

A Virtual Space for Children to Meet and Practice Chinese

Mei Si

Published online: 7 February 2015

© International Artificial Intelligence in Education Society 2015

Abstract Second language acquisition after the students have learned their first language is a unique process. One major difference between learning a foreign language and one's mother tongue is that second language learning is often 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 virtual learning environments suitable for children to practice their language and culture skills. We present a unique virtual environment that contextualizes the practice and engages the learners with narratives, encourages group work and leverages the power of embodied cognition in language learning. Our system has been deployed in an afterschool program for children from 6 to 8 years old. We report our evaluation results and reflections on the deployment process, followed by discussion and future work.

Keywords Language and culture learning · Virtual environment · Kinect · Embodied cognition

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.

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. It has been researched by the *New York Times* and determined that the number of students taking the Advanced Placement test in Chinese has grown so fast that it is likely to pass German as the third most-tested A.P. language, after Spanish and French (Dillon 2010). On the other hand, Chinese is a Category IV language classified by the Defense Language Aptitude Battery (DLAB), which means it requires a lot of practice for learners born in the US.

M. Si (✉)
Rensselaer Polytechnic Institute, Troy, NY, USA
e-mail: sim@rpi.edu

In this project, we investigate using computer game technologies for engaging children in language practice. We present a unique virtual environment that contextualizes the practice and engages the learners with narratives, encourages group work, and leverages the power of embodiment in language learning.

In our language practice environment, learners practice their language and social skills in the context of social interactions. We use computer simulated virtual environments and virtual characters to provide a base and guidance for these interactions.

The practice environments were designed for children. Different from the typical mouse and keyboard or touch pad software interface for adolescences or adults, the users can use Kinect cameras to interact with the system and with each other. Using their own body movements, the user can control the body movements of a virtual character or the user's view point—the position of the camera—in the game. They can do so and simultaneously have voice chat with other users.

The reason for encouraging users to move their hands and body during learning is two-fold. First, for young children this eliminates the needs of teaching them how to use a computer and can hopefully make the language practice process more like a game. Secondly, we expect this to be a more effective approach for practicing languages for all users. Freeing the user's hands from operating the computers will allow him/her to talk in a more natural way. Moreover, 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 (Barsalou 2008; Clark 2008; Noë 2004). Gesturing is a perfect example of embodied cognition. It is well-known that expressive gestures are an important aspect of language use and communication (Capirci et al. 1998). On the other hand, spontaneous gesturing, which does not directly relate to language use, has also been shown to play an important role in learning and the recall of abstract concepts (Goldin-Meadow 2003; Stevanoni and Salmon 2005). An explanation of this phenomenon is that memory can be off loaded to body-environment relationships that are “artificially” created by us.

In this project we created four Kinect enabled virtual environments with narratives and puzzles embedded in them. All of these virtual environments allow multiple users to log in from different locations using the Internet. We have also developed a mouse and keyboard version for users who do not have or want to use a Kinect camera.

Our games are designed to be supplemental practice materials to the learners' language classes. We hope that through integrating a narrative environment and a Kinect enabled embodied interface, this novel platform will provide the learners an engaging experience. We have performed a preliminary evaluation of the games with learners in the age range of 6 to 8. The results positively support our hypothesis.

In the next sections, we will review computer based learning systems and the theories of embodied cognition which lead to our hypotheses. We will then describe our system in detail followed by the experimental study and results.

Related Work

Computer Based Learning Systems

With the rapid development of computer technologies in recent years, there is an increasing trend of supplementing face to face teaching with digital media. Initially non-interactive media forms—audio and video recordings—were used. Presently, interactive media have been designed for language learning. Some of them have been deployed commercially and have achieved considerable success, such as the Tactical Language Training System (Alelo) and Rosetta Stone. Next, we review three different ways digital media have been incorporated into second language teaching.

First, a lot of software that was not designed explicitly for language teaching has been explored. For instance, students can use software in the target language (such as Microsoft Word, Adobe Photoshop, FinalCut Pro) to accomplish a specific task with the use of the software, and use digital storytelling software, such as VoiceThread, Audacity, PulpMotion, ComicLife, YackPack, among others, to tell stories, make presentations, or exchange information.

Secondly, Second Life hosts many language teaching/learning spaces. It has been noted that “Language learning is the most common education-based activity in Second Life” (Miller 2009). For example, a multi-year funding project by the European Commission yielded several projects (Hundsberger 2009). More specific simulation platforms have also been created. Digibahn is a 3D digital game environment for studying German (Neville 2010). The Center for Applied Second Language Studies (CASLS) and Avant Assessment, Centric, and Soochow University, created a MyChina Village Virtual Chinese Immersion Camp in 2009 using Second Life (MyChina Village 2009).

