

Interaction Challenges in AI Equipped Environments Built to Teach Foreign Languages Through Dialogue and Task-Completion

Rahul R. Divekar¹, Jaimie Drozdal¹, Yalun Zhou¹, Ziyi Song¹, David Allen¹,
Robert Rouhani¹, Rui Zhao¹, Shuyue Zheng¹, Lilit Balagyozyan¹, Hui Su^{1,2}

Rensselaer Polytechnic Institute¹
Troy, NY 12180
{divekr, droz dj3, zhouy12, songz2, allend5,
rouhar2, zhaor, zhengs6, balagl}
@rpi.edu

IBM Thomas J. Watson Research Laboratory²
Yorktown Heights, NY 10598
{huisuibmres}
@us.ibm.com

ABSTRACT

As cities around the world become more diverse in culture and language, there is a growing need for learning foreign languages. To further this excitement, we have built a human-scale, immersive room with a virtual AI agent that aids foreign language learning. Our system aids the language learning process through task-completion exercises using multi-modal dialogue. The Cognitive and Immersive Room (CIR) is developed as an immersive Chinese restaurant to teach Mandarin, but the interaction challenges and solutions can be reasonably generalized to other languages taught using similar techniques. As users interact with the immersive environment and the virtual AI agent, they face several user interaction challenges. These challenges arise from new learners' lack of proficiency in the foreign language. By studying user interactions in the CIR, we were able to articulate some of the interaction challenges. We have enhanced the AI agent, virtual environment, and the on-boarding process for new users to mitigate these challenges. The enhancements and the results which show that they were effective are discussed here.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation (e.g. HCI): User Interfaces

Author Keywords

Language teaching; Mandarin teaching; Immersive room; Intelligent agent; Multi-modal interaction; Mandarin as a foreign language; Three dimensional virtual worlds

INTRODUCTION

In this work we are addressing a universal problem. People coming from different backgrounds want to learn a foreign

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DIS '18, June 9–13, 2018, , Hong Kong

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-5198-0/18/06...\$15.00

DOI: <https://doi.org/10.1145/3196709.3196717>

language. A common and effective way to learn a foreign language, after learning the words and grammar structures, is by speaking and listening to it [35]. Practicing dialogues in the context of completing a task that one might face in real life, like ordering food at a restaurant, can help the language learning process [38]. However, the avenues to do this are limited because fluent, reliable speakers who are willing to help are hard to find.

Currently, our immersive Chinese restaurant, also referred to as the Cognitive and Immersive Room (CIR), targets college-level students who are primarily English speakers and are learning Mandarin as a foreign language. The terms *users*, *learners* and *students* are used interchangeably hereon. We refer to L1 and L2 in this paper, which are the language students know well and the language students are newly learning, respectively. In our immersive system, L1 is English and L2 is Mandarin.

In this work, we give an overview of Mandarin as a language, highlighting its complexities for learners. We discuss current state-of-the-art technology available for learning Mandarin, and we emphasize the effect, but lack of immersive, human-scale, dialogue enabled, and task-based systems. After outlining the need for an advanced system for the purpose of Mandarin learning, we define expectations of our system and discuss the interaction environment, AI agent, and task (scenario) developed.

The CIR has a special technology relationship between artificial intelligence and immersive technology. The room is equipped with a virtual AI agent that can listen, speak, and "see" its users. The intelligent and immersive capabilities serve two major purposes. The AI agent allows students to practice their language skills through conversation, and the immersive environment gives more visual context to the task they are trying to complete with the agent. In the scenario we have developed, the immersive environment resembles a Chinese restaurant. The users are guests ordering food, and the AI agent is a waiter referred to as *AI Agent*, *AI Waiter* or *waiter* hereon.

This is a new paradigm of human-scale visuals and AI agents combined in the context of language learning through task completion. However, there are gaps between the new tech-

nologies and the users that we need to fill. In this paper, we discuss two main themes of interaction challenges: the system not understanding the *imperfect* language of the students, and the students not understanding the *fluent* language of the AI agent. Both are due to the lack of L2 proficiency in the users. We address the communication gaps between the AI agent and the users by enabling multi-modal interactions and verbal help features.

We have run many beta-tests of the system on native and non-native Mandarin speakers from a local university. The core of this paper describes the details of these tests and the enhancements built as a result of them. The paper concludes with an analysis of a formal user-study. Sixteen students from a Chinese Level-1 class taught at a local university were invited to interact with the system. The observations and feedback from the students further motivate the direction for next steps in language teaching via AI and immersive technologies.

In this new learning paradigm, there still remain several other interaction challenges that are articulated here. Natural interactions do not follow a script, and therefore, it is challenging to create a system that teaches language through natural interactions. Nonetheless, these challenges are a motivation to the research community to build AI agents that, especially in the language education domain, can pro-actively mitigate the challenges between users and new technologies.

MANDARIN - THE LANGUAGE

In this section, we provide an overview of Mandarin as a language so that readers are able to understand the interaction challenges and the help functions described in later sections. Mandarin is an increasingly popular language in the United States[44]. The Foreign Service Institute rates Mandarin as a category 5 language, making it one of five most difficult languages in the world [3]. Because of its level of difficulty and high demand for learners, Mandarin language-learning programs, now more than ever, need effective language-learning technology.

Mandarin is a complex, intricate language that can be represented by characters (*Hanzi*) or by *pinyin* [22]. *Pinyin* uses the Roman alphabet to represent Mandarin Chinese sounds and allows non-native speakers to grasp the Chinese pronunciation and tone system better [34]. Fig. 1 shows *pinyin* in red on line 1 and *Hanzi* on line 2. Understanding only one of these representations will cause a severe deficit in communication. Mandarin is a tonal language, which means the tone of a syllable carries meaning. It is an especially difficult language for non-native speakers whose first language is not a tonal language [40]. Each syllable in Mandarin carries one of five

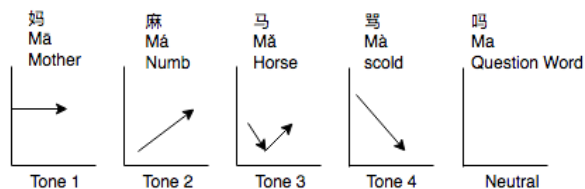


Figure 1. The Five Tones of Mandarin

possible tones, and each tone has a specific pitch, modulation, and duration. In written *pinyin*, a diacritical mark above each

syllable specifies what tone is associated with the syllable [22]. As shown in Fig. 1, depending on the tone, the syllable 'ma' can translate to English as mother, numb, horse, to scold, or as a question particle. Because of these subtleties, accurate tone recognition and pronunciation are critical to Mandarin speakers and new learners. Furthermore, there are many grammatical conventions and structures that rely on tones (tone sandhi[22]) and influence the meaning of a statement.

New learners of Mandarin, especially new learners with a non-tonal first language, have a wider pitch range than native Mandarin speakers. The mispronunciations associated with their novice and inarticulate speech lead to mis-communications [26]. Only with practice and exposure to the language can a new learner of Mandarin become fluent in their tone pronunciation and recognition. For more information on Mandarin as a language see [13]

In this overview, we see the numerous ways learners of Mandarin can go wrong with communication. In the next sections, we will describe our technology which is an avenue for the students to practice spoken Mandarin. This section showed why it is easy for the building blocks of our technology such as one of the most sophisticated natural language processing systems ([18]) to misunderstand or misinterpret speech input from new learners. The challenges of learning Mandarin as a foreign language also explain why users misinterpret or completely do not understand what they hear from the system. There is clearly a communication gap between the system and the users while communicating in L2. This work aims to ease this communication gap.

