 26 (2), 2009.
9. Rehm, M., Leichtenstern, K., Plomer, J. and Wiedemann, C.: Gesture activated mobile edutainment (GAME): intercultural training of nonverbal behavior with mobile phones. In proceeding of *the 9th International Conference on Mobile and Ubiquitous Multimedia*, 2010.
10. Rizzo, A., Newman, B., Parsons, T., Reger, G., Difede, J., Rothbaum, B.O., Mclay, R.N., Holloway, K., Graap, K., Newman, B., Spitalnick, J., Bordnick, P., Johnston, S. and Gahm G.: Development and Clinical Results from the Virtual Iraq Exposure Therapy Application for PTSD. In proceedings of *IEEE Explore: Virtual Rehabilitation*, Haifa, Israel, 2009.
11. Traum, D., Swartout, W., Marsella, S., and Gratch, J.: Fight, Flight, or Negotiate: Believable Strategies for Conversing under Crisis, In proceedings of *the 5th International Conference on Interactive Virtual Agents*, Kos, Greece, 2005.
12. Ventura, M.J., Franchescetti, D.R., Pennumatsa, P., Graesser, A.C., Jackson, G.T., Hu, X., Cai, Z., and the Tutoring Research Group.: Combining Computational Models of Short Essay Grading for Conceptual Physics Problems. In J.C. Lester, R.M. Vicari, & F. Paraguacu (Eds.), *Intelligent Tutoring Systems 2004*, pp. 423-431. Berlin, Germany: Springer, 2004.
13. Zhang, S., Banerjee, P.P., Luciano, C.: Virtual exercise environment for promoting active lifestyle for people with lower body disabilities. In Proceedings of *the 2010 International Conference on Networking, Sensing and Control (ICNSC),* pp. 80-84, Chicago, IL, 2010.