Finally, while most computer-based language learning systems are aimed at a standard personal computing environment, the effective use of language training has also been demonstrated in fully immersive virtual environments such as the Tactical Language Training System (Johnson et al. 2004). 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 et al. 2009). Chang, Lee and Si have investigated using immersive narrative and mixed reality for teaching Mandarin Chinese to college students (Chang et al. 2012). Sykes created a 3D multiuser virtual environment called Croquelandia, in which the user can practice making apologies in different scenarios in Spanish (Sykes 2013). Holden and Sykes also created a place-based mobile game called Mentira, which allows the users to practice Spanish in a task-based, goal-oriented interaction (Holden and Sykes 2013).

The enthusiasm for supplementing face to face teaching by a computer program goes beyond its use in language training. In recent years, computer aided pedagogical systems and intelligent pedagogical agents have been widely used for tutoring and training purposes, ranging from math (Beal et al. 2007) and physics tutoring (Ventura et al. 2004) to cognitive and social skills training (Traum et al. 2005), and from life style suggestions (Zhang et al. 2010) to PTSD (Rizzo et al. 2009) and autism interventions (Boujarwah et al. 2011).

Compared to traditional classroom training and human tutors, computer-based systems have several advantages. First, they can be used at the user's convenient time and location, and therefore increase the overall time the students can spend on the subject. Secondly, the system can be adaptive to each individual's need and progress, and therefore create a personalized learning environment. Thirdly, a computer-based system can foster learning and encourage interactions from users who are not comfortable interacting with other humans. Fourthly, a computer-based system can be used to record and keep track of the students' work, allowing for a more accessible and more organized re-presentation of the student's work (which helps both the student and instructor in their respective tasks), the taking care of various necessary but relatively unimportant bookkeeping tasks, and the collecting of user data that can be used to research the effectiveness of the program. Fifthly, computer-based systems can be copied and deployed at multiple locations at the same time, and therefore can often be a less expensive choice than human instructors or tutors.

On the other hand, using a computer based system may hinder some of the natural processes in language learning. Language and culture training involve more than reading, writing and speaking. In real life, people use non-verbal behaviors—gazes, gestures, and body movements—to accompany their speech. Most existing computer based learning systems require the user to sit in front of a computer and use keyboard and mouse to interact. Because their hands are tied to the task of operating a computer, using an interactive media in fact suppresses the learners' natural tendency of using their hands in language practice.

Gesture based natural user interfaces have been explored in cultural training (Rehm et al. 2010; Kistler et al. 2012). However, studies on integrating such interfaces with language and culture training are still rare. In this project we want to explore combining gesture based user interface with narratives and puzzles to provide the users a platform to practice their verbal and non-verbal skills together. As mentioned earlier, this design was also motivated by the theories of embodied cognition.

Theories of Embodied Cognition

Early theories of human cognition view cognition as a collection of mental processes and functions. Herbert Simon defines cognitive science as “the study of intelligence and intelligent systems, with particular reference to intelligent behavior as computation” (Simon and Kaplan 1989). This definition makes the computer, i.e. disembodied computation, a metaphor for how our mind functions. Gardner goes even further and argues that it was necessary (and sufficient) to “posit a level of analysis wholly separate from the biological or neurological, on the one hand, and the sociological or cultural, on the other” (Gardner 1985). This view suggests that human consciousness and cognition can be simulated or created from non-physiological forms using representations unrelated to our physical body.

The theories of embodied cognition, on the other hand, highlight the tight relationship between our body, mind, and the environment. It points out that many of our concepts are intimately related to aspects of our body and bodily movement. For example, according to this theory, our ideas and understanding of “up”, “down”, “over”, etc. are shaped by the nature of our body and its relationship to the environment. In fact, even many of our mathematical concepts, such as the continuity of a

function, which are supposed to be highly abstract, can be traced back to very basic spatial, perceptual, and motor capacities (Lakoff and Nunez 2001). Lave and Wenger explicitly claim that “there is no [mental] activity that is not situated.” (Lave and Wenger 1991).

Because of the tight relationship between action and cognition, physical movements help us to think, solve problems, and make decisions. For example, people often walk around a room while trying to decide where the furniture should go; and we do not rotate Tetris pieces because we know where they can go, but rather as a method to find out where they can go (Kirsh and Maglio 1992). Moreover, Margaret Wilson points out that off-line cognition is also body-based, and our memory is grounded in mechanisms that evolved for interaction with the environment (Wilson 2002).

Even though the theories of embodied cognition are not always cited, there is a large body of research which demonstrates the important role gesture plays in language use and learning (Kellerman 1992; McCafferty 2004; Morett et al. 2012). Goldin-Meadow and Alibali pointed out that gesture both reflects speakers’ thoughts and can change speakers’ thoughts; and therefore encouraging gesture can potentially change how people learn (Goldin-Meadow and Alibali 2013).