REVIEW OF CURRENT TECHNOLOGY

In this section, we review the existing applications and technology methods out there to aid Mandarin language learning. We then motivate what is needed to add on top of the current state-of-the-art.

Available Applications

There are numerous mobile applications and Internet resources that new learners can utilize to practice and improve their Mandarin. The most common applications for language learning focus on vocabulary and grammar learning [1]. A student is expected to rehearse and memorize the vocabulary through simple activities and is then tested on their knowledge. This method of learning relies largely on memorization and does not provide the foreign language exchanges, specifically opportunities to engage in authentic linguistic and social practices, considered necessary for the acquisition of a second language [43][38][24][54].

Other language teaching applications allow users to interact with native speakers from all over the world. These online programs, where a new learner can chat with someone and receive feedback, have a more immersive element because the user is able to speak directly with a native speaker. Some applications ([12]) allow anyone to register as a native speaker and chat with other users. The interactions occur through a text messaging system where users can edit and provide feedback on the grammar and vocabulary of the other person's messages in the conversation.

On the other hand, there are applications (e.g. ChinesePod[9],

eChineseLearning[10], busuu[7], tutor Mandarin[15]) that provide Mandarin-teaching experts, not just anybody, to interact with users. As Mandarin is a tonal language, pronunciation is a focus of many Mandarin learning applications (Wai Chinese[16]). Liao et al. [40] developed an unreleased, prototype of an adaptive Chinese pronunciation training system that analyzes a users pronunciation and provides feedback on phonetic and tonal errors. This prototype, along with many other computer-assisted pronunciation training (CAPT) systems, shows tone-analysis charts based upon the speaker's utterance. The user receives a real-time visualization of their pronunciation and can adjust their tone(s) to match a the tone-analysis chart of a native-speaker. However, unlike commercially available CAPT systems, Liao et al.'s prototype allowed for real-time feedback on pronunciation. Currently, there are no published papers on any user studies with this prototype. While these resources are beneficial in the education domain, we see an opportunity for enhancing these experiences by simulating real-word environments and scenarios.

Available Immersive Environments

Traveling to a foreign country is a highly praised approach for second-language acquisition [49]. When someone is immersed in a foreign culture, they are surrounded by and speaking the target language. Out of necessity, they will have to learn how to function in society and interact with others with a language they are newly acquiring. Studies have shown linguistic gains in people who study abroad [23][28][33][31]. However, Wright & Cong [59] reference multiple studies that show conflicting results, and they emphasize the lack of consistent methodologies in research on studying abroad. Due to geographical and financial constraints, it is often a challenge for new learners of a language to travel abroad and immerse themselves in that language. Perhaps a virtual, culturally immersive environment is the next best thing.

Lin & Lan [42] report two case studies, Vogel et al. [56] and Liao & Chen [41], that analyzed a total of 53 studies. The meta-analyses show positive findings (e.g. superior cognitive outcomes and more positive attitudes toward learning) associated with the use of interactive simulations and games compared to respective traditional teaching methods. Similarly, Cok [27] performed a meta-analysis of 37 papers and reports linguistic gains associated with using computer and Internet-based technology. She suggests that these gains result from allowing users to be in touch with authentic speaking situations and to experience interaction, neither of which can be achieved in traditional learning environments. The positive findings associated with computer-assisted language learning (CALL) highlight the potential of creating immersive, computer simulated environments for education purposes.

Virtual Worlds

In the context of this discussion, there are many ways to think of the term 'immersive' in regards to different environments. Systems that are virtually immersive are equipped with a 3D virtual world (VW) and are used across a variety of domains for multiple purposes. Berns[20], Gonzales-Pardo, & Camacho [38] define virtual worlds as '3D environments where users' graphical representations, called avatars can interact

with other avatars as well as objects within the environment'. They argue that VWs are 'highly immersive' environments because of their ability to reproduce or simulate a tremendous amount of real life activities. Furthermore, VWs provide great value to foreign language learning because of the interactions and inferences they allow. A 3D world allows users to explore linguistic concepts such as 'asking for directions' or 'ordering a meal' and learning propositions, among others [57].

An example of this type of immersive environment is SAIL (situated avatar-based immersive learning), a habitable 3D virtual space where users are able to control avatars and manipulate objects in the virtual world in the context of problem solving activities [52][39]. VEC3D, a virtual English classroom, is an interactive English learning environment in which users role-play open ended scenarios. Users conduct synchronous communication and can have real-time interactions in written or spoken format [52]. Most recently, a prototype of VEC3D 5.0 was developed linking the virtual world with Web 2.0 tools, but the efficacy of this system to learners studying English as a foreign language (EFL) has yet to be studied [53]. Roger Schank's DUSTIN [50] includes a digital environment that provides non-English speakers the opportunity to practice their English in a handful of scenarios. This system was developed for new employees at the company Accenture who are non-native English speakers that attend meetings at the company's U.S. training center near Chicago. To help with encounters the employees will have, scenarios enabled in the environment include getting through U.S. customs at the Chicago O'Hare Airport, checking into a hotel, and ordering food. Instead of computer generated graphics, the digital environment consists of videos of actors in the real-world locations [48].

Two of the most popular and commercial virtual learning environment systems are Active Worlds (AW) [4] and Second Life (SL) [5][52]. These systems are both Internet-based 3D worlds that are game like and include multiple users. Users choose and control avatars that they use to meet and chat with other users. In these virtual worlds instructors can host online projects and set up classrooms. These systems are often researched in the context of language learning because of their environments and the interactions they allow. A study of new learners of Hungarian in SL showed that '...the environment can generate a large quantity of productive and genuinely communicative language by learners, far more than would be possible in the normal classroom' [43].

Many studies on Second Life have focused on collecting qualitative and quantitative data to analyze student communication and interests and motivations within the virtual world. The studies of Toyoda & Harrison [55], Peterson [47], and Deutschmann et al. [30] focused on negotiation of meaning, strategy use during task-based interaction, and participation patterns during voice chat, respectively. While these studies did not assess learning gains, they shed light on the implementation of task-based learning in virtual worlds[47]. A number of studies measure learning gains associated with 3D virtual worlds (e.g. Canto & Ondarra [24]; Guzel & Aydin [35]; Milton et al. [43]; Henderson et al. [36]; Chen [25]), and a number of those studies have found significant effects of SL

on language performance (e.g. Canto & Ondarra [24]; Guzel & Aydin [35]).

Studies on virtual worlds other than SL and AW have also shown significant linguistic gains in users. Taguchi, Li, & Tang [50] developed a scenario-based interactive environment to improve student's knowledge of Chinese formulaic expressions. Results showed that students made significant gains in their knowledge of formulaic expressions after the practice, and this knowledge was retained two weeks later.

Yamazaki [61] studied Japanese learning throughout a 15-week curriculum in a 3D virtual world. Quantitative analysis confirmed statistically significant improvement in students' natural vocabulary acquisition, and qualitative analysis revealed statistically significant improvement in 'contextualized communicative competence' (e.g. persuasive talk, awareness of audience, and collaborative communication).

We plan to use these two studies, and others like these (e.g. Lan et al. [38]; Liao et al. [40]; Hismanoglu [37]), to motivate our research into applying the appropriate tools and pedagogy behind the CIR to produce language gains in the users.

Immersive Spaces

Apart from virtual worlds, immersive environments can be characterized by their physical environment and technologies. Katie Salen and David Birchfield's SMALLab (the Situated Multimedia Arts Learning Laboratory) is a room-sized interactive learning environment that integrates precise motion-capture technology and immersive floor projections. Currently, the system includes 30 scenarios, mostly related to science and mathematics, that support thousands of activities [14]. Research on this environment has linked it to student learning gains, improvement in teacher performance [14], and increase in student-led learning [21].

CAVE environments provide a high level of immersion through multi-screen projection systems that point to between three and six walls of a room-sized cube [45]. The first CAVE (Cave Automatic Virtual Environment) was developed in 1992[8], but since then many have been developed for applications in military training, medical education, manufacturing, and design. These environments are equipped with tracking systems that can track multiple users and their interactions in the environment. O'Brien, Levy, and Orich [46] compared the effectiveness of CAVE and PC technology in the language learning domain and found that complete immersion via CAVE technology did not provide significantly better experience for students. However, this study should not pose a threat to further research as linguistic gains were solely measured by the ability of a student to follow commands in the L2.

We believe that the immersive nature of virtual worlds can be enhanced by using panoramic displays and interactive 360 degree environments. The main inspiration for our CIR comes from a previous project that some members of the research team worked on. This project is an unreleased mixed-reality environment and learning game. Designed for a college-level Chinese language course, the game combines virtual reality with real props to create a mixed-reality environment where students must navigate through Beijing to find a missing book. This game provided a foundation and pedagogical backbone

that inspired us to implement a new interactive scenario. Currently, we have been unable to find any other studies on rooms with immersive technologies that attempt to improve a user's Mandarin through the use of task-based scenarios in a 3D world.

In the next section we describe the room we have built and set our readers up to understand the interaction challenges in it.

THE CIR ENVIRONMENT

This section describes describes the CIR. It provides detail on the display set-up, the visuals presented to the users, and the AI capabilities.

Physical Environment

The CIR holds a 360 degree panoramic display, as shown in Fig. 2. The computer render for this configuration is shown in Fig. 3. The edges, or lines, in the image show the start and end of each projector display. However, to a user in the room, it appears as a round and seamless screen, see Fig. 2. Each projector display has a diameter of 12 m and height of 3.8 m. We use five projectors to project on to the 360 display using special warping software. This panoramic display allows us to build environments that can visually immerse users.

The CIR is equipped with several Kinects and Point-To-Zoom cameras that act as the "eyes of the room". Software that runs on top of these sensors allows the room to recognize certain gestures made by the users, such as pointing. We made the design decision to have no intrusive devices that need to be worn or hand-held/laptop devices that need to be used to interact with the room. This is to make the virtual dining experience as real as possible with only the panoramic display. The CIR is also equipped with microphones, so the users can "talk" to the system. Natural Language Processing (NLP) software interprets what the users are saying and helps the system provide an appropriate response back to the users. Specifically, we use the Watson Assistant service [17] which we trained with pre-defined intents, examples, and [18] for speech-to-text processing. The utterance from the user, after being converted to text, is mapped to one of these intents. Based on the intent received from the NLP services and other contexts maintained locally, appropriate graphical and visual responses are chosen to execute. All of these individual intelligent services coming together in context constitute the AI.

The engineering and research behind the development and integration of displays, computer vision technologies, and natural language processing technologies are massive and scalable to other use-cases. We urge the more enthusiastic readers to see [32] for a sense of the entire technology stack and [62] for specifics on computer vision systems.

Ideal User Experience

Using the described physical environment, we have completed a running and integrated system based on the beginner college level curriculum from Modern Chinese: Simplified Characters Textbook 1A[19]. This curriculum is divided into different units, each one highlighting the Mandarin Chinese vocabulary, grammar, and culture behind the unit topic. The CIR currently focuses on Unit 4: Food and specifically aims to assist a student in getting through a common restaurant dialogue



Figure 2. A user interacting with the immersive Chinese restaurant in the 360 degree panoramic display. Best view in color. Demo Video: <https://youtu.be/LZWtDqhF1Ac>

in Mandarin. We chose this unit because Chinese restaurants are located all over the United States and can provide culturally rich environments. By increasing students' ability to successfully have a restaurant dialogue in Mandarin, they may be more comfortable or willing to carry out a similar conversation in their new language at a real Chinese-American restaurant.

Part of the curriculum as taught in the Chinese-1 class at the university involves students role-playing scenarios in pairs or groups. Specifically in Unit 4, students practice dialogues related to communication goals which represent conversations that may occur at a restaurant. These dialogue exercises are critical because they allow students to practice their pronunciation, which is critical for beginners of this tonal language, and because multi-player classroom course designs can enhance student motivation, attitude, and performance [60]. The CIR allows students to practice their dialogue through an exercise based on ordering food at a Chinese restaurant. This dialogue includes sub-tasks like asking for seating, ordering drinks and food, confirming the order for any mistakes, giving feedback on the taste of food, and paying the check. The user is not given a script to follow. To give the readers an example of what the interaction between a user and the waiter can look like a sample dialogue interaction is shown in fig. 4. A typical dialogue is longer and has more turns.

Using the L2 proficiency of the users and the capabilities of the physical environment, we have developed a cognitive and immersive system that simulates the environment of a Chinese restaurant. The waiter of the restaurant is a panda that appears to be floating. Users can have a conversation with this panda in the context of dining while they are surrounded by the visuals of the Chinese restaurant displayed on the 360 screen.

To build this experience, we carefully crafted sub-tasks, like asking for a table or ordering a menu item. Using the Modern Chinese textbook[2] and inputs from Chinese professors at a

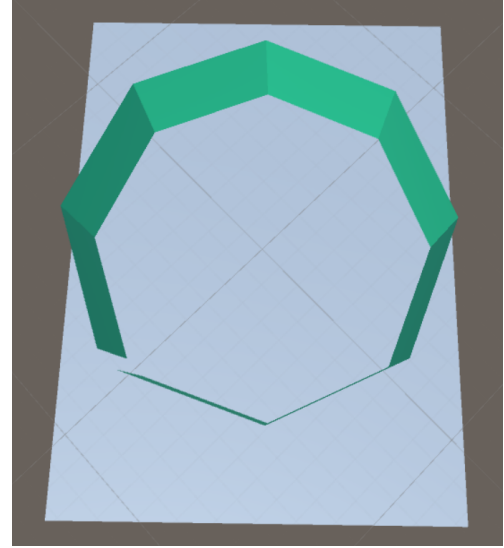


Figure 3. Computer render of the CIR display

local university, we designed dialogue inputs and responses that include vocabulary and grammar at the existing language level of the new learner. The system can accept a range of inputs that map to an intent. For example, "I want water" and "Can I get some water" map to the 'ordering' intent. For each intent that is detected by the system, the system has multiple sentence variations as possible outputs, and it picks one to randomly output. All of these possible outputs for a given intent mean the same thing. E.g. "I will get your food", "Your order will be right up", etc. as output for when food is ordered. This makes the system less repetitive and adds versatility to the dialogue. To test the pedagogical value of the system, we added a few words that are outside of the new learners' known vocabulary. Even though the goal of this literature is not to validate the pedagogical model, we mention this because it is a possible hindrance for a smooth user experience.