Gesturing plays an important role in learning, not only for language learning, but also for learning abstract concepts. This is because memory can be off loaded to body-environment relationships that are “artificially” created by us. Kinesthesia is such an example. In Goldin-Meadow and colleagues’ study, one group could freely gesture during an intervening math task, while the other group was asked not to move while conducting the same task. The results showed that the group that was not allowed to gesture performed significantly lower in the memory recall test than the group that was able to move around during the intervening math task. As Goldin-Meadow concluded, “the physical act of gesturing plays an active (not merely expressive) role in learning, reasoning, and cognitive change by providing an alternative (analog, motoric, visuospatial) representational format.” (Goldin-Meadow 2003) It appears that gestures continuously inform and alter verbal thinking, which continuously informs and alters gestures, forming a coupled system, in which the act of gesturing is not simply a motor act expressive of some fully neutrally realized process of thought, but instead a “coupled neural-bodily unfolding that is itself usefully seen as an organismically extended process of thought” (Clark 2008). In this project, we hypothesize that embodied interactive environments, which provide affordance for the natural use of gestures and body movements will benefit language learning.

Project Description

This project is aimed at leveraging computer game technologies and Microsoft Kinect camera to create engaging and embodied virtual environments for children to practice their language and culture skills in Mandarin Chinese. Next, we describe how the virtual environments function and the techniques implemented.

Overview of the System

Figure 1 provides an overview of how the system functions. This system contains a central server and multiple clients. The server handles most of the computation, and



Fig. 1 Multi-user multi-input support

therefore low-end machines can be used as clients. Each client needs to be equipped with a Kinect camera, a microphone and a 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.

The system supports both Kinect camera based interaction and traditional mouse and keyboard based interaction. Using either types of inputs, the users can control a character's body movements and simultaneously have voice chat with other users. Multiple users can log into the system from different physical locations and interact with each other in the virtual spaces just like in other multi-player games. When the user first logs in, a unique ID is assigned, and the server keeps track of each ID's activities across multiple sessions. By default, the clients connect to a server set up in our lab. With requests, we can also configure the system to allocate the server elsewhere.

System Development

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, such as logging into Skype and adding each other as contacts. Considering the age group of our users, we believe this is an especially important feature.

The Unity game engine and most other commercially available game engines do not support voice chat. 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 if needed. Similar to the main server for the games, this voice chat server can be located somewhere else if needed.

Environments and Characters

Four Chinese practice environments were created, as shown in Figs. 2, 3a, b, 4, and 5. These games are designed as supplementary materials to the learners' regular Chinese classes. Currently, the menus and the instructions are displayed in English to help the users understand how to play the game. To fit learners of different ages and different Chinese language skills, these environments were designed to support different levels of open ended conversations. The social interaction component is emphasized in all of the scenarios. The users are encouraged to and most of the time have to discuss and collaborate with each other to solve the problems presented in the virtual environments. We will describe each of the environments below and how they can be used by the teachers.

Environment 1: Cao Chung Weighing an Elephant

This scenario is designed to help the learners practice sentences and words in three different categories: numbers and simple math; directions such as up, down, left, right; and simple movements for manipulating objects such as picking up and dropping down.

We designed a 2.5D virtual environment (Fig. 2) based on a traditional Chinese children's story—Cao Chong Weighing an Elephant:



Fig. 2 Environment 1



Fig. 3 a Environment 2 showing part of the object. b Environment 2 showing the full object

This happened about 1700 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

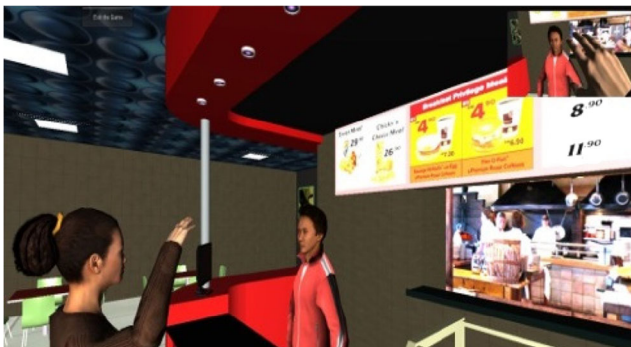


Fig. 4 Environment 3



Fig. 5 Environment 4

transported 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 determined the weight of the elephant.

In this virtual world, the users are asked to play the children in the story and to find the right way to weigh an elephant without hurting it. The users are provided multiple tools, such as a knife, which is not big enough to chop the elephant into pieces but could 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.

In this scenario, the learners can practice sentences and words related to directions and movements, such as up, down, left, right, inside, outside, pick up and drop down. The learners can practice these phrases while performing the corresponding actions, which provide a positive feedback loop. Moreover, because controlling a character's movements through a Kinect camera requires practice, these activities add additional game playing features to the practice.