Based on our dialogue design, our expectation is that most of the time new learners will be able to speak fluently with the waiter using sentence structures they know without needing any help. After practicing a dining scenario in the immersive Chinese restaurant using their new language, we expect the students may be more confident going to a real Chinese restaurant and having similar conversations with a human waiter.

INTERACTION CHALLENGES

To better understand how to develop our system, students from a Chinese III class at a local university were asked for their informal feedback throughout the development process. Over the period of several months, nine students visited the CIR and interacted with the system. Every time users interacted with the system, a native Mandarin speaker and the development team was standing by to help with any issues that may have arose, identify them as interaction challenges, and translate that into a technology problem.

	Mandarin	English
AI Agent	你好！欢迎光临中餐厅！你们几位？	Hello! Welcome to the Chinese restaurant! How many people?
Customer	我们两位	We are two people
AI Agent	这边请，请问您想喝点什么？	This way please. What would you like to drink?
Customer	我想要水	Can i get some water?
AI Agent	好的马上就来。饮料怎么样？	Sure. Coming right up. How was your drink?
Customer	很好喝。	It was great!
AI Agent	你想吃点什么？这是我们的菜单。	Would you like something to eat? Here's the menu
Customer	我想要熏鱼。	I want smoked fish
AI Agent	好的。还要其他的菜吗？	Okay. Anything else?
Customer	北京烤鸭是什么？	What is the peking duck?
AI Agent	<i>(story about peking duck)</i> 你想要那个吗？	<i>(story about peking duck)</i> Do you want that?
Customer	好的。	Ok
AI Agent	好的。记下来了。	Sure. I'll put your order in
AI Agent	<i>(after a moment)</i> 菜的味道怎么样？这是您的账单，账单对吗？	<i>(after a moment)</i> Was everything okay? Here is your check, is it correct?
Customer	很好吃	Yes! it was great
AI Agent	谢谢光临	Thanks for coming to the restaurant!

Figure 4. Sample dialogue interaction. Best view in color.

The first group of students to ever test the system could not complete more than three turns of dialogue, even at their level (three) of proficiency. This was a key indicator that there were major communication gaps we needed to identify and fill. Since then, we have learned and addressed several interaction challenges and reached a level of system development where new learners are able to get through the restaurant dialogue in a reasonable time (5-10 minutes.)

Here, we describe the interaction challenges we observed through beta-tests of the system, and later, we specify the incremental enhancements made for these challenges. Although these interactions are presented in the context of a dining scenario in the immersive environment, the challenges can be generalized a good extent to other scenarios and environments where an AI teaches a foreign language through dialogue.

1. Misrepresentations in the speech-to-text engine

Mandarin is a complex language with tones and pronunciations different from the L1 of a majority of our intended users. Given its intricacies, a range of errors are common in new learners. We observed that even the slightest errors in users' pronunciations threw off the speech-to-text engine and led to a dead-end in their interactions. In addition to these errors, the text-to-speech engine expects continuity and fluency in user's speech. So, if the user's timing and fluency is not right, syllables that make up one word may be recognized as multiple words. This becomes more of an issue as it is common for words to have similar pronunciations but different meanings in Mandarin.

2. New learners are used to multi-lingual communication

In a classroom, when students learn from a teacher or role-play with their peers, they often use their L1 to help move the conversation along. A new learner may use their L1 to confirm what they understood, to correct what they meant, or to continue the conversation more comfortably. In the early stages of our environment, the system only supported Mandarin. This limited the users' abilities to complete the task if they were stuck at a stage in the dialogue. The native expert that was in the room during system tests had to give pronunciation feedback and encouragement to the students in order for them to make progress with the AI agent.

3. Agent's response was not understood

Although we use a standard Mandarin text-to-speech engine, it is possible that the speed of the AI agent's speech is too fast for new learners. It was clear during our beta-tests that the users either did not hear the waiter or did not understand the waiter. In these situations, the users struggled to move forward in the dialogue and had to ask the native expert and research team to reiterate what the waiter said. We did not want to slow down the speech of the waiter because in a real-world situation, new learners would be expected to understand sentences spoken at a standard pace.

4. Words in the dialogue are outside of the new learner's vocabulary

Words outside of the students' vocabularies were incorporated into the interactive dialogue through menu items and words spoken by the AI waiter. This was done intentionally, as mentioned earlier, and was based on the unit vocabularies from Modern Chinese textbook [2]. In the real world it is expected that the new learners will hear words they do not know. However, by using context and social interactions, the new learner might be able to understand the meaning conveyed. Yamazaki's [61] study is a successful example of this.

5. Users do not know how to respond to the waiter

We observed during the beta-tests that students may have understood what was said by the AI waiter, but they were unsure of what would be a good response. For example, during one interaction, the waiter asked for the user's order, but the user did not want to order anything else. The only way to move past this challenge was for a researcher in the room to suggest what the user could say. Overall, students get stuck in the interaction because they do not know what to say due to either their beginner-level language proficiency or not knowing what the AI waiter is capable of understanding.

6. Users do not know where they went wrong in the interaction

Especially in the beginning stages of our system, there were times when the AI agent did not respond to the users after they had spoken. This was due to the communication gap of the speech-to-text system, explained in interaction challenge 1. When this happened, students did not know what the system had heard and had no idea where the miscommunication was coming from. The system provided no feedback on what word or syllable the user mispronounced.

There was also little feedback on what the system interpreted from the user's input. The users relied on the native speakers in the room to explain their pronunciation errors.

Enhancements to Ease Interaction Challenges

This section describes help features/functions we designed and incorporated to bridge the communication gaps between the users and the system. We have been careful not to make the user interactions with the waiter too easy to the extent that there is no learning benefit. However, we still need to give the feeling of accomplishment to every student. The help functions include multi-modal interactions that feel natural to the user and the interaction at hand. Meaning, the help features are interactions that are likely to occur in the real world when there is a communication gap. We use the subsection numbers from the previous section to match the interaction challenges with the current, implemented solutions.

1. *Enabling multi-modal interactions*

Under interaction challenge 4, we discussed that the students in our beta-tests did not know how to pronounce a majority of the items on the menu. Some items themselves are popular in Chinese-American restaurants, the students did not know the Chinese names of them. Thus, in this section, we describe the two multi-modal dialogue interactions available in the room using pointing. In the real world, when one does not know how to pronounce something on the menu, they would likely point at the item on the menu and say "I want to order this" or "What is this?". This is a real-world solution to the communication gap and is common to almost all languages.

In our immersive restaurant, there is a menu on the panoramic display, as seen in fig. 2. To implement a solution to the communication gap, we enabled the room with Kinect devices and software that recognizes pointing gestures. By using this technology, we serve our aim of providing an experience without intrusive devices. Our software fuses data from multiple Kinects and multiple speech channels to interpret the menu item being referred to in the deictic speech. This is a deep research effort of its own that is not a scope of this paper. Now, users can point at an item on the menu and say "I want this" to order it.

Not only can users point at and order a dish on the menu, they can also ask for the cultural history and significance of a dish. Chinese cuisine is an important part of Chinese culture [58], and giving new learners of Mandarin the opportunity to explore the history of dishes on the menu in L1, culturally immerses them further and enhances the learning process. According to research, language and culture are inextricably linked, and Brown [29] argues that the acquisition of a second language is the acquisition of a second culture. In the case that a new learner does not know about a dish, they can point at the dish on the menu and ask "What is this?". The system then provides background information on the dish. The AI agent can also display a tone-chart to show the exact pronunciation of a dish. These two features can be utilized to introduce new words into the user's vocabulary.

The above actions, pointing and saying "I want this" or "What is this?", can alternatively be accomplished with just

verbal input. Instead of saying the command and gesturing at the menu, users would say the command including the menu item, assuming they know the correct pronunciation.

2. *Allowing the users to switch between languages*

At any point in the digital dining experience, if the user is having difficulties communicating with the AI agent, they can ask the AI agent to switch from L2 to L1. In our Chinese restaurant, this means the user can switch from Mandarin to English and vice versa. The user can carry on the dialogue for as long as they like in L1, or they can switch back to L2 at any point. Again, this explicit request for help identifies areas in the language that new learners may be struggling with. Allowing the user to switch between their L1 and L2 specifically mitigates interaction challenges 1 and 2, but is a fall-back for any communication failure in Mandarin.

3. *Asking the AI agent for clarification*

To address interaction challenges 3, 4 and 5, when a user does not grasp what was said by the AI Agent, a user can ask the AI agent, in their L2 or L1, "What did you say?" and "What am I supposed to say?". Modern voice-interface devices, such as Amazon Alexa[6] and Google Assistant[11], also include the feature where a user can ask "What did you say?". However, with these systems, the user is most often looking for the device to repeat what it just said most likely because the user did not hear it or misheard it. When the user is new to a language, repetition without any further information can be useless.

Therefore, in our immersive restaurant, when a user asks the AI agent "What did you say?" the agent repeats the phrase, gives an explanation/translation of the phrase in L1, and suggests a possible response in L2 to the user. The explanation and suggested response is manually encoded in the system based on the speech input. The spoken phrase and its suggested response are displayed on the screen in *pinyin*. This visual allows the new learner to see and hear the words simultaneously. The explicit asking for help provides insight to teachers and system developers on areas of the language that new learners are struggling with.

Although a user may understand what the agent said to them, they may not know how to respond to the agent. In this case, the user can ask "What am I supposed to say?". This triggers the same response as asking "What did you say?". The user receives the phrase repeated in L2, the phrase explained in L1, and a possible response in L2.

4. *Displaying the live transcript*

As described in interaction challenge 6, users may not know where the miscommunication is in the interaction. Therefore, a transcription of the dialogue is always shown on the panoramic display. Users can reference this to see what the system thinks they said or what the AI agent said. This provides the users immediate feedback by showing that the system is listening and what the system interpreted from the speech input. The users can use this information to determine which word, or words, was mis-interpreted (if

any) and correct it. This help feature also helps mitigate challenge 1.

5. *Designing a system-initiated dialogue*

Students may not know the dialogue capabilities of the system in the immersive restaurant, and consequently, they may not know what a reasonable response in the dialogue is. Imagine the situation where a student walks into the immersive restaurant and the AI agent says "Welcome to the restaurant." At this point, the user may be unsure of what to say next. They may wonder - do I say hello?, do I say thank you?, or do I ask for a table? In our beta-tests we saw students were hesitant at points in the dialogue because of these confusions. This interaction challenge is a general UX problem in most voice-based interfaces where the entire "menu" or "capability list" is not shown or known to the user.

In a restaurant scenario where communication is limited to a specific set of vocabulary, this challenge is easier to solve than in the general voice-bot space. We address this problem by ending all agent responses with a question that hints at what response is expected. For example, if the AI agent says "Welcome to the restaurant", it does not give an indication of what answer is expected, and the branches from this turn in the dialogue could grow wild. To avoid this, the AI agent specifically says "Welcome. How many people are you?" This question indicates the user should reply back with information containing a number. Additionally, creating a system-initiated dialogue limits the scope of the dialogue that is managed on the technological end because a user-initiated dialogue could grow into a really huge space. The system-initiated dialogue was planned in collaboration with the language teachers and the textbook [2] to elicit specific vocabulary words and sentence structures in users' responses. If the dialogue needs to be expanded, we will still follow the principles of a waiter-initiated dialogue. As an example, if multiple dining options were incorporated into the immersive restaurant, then the opening question from the AI agent might be "Welcome! Would you like a table or takeout?" Each line from the waiter ends in a question, initiating a specific response from the users, as shown in Fig. 5.

Table 1 shows a mapping of challenges and solutions numbers (from the list above) to summarize the previous two sections. Once the system was enabled with these help functions, we needed to get information about these help functions to the users and set-up their expectation of the capabilities of the system. We added the three main help functions (asking for help, asking what the waiter said, and asking what can be said next) to the display so that they are always visible to the user. The help features are shown in both English and Mandarin. Currently, Mandarin and English are the two languages enabled in our immersive restaurant; however, the interaction challenges and our enhancements to mitigate these challenges are not limited to only these languages. The interaction challenges and respective help features are universal.

Table 1. Summary of interaction challenges and solutions

Interaction Challenge	Solution
Misrepresentation in Speech-to-text	2,4
Need for multi-lingual communication	2
Agent's response not understood	2,3
Words are out of users' vocabulary	1,2,3
Users do not know how to respond	2,3,5
Users unaware of what went wrong	4

USER STUDY

We invited 16 students who were taking a Chinese I course at the same university this system is being developed at to interact with our immersive Chinese restaurant. Out of these 16 students, 12 listed English as their first language, with other first languages including Spanish, Taishanese, and Cantonese. Fifteen of the 16 students indicated having friends that speak Chinese, and four of the 16 indicated having family that speak Chinese. As mentioned earlier, the content of the immersive Chinese restaurant is based on the Modern Chinese: Simplified Characters Textbook 1A [2]. This is the textbook that the recruited Chinese I class uses. In the class prior to the study, students learned the new vocabulary and grammar structures in the food unit. The following class, instead of role-playing scenarios with peers (as is usual), the students came to the CIR to practice spoken communication with the AI agent. The professor of this Chinese I class is a member of our research team and was present during the user study to provide instruction when necessary.

Set-up of the User Study

A week before coming to the CIR to experience the system, the students were e-mailed a document explaining the immersive environment and restaurant scenario. The e-mail also included a link to a video that described how to interact with the system and how to use the help features. When the students arrived at the Cognitive and Immersive Room, they were given a demonstration of how to interact with the system by members of the research team. The demo showed the students what an entire run-through of the dialogue could look like. The students were also provided with the same document explaining the system and help features that was e-mailed to them a week before the study.

After the demonstration, the students were randomly assigned into pairs. We designed the study with student pairs to promote cooperative learning because it has been shown that learners who work together acquire more language and social skills than those who work individually [63][51]. The student pairs were randomly assigned a turn number. One at a time, each pair went to the front and center of the immersive restaurant to interact with the AI agent. Because this study ran during the Chinese I class time, each pair was given five minutes to get through as much of the dialogue as possible. This time limit was specific to this user study and ensured that all pairs of students would get a chance to interact with the immersive restaurant. The students in the pair took turns talking to the waiter and ordering their drinks and dishes.