This scenario also helps the learners practice doing simple math in Chinese, i.e. they have to remember the weight of each stone they have weighed and calculate the total weight. Of course, before the learners figure out the right way to weigh the elephant, a general unrestricted discussion among the learners might also take place.

For learners who have already successfully completed the game before, we enable replay by changing the initial locations of the tools and sometimes their functions. For example, the first time the game is played, the elephant will only follow the learner if the learner gives a banana to the elephant. In later trials, the banana is replaced by an apple or other fruits.

Environment 2: The Elephant and the Blind Men

This scenario is designed to help the learners practice sentences and words related to color, shape, and the type of an object, such as whether it is a fruit or an animal. The practice scenario is created based on another traditional 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 difficulty for this practice. Figure 3a shows an example of what a learner sees in the game. 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. However, the user will never be able to see the whole object. We designed this function to encourage body movements in all of our games.

When the learners have trouble naming the object, the name of the object can be displayed in Chinese as a hint. As shown in Fig. 3a, the Chinese characters in the lower right corner are the right answer to the current question. Another hint the system can give is to switch the camera view, and allow the user to see the object from an alternative perspective. Once the users provide the correct answer, the whole object will be displayed with a congratulatory message, as shown in Fig. 3b.

The vocabularies used in this game are mostly related to color, shape, and the type of objects, i.e. food, plants and animals. We provide a background story—the elephant and the blind men—to motivate the learners to engage in the practice. However, the actual conversation practiced in this scenario is not related to the story.

This game is configurable. By default, we have about 50 objects in the game, covering basic vocabularies related to food, animals and plants. If the teachers

want to have additional objects, or control the sequence of how the objects will be shown to the students, the game can be configured accordingly. Currently, the configuration process is not fully automated. We have to rebuild the game after adding new materials. We plan to improve on this in our future work.

Environment 3: A School Café

We have created two 3D environments with 3D characters in them, as shown in Figs. 4 and 5 below. Environment 3 is a student café with realistic models of two human characters in it. It is designed to help the learners practice talking about food and school activities. The users can control either character, “walk” around in the environment, talk with the other character, and physically interact with objects in the room. For example, the user can pick up a cup and move it to the counter. The user can also collaborate with another user and engage in physical activities together, such as pushing the tables around or moving a heavy box on the ground. These activities allow the learners to in addition practice sentences and words associated with movement coordination. They are also designed to add the “fun” factor to the game.

Because the characters are modeled in 3D, when Kinect camera is used, the characters directly mimic the user’s body movements. This gives the characters a richer set of movements compared to the 2D characters in Environment 1. When mouse and keyboard are used, only a restricted set of movements are supported.

Environment 4: A Chat Room

Finally, the last virtual environment is a chat room. The characters in it are modeled in cartoon style. The characters can’t move around in the room. Other than that, the user has similar controls over the characters as in the school café environment.

This environment supports more open ended conversations. Any face to face conversation can be moved to this virtual space. This gives advanced learners a chance to explore and even invent a game with their own rules. Alternatively, teachers can design scripts for students to act out in the virtual environment.

The characters are positioned in a way that gives them good views of each other’s body and arm movements. We hope this environment provides an easy to use alternative to Skype or Google Handout which also support video chat with multiple participants. Other than inputting a user ID at the very beginning, there is no additional step for setting up a conversation in this environment. As we have discussed before, we believe this level of simplicity is important for our user group. Moreover, in this environment, each user can have a good view of all other users instead of just the speaker. This allows them to see both the speaker’s non-verbal behaviors and those of the listeners’. We hope this can help the users learn culture related gestures and other non-verbal behaviors easier.

Deployment and Evaluation

Motivation is essential for learning. We hope that by using multi-player game, narrative, and Kinect camera enabled user interface, our learning environments can provide the learners an engaging experience, and therefore improve their learning results. We also hope that using the Kinect camera for controlling the characters will provide the learners a more embodied experience, and enhance their learning results compared to using the mouse and keyboard interface.

Children at different ages have different cognitive and motor skills which may affect the effectiveness of our learning environments. Younger children may behave differently and have different experiences from older children and adults using the same learning materials. As an initial evaluation, we conducted a month long study on the usability and effectiveness of our learning environments for young children (6 to 8 years old). Overall, the results are encouraging. Both the children and their teachers reported that they greatly enjoyed the practice, and the children's learning results were improved. Next, we report the evaluation process with discussions and planned future work. In particular, we report what we have learned from working with children in this age range.

Subjects

We performed our evaluation at Bethel Tutoring Center in Rosemead, California during July, 2013. Bethel Tutoring Center runs summer and afterschool programs for K-12 students. Many of their students come from Asian families, and there is already a need for Chinese classes.

Twenty children between the ages of 6 and 8 participated in our study. Most of these children come from Asian families, e.g. Chinese, Vietnamese, Cambodian and Malaysian and their parents or grandparents can speak Chinese. Many of them can understand simple Chinese, but none of them could speak or read before coming to the center.