The pair's interaction ended when they completed the dialogue and the waiter said goodbye, or it ended when the five minute time limit was up.

The professor and some native speakers were present in the room to help them with any language difficulties. All students were in the CIR for the entire study and could watch the other students interact with the system. After all the pairs had completed their turn with the immersive restaurant, they were asked to fill out a questionnaire; the questions and responses are discussed in the next section. A week following the study, a follow-up questionnaire was sent to the students to fill out online.

Results and Reflections

Questionnaire Analysis

The post-study questionnaire included questions about users' age, gender, first language, and if they have friends or family that speak Chinese. They were asked to rate the available help functions on a 5 point Likert scale of very unhelpful to very helpful. The questions asked how helpful the help functions are in the real-world and in the immersive-environment (CIR). The results of this portion of the questionnaire are shown in Fig. 5. The y-axis represents the different help features, and the x-axis shows the helpfulness score from 0-5 (very unhelpful to very helpful). The results show that the users consider all of the help functions useful in a real-world scenario. Almost all of the help functions were perceived as more useful in the real world compared to the immersive environment. We believe this is because students did not use the help features in the immersive room that often. For each of the help functions, excluding pointing, only one in four students remembered that they could use that help function. This is with the help functions demonstrated before the experiment and constantly displayed in their field-of-view throughout the study. The pointing feature was remembered by 15 out of the 16 students. Because large-scale, digital, and immersive environments like the CIR are sparse and not available to most of the general public, we assume the users have not interacted with a system like the CIR. Not only are we introducing users to the immersive environment, but we are also introducing them to the restaurant dining scenario. Therefore, we need to effectively familiarize the users to the environment and the task. Before the Chinese I students came to the CIR, they were shown a video of how the system looks and how to interact with it. They were also given an information sheet that briefly introduced the system and mentioned the help functions. This gave them information on the environment and the dialogue task. The help document was given to the students during the study to use as a reference during their interactions with the system. Interestingly, during the study we did not observe any of the students looking at their help sheet during their interaction in the immersive restaurant.

Because the user study was run during class time, we were unable to give in depth questionnaires at the time of the study. We followed up via e-mail a week after the study with a questionnaire about introducing the users to the system and what additional questions about which help features they liked. Only six subjects participated in this portion of the study, but they

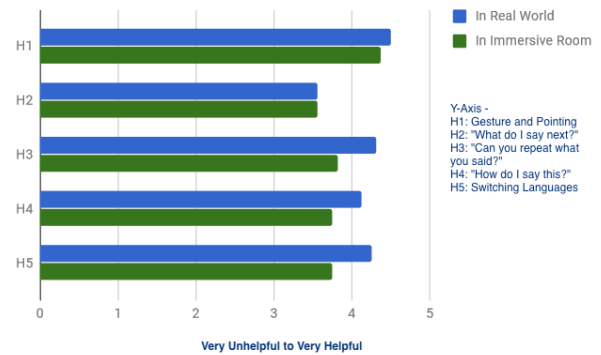


Figure 5. User ratings on helpfulness of help features in the real world vs. immersive environment

indicated that the video, the demo, and the document were helpful to them and that they liked all of the help features. Although the video, the demo, and the help document were received positively by the Chinese I students, we aim to develop more effective and efficient ways to introduce the users to the immersive restaurant. The goal is for users to need no outside help from a teacher or member of the research team to successfully communicate and interact with the system.

Evaluation of the Help Functions - Researcher Observations

The help feature that was used most frequently was the pointing gesture. Many times when a student was struggling with the pronunciation of a menu item, they would instead point to the item and order it. The students resorted to using this interactive help function on their own. Otherwise, the professor had to encourage students to try some of the help functions. For example, at one point in the dialogue a student asked the professor what to do. Instead of telling them, the professor told the student to ask the system what to say next.

Students seemed to rely on help from their professor, classmates, and members of the research team when they were stuck, instead of the help functions enabled in the system. The professor and members of the research team provided help even when the students did not explicitly ask for it. For example, a majority of the users had to repeat themselves multiple times in order for the system to pick up the correct interpretation of their pronunciation. In these instances, the professor and members of the research team would chime in with the correct pronunciation or would encourage the student to keep trying. This is typical to any language learning classroom and the native Mandarin speakers were able to guide the students on their timing and pronunciations. Our goal is to enable the AI agent to provide the same help that the professor and native speakers did. This gives us additional motivation to integrate a more pro-active help system into the immersive restaurant.

Using video recordings of the study, we observed the help that the professor and native speakers provided to the students. From these observations we can further develop our help features. Every student was given help at some point in their interaction. The different kinds of help given were classified as the following:

- Encouraging the student to try the phrase again

- Suggesting pronunciations of words and phrases
- Suggesting the student try a different word or menu item (e.g. instead of Panda say Waiter)
- Suggesting the correct tones for the pronunciation including hand gestures of the tones
- Explaining the system bugs (usually related to timing of speech)
- Confirming they understood what the student said, although the system did not
- Repeating what the waiter said or hinting to the student what the waiter was expecting them to say

These observations motivate us to enhance the ability of the AI agent by teaching it to observe when a user needs help and what help to provide or suggest.

Other Observations

Further observations of video footage from the user study, showed the students consistently looking at the transcript to get feedback on their pronunciations and to understand the stage of interaction. We noted in our communication with some students that they found the displayed transcript of the waiter's response and their own response useful. In the future, we want to explore the pedagogical relevance of this. The students might have understood the written Chinese better than the AI agent's spoken response, or they might have used the transcript to confirm what they heard. Either way, from an interaction perspective, it made things easier for the user. However, a live transcript is not something that is available to a new learner in the real world. We would like the students to use the system without it, but in a student's initial interactions with the CIR, having the transcript available could have a positive impact or a moral boost. In future CIR developments, it may be possible for the transcript to be turned on or off by the user in order to set the difficulty level of the system.

We observed that most students used the shortest path to complete the dialogue and each step within the dialogue. Students only ordered one drink and one dish each, they always said the food was good without any additional remarks, and they did not, in general, explore dialogue paths not demonstrated to them (e.g. asking for a dish recommendation).

The goal of the immersive Chinese restaurant is to encourage students to explore and have detailed conversations in Mandarin. However, we believe that even with a sufficient demonstration of how complex the dialogue can be, students would likely take the shortest path to complete it. A further enhancement of the system is to utilize the waiter-initiated dialogue design to encourage students to expand the conversation. For example, if a user says that the food was good, the waiter could say "Did you like the spicy dish or the sweet one more?". There are many questions the waiter can ask throughout the dining experience that will challenge the new learners to expand their Mandarin skills.

We also noted that the pairs of students who interacted with the system later in the class got further in the dialogue than the pairs that went first. For example, the last two pairs of students made it farther in the dialogue than any other pair

before them, and the last pair was the only pair to complete the entire dialogue in the five minute limit. This suggests that being in the CIR and watching other students had positive effects on student performance in the CIR.