Before this experiment, the students had taken 3 months of Chinese classes at the Center from April to June. The classes were twice per week for 30 min each. The class did not use a textbook. Instead, the teachers used flash cards to teach frequently used vocabulary words. The students were taught both how to say the words and how to read from written (simplified) Chinese. The teachers also practiced with the students common phrases and simple sentences they can use in everyday lives, such as “eat lunch” and “go to rest room.” Typically the first twenty minutes are used for teaching new vocabulary and phrases. The last 10 min are used for rehearsing contents learned in previous classes. Thus, the students learn in an iterative fashion.

There was no formal exam. The students were evaluated based on their recalls of learned vocabulary and skills of communicating in Chinese during the review phases of the classes. We learned from the teachers that the learning results of students at this age range are much less stable than those of older children. The students' performances are largely affected by how much they have practiced recently. To fully engage students in learning and practice over time is often a challenge. It is often the case that only 60 to 70 % of the students actively participate in learning and practicing in class. Therefore, the teachers are very interested in trying out our games and hope the games can engage the students better.

Materials and Procedure

Two teachers participated in this study. One is the student's Chinese teacher. The other teacher helps to coordinate the class and only speaks very little Chinese.

The evaluation lasted a month. For conducting the study, the teachers replaced their regular Chinese classes with using our system. The students took two thirty-minute Chinese classes per week. Similar to how time was spent in their regular Chinese classes, during the first 20 min the teachers used our system to introduce new vocabulary and phrases. Then the teachers reviewed old content with the students for the remaining part of the classes.

The classes in the first week were mostly spent on teaching the students how to use the system. This step also provided us feedback on the usability of the games for children in this age range. During this informal usability evaluation, the teachers decided that Environment 1 and Environment 2 fit the level of the class. Environment 3 and Environment 4 require students to conduct longer conversations, which was too advanced for the students. Moreover, the movement coordination tasks, e.g. two persons pushing a table together in Environment 3 were too hard for the students at their ages. The teachers also made modifications to the game play to fit their class better. They considered both the limitation in resources, i.e. the class only has two teachers and can only occupy one classroom, and the students' age group when making the modifications. We will present these modifications in the [Evaluation Results](#) section.

For the remaining 3 weeks (six classes), Environment 1 was used during the first three classes and Environment 2 was used for the second three classes. Figures 6 and 7 show how the students interacted with our system. Evaluation data were obtained through the teachers' subjective reports after the evaluation period.

Evaluation Results

We report two types of evaluation results here. The first is how the learning environments were used in the classes, and the modifications made by the teachers. The second is a comparison of the students' learning results with and without using our system.

Fig. 6 Students practicing in environment 1



Fig. 7 Students practicing in environment 2



Modifications

The main modification made by the teachers is that instead of individual or small group practice, which is what our system is originally designed for, the whole class worked together as a large group. The teachers found this to be a more effective way of engaging the class. Typically the students were very excited, and they all want to play the games. Instead of having the students play the games one by one, the teachers picked one or two of them for each game session, and the whole class contributed to the solutions by discussing with the selected students.

Other modifications are made in the direction of making the learning environment multi-purpose—teaching the students multiple subjects at once. We will describe these modifications below as we describe the typical procedures of how the two learning environments were used in the classes.

Using Environment 1

Because of the ages of the students, the teachers told them the right solution to the problem at the very beginning—they have to find the stones that will make the boat sink to the same level as the elephant and then weight the stones. So the game became looking for stones to be put into the boat, remembering the weights of the stones and summing the weights up. For conducting this evaluation, we helped the center set up two laptops with Kinects. Though the game supports up to four players at the same time, the teachers did not have additional computers to use. Therefore, typically two students were selected to play the game together. The Kinect version of the game was used. When the two selected students controlled the characters moving in the virtual world, the teacher led the class in reciting the names of the corresponding actions, such as “picking up” and “dropping down”, and the names of the objects, such as stone and banana.

The teachers also used this scene to teach the students how to do math. For older children and adults, summing up a few numbers is not a hard task. But for this age group, it is. The class did the practice in both Chinese and English. In fact, one of the suggestions from the teachers was to enrich the math component of this learning environment, so that they can teach the students both Chinese and math at the same time.

Using Environment 2

For this game, the whole class worked together, and only one computer was used. The mouse and keyboard version was used because they didn't find using a Kinect camera for rotating the camera angle in the game to be an attractive feature.

Each time a new picture was shown, the teacher asked the class to guess what it was. The students then had a discussion that was semi-moderated by the teacher. A lot of such discussions were in English, and the teachers reminded the students about words that they had already learned in Chinese. If the discussion couldn't reach a conclusion after a while, which happened very rarely, the teachers would look at the hint and tell the students the answer. Otherwise, once the whole group settled upon one answer, either a teacher or a student would enter the answer into the system. The teachers would then lead the class reciting the words a few times.

Similarly as when using Environment 1, the teachers wanted to multi-purpose this practice. Our system can display the name of the objects in Chinese and accepts answers in English. Our original design was to let the teachers or parents help children who can't spell inputting the answers. The teachers instead used this scenario to teach the students how to read in Chinese and how to spell the corresponding English words.

A Comparison of Learning Results

In general, the results are encouraging and consistent with our expectations. The games were well received by the students and the teachers. The teachers reported that the students were more active in practice and learned faster than in their regular classes.

Table 1 summarizes the teachers' reports of the students' progresses. The teachers compared the students' achievements during the evaluation month with their achievements during the first 3 months of the Chinese classes. The data in the table were estimated by the teachers. We used the teachers' subjective judgments because for this

Table 1 Comparison of engagement and learning outcome

	April to June (prior classes)	July (evaluation)
Engaged Students	60–70 % of the class	80 % of the class
Listening	50 % can understand simple sentences	50 % can understand simple sentences
Vocabulary	<ul style="list-style-type: none"> • 25 % learned about 100 words • 50 % learned about 50 words • 20 % learned about 20 words • A few students can only say a few words 	<ul style="list-style-type: none"> • 25 % learned 25–30 words • 50 % learned about 15 words • Everyone improved their recall of old vocabulary. On average, they can remember 20 % more of the words
Speaking	<ul style="list-style-type: none"> • Most students learned about 10 sentences • The students have different levels of fluency in speaking • Most students can't create new sentences by themselves 	<ul style="list-style-type: none"> • 60–70 % improved their pronunciation and sentence organization for the sentences they have learned before • 50 % can produce simple sentences with 2 or 3 words, such as “move up” and “drop off” through using Environment 1

age range their Chinese classes does not have formal exams and based on our conversations with the teachers a onetime assessment does not seem to be the best way for evaluating young children's language skills. In the future, we will consider working with the teachers to develop objective measurements for learning results.

As we have mentioned before, none of the students could speak or read in Chinese before coming to the center. After 3 months of classes, the students' language skills varied a lot. Most of the students were able to say a dozen of sentences for everyday uses with various levels of fluency. These sentences were taught by the teachers, and the students were not able to create new sentences by themselves.

In terms of vocabulary, 4–5 students (25 % of the class) knew about 100 Chinese words—they could read, speak and understand from listening; 10 students (50 % of the class) knew about 50 words; 2–3 students (25 % of the class) knew about 20 words, and 1–2 students could only speak a few words even after 3 months of classes.

For listening comprehension skills, half of the students could understand simple sentences before they came to the center. This is mainly because of their family background. In general, the students' listening comprehension skill did not change during the classes.

Most of the students improved on their vocabulary and skills of constructing and using sentences during the evaluation month. In addition, more students were engaged in the practices—they actively participated in the practice. The teachers told us that for young children, it is extremely important to keep rehearsing what they have learnt before. Otherwise, they will forget very quickly. The teachers believe that most of the improvements shown in the students' performances were resulted from the students being more engaged in the practices. The teachers also mentioned that by using the games, the students practiced together more often which is another reason for their improved performances.

The number of new words the students learned during the evaluation period is roughly equivalent to what they have learned on average for 1 month in the past. However, their recall of old vocabulary has improved. This is especially true for the students who had poor performances in the past. For the top students, they had already mastered most of the words taught before and therefore there was not much room for improvement. As a result, the class' vocabulary sizes were more even after the evaluation period. On average, the students remembered 20 % more of the old content after the evaluation month.

During the evaluation, the students improved their pronunciation and sentence organization for the sentences and phrases they were previously taught. Though not objectively measured, the teachers said that the students learned much faster than before, and learned more than the students in similar classes from previous years. By the end of the evaluation, 50 % of the students could produce simple sentences with 2 or 3 words, such as “move up” and “drop off.” They developed this skill through using Environment 1.

We did not observe an improvement in the students' listening comprehension skills in this study. One month is probably too short for improving this skill.

We consider our games being engaging to the children as our major achievement. Motivation is essential for learning. Based on the teachers' reports, the improvements in the student's language skills are largely due to their enthusiasm to participate and

practice. The teachers liked using our games as a teaching platform a lot. As side evidence, after the evaluation period, they continued using the games in their classes for another month until most students had mastered the content in Environments 1 and 2.