Encouragements

According to the class textbook [2] and information from the professor of the class, upon arriving to the study students had not yet learned how to ask or pay for the check in Mandarin. In the open ended portion of the questionnaire, one student wrote that after their time in the CIR they knew how to pay the bill. This is an encouraging feat for the research team. Although the students were receiving some guidance from the professor and other Mandarin speakers in the room, a student learned something new through their experience with the Cognitive and Immersive Room. Another student wrote that our system is 'a good proof-of-concept', and they 'hope to see more technology like this integrated into education'.

CONCLUSION AND FUTURE WORK

In this work, we have motivated the need for an immersive room where users can practice spoken language through real-life task completion exercises. We have presented the Cognitive and Immersive Room developed as an immersive Chinese restaurant by collaborating with experts from a variety of domains (e.g. computer science, HCI, foreign language education, graphic design, game development, software development, etc.) The immersive Chinese restaurant is an environment where new learners of Mandarin can role-play with a virtual waiter. We evaluated interaction challenges of new learners and new CIR users through beta tests. These challenges motivated the development of multiple help features, including pointing at an on-screen menu item to order it. A formal user study showed the effectiveness and likability of the features developed to mitigate the interaction challenges, and the study also shed light on further developments.

Our goal is to have an immersive environment where a user can interact with the AI agent as though it were a real human. We are constantly developing new technologies in the room that we think would be helpful in bridging the communication gaps. In the future steps, the AI agent will be able to provide productive feedback to the user and help them improve their L2 proficiency. This includes integrating a pitch contour analysis graph that will show the users a plot of their tones in real time. We also plan to enable the immersive restaurant with a society of agents playing different roles (e.g. customers, chefs, etc.). Once we have confidently tackled the restaurant scenario, we hope to enable multiple scenarios in the CIR. In this direction, we hope to find and solve new interaction challenges that are not yet foreseen.

ACKNOWLEDGMENT

This work is supported by the Cognitive and Immersive Systems Laboratory (CISL), a collaboration between IBM Research and Rensselaer Polytechnic Institute, and also a center in IBM's AI Horizon Network. The authors would like to thank all of the researchers at CISL and the colleagues at IBM Yorktown Heights and IBM China for their contributions to this research. Without their help, this project would not have been possible.

REFERENCES

1. 2009. *Computer Assisted Language Learning: Critical Concepts in Linguistics*. Vol. I. Routledge.
2. 2013. *Modern Chinese Textbook 1A*. Better Chinese.
3. 2017. Effective Language Learning. (Oct 2017). <http://www.effectivelanguagelearning.com/language-guide/language-difficulty>
4. 2018. (2018). <http://www.activeworlds.com/>
5. 2018. (2018). <http://secondlife.com/>
6. 2018. Amazon Alexa. (2018). <https://developer.amazon.com/alexa>
7. 2018. Busuu. (2018). <https://www.busuu.com>
8. 2018. CAVE Automatic Virtual Environment. (2018). <http://www.visbox.com/products/cave/>
9. 2018. Chinese Pod. (2018). <https://chinesepod.com>
10. 2018. eChineseLearning. (2018). <https://www.echineselearning.com>
11. 2018. Google Assistant. (2018). <https://assistant.google.com>
12. 2018. Hello Talk. (2018). <http://home.hellotalk.com>
13. 2018. Mandarin (Chinese) | About World Languages. (2018). <http://aboutworldlanguages.com/mandarin>
14. 2018. Smallab Learning. (2018). <https://www.smallablearning.com/>
15. 2018. Tutor Mandarin. (2018). <https://www.tutormandarin.net/en>
16. 2018. Wai Chinese. (2018). <https://waichinese.com/features.html>
17. 2018. Watson Assistant. (2018). <https://www.ibm.com/watson/services/conversation/>
18. 2018. Watson Speech To Text. (2018). <https://www.ibm.com/watson/services/speech-to-text/>
19. Muna Mohammed Abbas Alkhateeb. 2017. An Analytical Approach of the Cultural Context in Language Teaching and Learning. *Journal of University of Babylon* 25, 2 (2017), 832–840. <https://doaj.org/article/d1e436f141a6483b832b2895c02dd40e?>
20. Anke Berns, Antonio Gonzalez-Pardo, and David Camacho. 2013. Game-like language learning in 3-D virtual environments. 60 (01 2013), 210–220.
21. David Birchfield and Colleen Megowan-Romanowicz. 2009. Earth science learning in SMALLab: A design experiment for mixed reality. *International Journal of Computer-Supported Collaborative Learning* 4, 4 (01 Dec 2009), 403–421. DOI: <http://dx.doi.org/10.1007/s11412-009-9074-8>
22. Lynne Bolduc. 1997. Mastering Difficult Concepts Using Software to Support Learning: A Case Study in Teaching Mandarin Chinese. *Journal of Science Education and Technology* 6, 2 (1997), 91–102.
23. Richard Brecht, Dan Davidson, and Ralph Ginsberg. 1995. Predictors of foreign language gain during study abroad. *Second language acquisition in a study abroad context* 9 (1995), 37.
24. Silvia Canto and Kristi Jauregi Ondarra. 2017. Language learning effects through the integration of synchronous online communication: The case of video communication and Second Life. *Language Learning in Higher Education* 7, 1 (2017), 21–53.
25. Dongdong Chen. 2010. Enhancing the learning of Chinese with Second Life. *Journal of Technology and Chinese Language Teaching* 1, 1 (2010), 14–30.
26. Nancy F Chen, Darren Wee, Rong Tong, Bin Ma, and Haizhou Li. 2016. Large-scale characterization of non-native Mandarin Chinese spoken by speakers of European origin: Analysis on iCALL. *Speech Communication* 84 (2016), 46–56.
27. Tina Cok. 2016. ICT-supported language learning tools for Chinese as a foreign Language: a content review/Jezikovnotehnoloska orodja za učenje kitajščine kot tujega jezika: pregled vsebine. *Revija za Elementarno Izobraževanje* 9, 3 (06 2016), 103–120. <http://libproxy.rpi.edu/login?url=https://search.proquest.com/docview/1799383943?accountid=28525> Copyright - Copyright Univerza v Mariboru, Faculty of Education/Pedagoska Fakulteta Jun 2016; Document feature - Tables; ; Graphs; Last updated - 2016-06-25.
28. Dan E Davidson. 2010. Study abroad: When, how long, and with what results? New data from the Russian front. *Foreign Language Annals* 43, 1 (2010), 6–26.
29. Oxana Dema and Aleidine Kramer Moeller. 2012. Teaching culture in the 21st century language classroom. *Faculty Publications: Department of Teaching, Learning and Teacher Education* (2012), 75–91.
30. Mats Deutschmann, Luisa Panichi, and Judith Molka-Danielsen. 2009. Designing oral participation in Second Life—a comparative study of two language proficiency courses. *ReCALL* 21, 2 (2009), 206–226.
31. Wenhao Diao. 2011. Study abroad, participation and turn taking: A case study. In *Selected Proceedings of the 2010 Second Language Research Forum. Cascadilla Proceedings*.
32. Rahul R. Divekar, Robert Rouhani, Matthew Peveler, Rui Zhao, Jeffrey O. Kephart, David Allen, Kang Wang, Qiang Ji, and Hui Su. CIRA - An Architecture for Building Configurable Immersive Smart-rooms. In *To appear in proceedings of Intellisys 2018*.
33. Hang Du. 2013. The development of Chinese fluency during study abroad in China. *The Modern Language Journal* 97, 1 (2013), 131–143.
34. Michael E Everson. 2011. Best practices in teaching logographic and non-Roman writing systems to L2 learners. *Annual Review of Applied Linguistics* 31 (2011), 249–274.

35. Serhat Güzel and Selami Aydin. 2016. The Effect of Second Life on Speaking Achievement. *Online Submission* 6, 4 (2016), 236–245.
36. Michael Henderson, Hui Huang, Scott Grant, and Lyn Henderson. 2009. Language acquisition in Second Life: Improving self-efficacy beliefs. (2009).
37. Murat Hismanoglu. 2012. Integrating second life into an EFL classroom: A new dimension in foreign language learning and teaching. 3 (01 2012), 100–111.
38. Yu-Ju Lan, Yu-Hsuan Kan, Indy Y.T. Hsiao, Stephen J.H. Yang, and Kuo-En Chang. 2013. Designing interaction tasks in Second Life for Chinese as a foreign language learners: A preliminary exploration. *Australasian Journal of Educational Technology* 29, 2 (2013). DOI : <http://dx.doi.org/10.14742/ajet.144>
39. Mark Lee and Barney Dalgarno. 2011. *Scaffolding Discovery Learning in 3D Virtual Environments: Challenges and Considerations for Instructional Design*. IGI Global, United States, 138–169. DOI : <http://dx.doi.org/10.4018/978-1-61692-825-4.ch008> Imported on 12 May 2017 - DigiTool details were: publisher = Hershey, PA: IGI-Global, 2011. editor/s (773b) = Shalin Hai-Jew; Issue no. (773s) = 8; Parent title (773t) = Virtual immersive and 3D learning spaces: emerging technologies and trends.
40. Hsien-Cheng Liao, Ying-Hua Guan, Jia-Jang Tu, and Jiang-Chun Chen. 2014. A prototype of an adaptive Chinese pronunciation training system. *System* 45 (2014), 52–66.
41. Yuen-kuang Liao and Yu-wen Chen. 2007. The effect of computer simulation instruction on student learning: A meta-analysis of studies in Taiwan. *Journal of information Technology and Applications* 2, 2 (2007), 69–79.
42. Tsun-Ju Lin and Yu-Ju Lan. 2015. Language Learning in Virtual Reality Environments: Past, Present, and Future. *Educational Technology & Society* 18, 4 (2015), 486–497.
43. James Milton, Sunniva Jonsen, Steven Hirst, and Sharn Lindenburn. 2012. Foreign language vocabulary development through activities in an online 3D environment. *The Language Learning Journal* 40, 1 (2012), 99–112. DOI : <http://dx.doi.org/10.1080/09571736.2012.658229>
44. Corey Mitchell. 2016. One million U.S. students could be studying Mandarin by 2020. (Aug 2016). <http://www.pbs.org/newshour/rundown/2020-u-s-aims-1-million-students-studying-mandarin/>
45. Max North and Sarah North. 2016. A Comparative Study of Sense of Presence of Traditional Virtual Reality and Immersive Environments. *Australasian Journal of Information Systems* 20, 0 (2016). DOI : <http://dx.doi.org/10.3127/ajis.v20i0.1168>
46. Mary G. O'Brien, Richard Levy, and Annika Orich. 2009. Virtual Immersion: The Role of CAVE and PC Technology. *CALICO Journal* 26, 2 (01 2009), 337. <http://libproxy.rpi.edu/login?url=https://search-proquest-com.libproxy.rpi.edu/docview/750619686?accountid=28525> Copyright - Copyright CALICO Jan 2009; Last updated - 2014-12-07.
47. Mark Peterson. 2011. Towards a Research Agenda for the Use of Three-Dimensional Virtual Worlds in Language Learning. *CALICO Journal* 29, 1 (2011), 67–80. <http://www.jstor.org/stable/calicojournal.29.1.67>
48. Thomas C Reeves and John G Hedberg. 2003. *Chapter 1: A Pragmatic Rationale for Evaluation*. Education Technology, 7–9.
49. Wei Ren. 2013. The effect of the study abroad on the pragmatic development in internal modification of refusals. 23 (12 2013), 715–741.
50. Roger C. Schank. 1990. Case-Based Teaching: Four Experiences in Educational Software Design. *Interactive Learning Environments* 1 (1990), 231–253.
51. Helen H Shen and Wenjing Xu. 2015. Active learning: Qualitative inquiries into vocabulary instruction in Chinese L2 classrooms. *Foreign Language Annals* 48, 1 (2015), 82–99.
52. Ya-Chun Shih. 2007. The Development of an Online Virtual English Classroom: VEC3D.
53. Ya-Chun Shih. 2017. *Weaving Web 2.0 and Facial Expression Recognition Into the 3D Virtual English Classroom*. Vol. I-III. IGI Global, 1375–1399.
54. Naoko Taguchi, Qiong Li, and Xiaofei Tang. 2017. Learning Chinese Formulaic Expressions in a Scenario-Based Interactive Environment. 50 (10 2017).
55. Etsuko Toyoda. 2002. Categorization of Text Chat Communication between Learners and Native Speakers of Japanese.
56. Jennifer J. Vogel, David S. Vogel, Jan Cannon-Bowers, Clint A. Bowers, Kathryn Muse, and Michelle Wright. 2006. Computer Gaming and Interactive Simulations for Learning: A Meta-Analysis. *Journal of Educational Computing Research* 34, 3 (2006), 229–243. DOI : <http://dx.doi.org/10.2190/FLHV-K4WA-WPVQ-HOYM>
57. Airong Wang. 2017. *Learning English in a Multi-User Virtual Environment: Exploring Factors Affecting Participation*. Ph.D. Dissertation. Mid Sweden University.
58. Endymion Porter. Wilkinson. 2015. *Chinese History: A New Manual* (4 ed.). Harvard University Asia Center for the Harvard-Yenching Institute distributed by Harvard University Press.
59. Clare Wright and Z Cong. 2014. Examining the effects of Study Abroad on L2 Chinese development among UK university learners. *Newcastle and Northumbria Working Papers in Linguistics* 20 (2014), 67–83.

60. Zhou Yalun. 2016. Digital Vocabulary Competition as Motivator for Learning in CFL Classrooms. *JOURNAL OF TECHNOLOGY AND CHINESE LANGUAGE TEACHING* 7, 2 (2016), 1–22.
61. Kasumi Yamazaki. 2018. Computer-assisted learning of communication (CALC): A case study of Japanese learning in a 3D virtual world. *ReCALL* (2018), 1–18. DOI : <http://dx.doi.org/10.1017/S0958344017000350>
62. Rui Zhao, Kang Wang, Rahul Divekar, Robert Rouhani, Hui Su, and Qiang Ji. 2018. An Immersive System with Multi-modal Human-computer Interaction. In *The IEEE Conference on Automatic Face and Gesture Recognition (FG)*.
63. Wenjing Zuo. 2011. The Effects of Cooperative Learning on Improving College Students’s Reading Comprehension. *Theory and Practice in Language Studies* 1, 8 (Jan 2011). DOI : <http://dx.doi.org/10.4304/tpls.1.8.986-989>