Discussion and Future Work

We had two hypotheses when creating this system. One is that by using multi-player game as a platform and combining narrative and Kinect camera enabled hand-free user interface, we can create an engaging language practice environment. This hypothesis is partially confirmed. Both the teachers and the students in our evaluation have shown great enthusiasm towards using our system, and the teachers believe that the improvements in the students' learning results are largely due to their improved engagement in class. However, the current evaluation was conducted in a group fashion—the whole class interacted with the games as a large group, and the teachers moderated the interaction. Whether the games are engaging when students practice alone and only virtually meet with their peers and teachers still needs to be investigated. In the current study, only Environment 1 and Environment 2 were used, and the students were only 6 to 8 years old. Additional evaluations are needed for studying learners' behaviors in other environments, and for learners in different age groups.

Our second hypothesis is that we hope using the Kinect camera for controlling the characters will provide the learners a more embodied experience, and enhance their learning results compared to using the mouse and keyboard interface. Our current evaluation does not provide a concrete answer to this question. The teachers and children's preference for playing the Kinect enabled game in Environment 1 suggests that the embodied interface is engaging. However, we are not sure of its impact on learning outcomes. We plan to conduct controlled empirical studies on how the forms of the interaction, i.e. mouse and keyboard vs. Kinect based interface affect the learners' learning behaviors and outcomes. We suspect that the learner's age and language skill may also mediate the effect. We think our games, and in particular Environment 3 and Environment 4 can also be used by college students. In addition to working with after school programs and weekend Chinese schools, we plan to extend our evaluations to using students who take colleague level Chinese classes. Rensselaer Polytechnic Institute, for example, offers Chinese in three difficulty levels, which allows us to have control over the learners' language skills before using the system.

In future evaluations with teenagers or colleagues students, we will work with the teachers for developing more objective criteria for evaluating the students' learning outcomes. In the current study, because the children were fairly young, their retention of vocabulary and phrases is less stable compared to older children and is very sensitive to the amount of practice they had recently. Therefore, we relied on the teachers' subjective evaluation for getting a general idea of whether our system can benefit these children. We consider our system a success for this age range because it effectively motivated the children to practice more.

In addition to conducting more formal evaluations, we hope to extend the current system. There are three directions that we are particularly interested in. The first one is suggested by the teachers that we can design learning scenarios that enable children to combine their study of Chinese with practicing other skills, such as math skills and even their English skills. The idea of Content and Language Integrated Learning (CLIL) has been explored by many researchers and for children of different ages (Baştürk and Gulmez 2011; Snow 1998; Willis 1996). However, integrated virtual environment for this purpose is still rare. In the future, we would like to work with educational experts and school teachers for developing such a system.

Secondly, we hope to add AI controlled characters into the virtual environments and allow children to interact with them. The AI characters will be configurable to the teachers. Using such characters, the system will be able to provide more consistent user experience over time.

Finally, a natural question for any language learning applications is how the system encourages learners to practice the new language, e.g. Chinese, and not using their mother tone, e.g. English. Currently, our system does not include any feature for this purpose. In the evaluation, we relied on the teachers to make sure the children speak in Chinese. For future work, we are considering adding a spoken language identification module which will warn the users if they speak in English. Combined with AI controlled characters, we will be able to leverage the power of social norm and narrative for keeping the user speaking in Chinese. For example, if the user starts to talk in English, the AI controlled characters can look confused or even mad at the user.

Conclusion

Learning a language requires a tremendous amount of practice both inside and outside of the classroom. One common challenge faced by teaching students of all ages and, in particular, young children is how to engage them in practice.

The rapid development of computer technologies in recent years enables a variety of user interfaces to be conveniently 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.

In this project we propose to attack the language training challenge by creating virtual spaces with narratives and puzzles embedded in them, and encouraging children to actively move their bodies during their language practices. We created four Kinect enabled virtual environments for this purpose. Our initial evaluation has received positive feedback. The current evaluation confirms the importance of motivation in learning. It also demonstrates that our system can be used by children as young as 6 years old. On the other hand, this evaluation is preliminary, and additional studies are needed to answer other questions we have regarding using such embodied virtual environments for language learning.

In summary, this work provides 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.

References

- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Baştürk, M., & Gulmez, R. (2011). Multilingual learning environment in French and German language teaching departments. *TOJNED: The Online Journal of New Horizons in Education*, 1(2), 16–22.
- Beal, C. R., Wallis, R., Arroyo, I., & Woolf, B. P. (2007). On-line tutoring for math achievement testing: a controlled evaluation. *Journal of Interactive Online Learning*, 6(1), 43–55.
- Boujarwah, F., Riedl, M. O., Abowd, G., & Arriaga, R. (2011). REACT: intelligent authoring of social skills instructional modules for adolescents with high-functioning autism. *ACM SIGACCESS Newsletter*, 99.
- Capirci, O., Montanari, S., & Volterra, V. (1998). Gestures, signs, and words in early language development (in The nature and functions of gesture in children's communication, J.M. Iverson & S. Goldin-Meadow, Eds.). *New Directions for Child Development*, 79, 45–60.
- Chang, B., Sheldon, L., & Si, M. (2012). Foreign language learning in immersive virtual environments. In Proceedings of *IS&T/SPIE Electronic Imaging*, Burlingame, CA.
- Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York: Oxford University Press.
- Dillon, A. (2010). Foreign languages fade in class—except Chinese. Retrieved March 2, 2014, from <http://www.nytimes.com/2010/01/21/education/21chinese.html>.
- Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.
- Goldin-Meadow, S. (2003). 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: MIT Press.
- Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257.
- Holden, C. L., & Sykes, J. M. (2013). Complex L2 pragmatic feedback via place-based mobile games. In N. Taguchi & J. M. Sykes (Eds.), *Technology in interlanguage pragmatics research and teaching: The archaeology of magic in Roman Egypt, Cyprus, and Spain (chapter 7)* (pp. 155–183). Amsterdam: John Benjamins Publishing Company.
- Hundsberger, S. (2009). Foreign language learning in second life and the implications for the resource provision in academic libraries. <http://arcadiaproject.lib.cam.ac.uk>.
- Johnson, W. L., Marsella, S. C., Mote, N., Si, M., Vilhjalmsson, H., & Wu, S. (2004). 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.
- Kellerman, S. (1992). 'I see what you mean': The role of kinesic behaviour in listening and implications for foreign and second language learning. *Applied Linguistics*, 13(3), 239–258.
- Kirsh, D., & Maglio, P. (1992). Some epistemic benefits of action: Tetris a case study. In *Proceedings of the Fourteenth Annual Cognitive Science Society*, Morgan Kaufmann.
- Kistler, F., Endrass, B., Damian, I., Dang, C. T., & André, E. (2012). Natural interaction with culturally adaptive virtual characters. *Journal on Multimodal User Interfaces*, 6(1–2), 39–47.
- Lakoff, G., & Nunez, R. (2001). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York: Basic Books.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation (learning in doing: social, cognitive and computational practices)*. Cambridge: Cambridge University Press.
- McCafferty, S. G. (2004). Space for cognition: gesture and second language learning. *International Journal of Applied Linguistics*, 14(1), 148–165.
- Miller, J. (2009). Linden lab vice president of platform and technology development <http://222.virtualwordsnews.com/2009/05/out-of-stealth-8d-taps-language-learners-bots-microtransactions.html>.
- Morett, L. M., Gibbs, R. W., & MacWhinney, B. (2012). The role of gesture in second language learning: Communication, acquisition, & retention. In proceedings of *CogSci*.
- MyChina Village (2009). <http://mychinavillage.uoregon.edu/learn.php?section=overview>.
- Neville, D. (2010). Structuring narrative in 3D digital game-based learning environments to support second language acquisition. *The Foreign Language Annals*, 43(3), 445–468.
- Noë, A. (2004). *Action in perception*. Cambridge: MIT Press.
- O'Brien, M., Levy, R., & Orich, A. (2009). Virtual immersion: the role of CAVE and PC technology. *CALICO Journal*, 26(2), 337–362.
- Rehm, M., Leichtenstern, K., Plomer, J., & Wiedemann, C. (2010). 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*.

- 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., & Gahm, G. (2009). Development and clinical results from the virtual Iraq exposure therapy application for PTSD. In proceedings of *IEEE Explore: Virtual Rehabilitation*, Haifa, Israel.
- Simon, H. A., & Kaplan, C. A. (1989). Foundations of cognitive sciences. In M. I. Posner (Ed.), *Foundations of cognitive sciences* (pp. 1–47). Cambridge: MIT Press.
- Snow, M. A. (1998). Trends and issues in content-based instruction. *Annual Review of Applied Linguistics*, 18, 243–267.
- Stevanoni, E., & Salmon, K. (2005). Giving memory a hand: instructing children to gesture enhances their event recall. *Journal of Nonverbal Behavior*, 29(4), 217–233.
- Sykes, J. M. (2013). Multiuser virtual environments: Learner apologies in Spanish. In N. Taguchi & J. M. Sykes (Eds.), *Technology in interlanguage pragmatics research and teaching: The archaeology of magic in Roman Egypt, Cyprus, and Spain (Chapter 4)* (pp. 71–100). Amsterdam: John Benjamins Publishing Company.
- Traum, D., Swartout, W., Marsella, S., & Gratch, J. (2005). Fight, flight, or negotiate: Believable strategies for conversing under crisis. In proceedings of *the 5th International Conference on Interactive Virtual Agents*, Kos, Greece.
- Ventura, M. J., Franchescetti, D. R., Pennumatsa, P., Graesser, A. C., Jackson, G. T., Hu, X., Cai, Z., & the Tutoring Research Group. (2004). 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: Springer.
- Willis, J. (1996). *A framework for task-based learning (Longman handbooks for language teachers)*. Harlow: Longman.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9, 625–636.
- Zhang, S., Banerjee, P. P., & Luciano, C. (2010). 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.