The Rensselaer Mandarin Project—a Cognitive and Immersive Language Learning Environment

David Allen,[1] Rahul R. Divekar,[1] Jaimie Drozdal,[1] Lilit Balagyozyan,[1]
Shuyue Zheng,[1] Ziyi Song,[1] Huang Zou,[1] Jeramey Tyler,[1] Xiangyang Mou,[1] Rui Zhao,[1]
Helen Zhou,[1] Jianling Yue,[1] Jeffrey O. Kephart,[2] and **Hui Su[1,2]**
[1] Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180 USA
[2] IBM Thomas J Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598 USA
{allend5, divekr, drozdj2, balagl, zhengs6, songz2, zouh, tylerj2, moux4,
zhouy12, zhaor, yuej2, suh5}@rpi.edu, {kephart, huisuibmres}@us.ibm.com

Abstract

The Rensselaer Mandarin Project enables a group of foreign language students to improve functional understanding, pronunciation and vocabulary in Mandarin Chinese through authentic speaking situations in a virtual visit to China. Students use speech, gestures, and combinations thereof to navigate an immersive, mixed reality, stylized realism game experience through interaction with AI agents, immersive technologies, and game mechanics. The environment was developed in a black box theater equipped with a human-scale 360° panoramic screen ($14'h$, $20'r$), arrays of markerless motion tracking sensors, and speakers for spatial audio.

Introduction

There are many reasons to learn a foreign language. The ability to communicate in multiple languages can lead to both a deeper understanding of cultural differences throughout our world, and the ability to thrive in the global economy. On that cornerstone, the Rensselear Mandarin Project (RMP) leverages intelligent and immersive systems to facilitate the rapid acquisition of second language skills, and to give students a competitive edge in a world with increasing lingual and cultural diversity.

Work in the field suggests that interactive environments provide significant benefits to learning a second language when compared to traditional teaching methods (Canto and Ondarra 2017; Güzel and Aydin 2016). Interactivity improves learning outcomes and facilitates positive learning attitudes in the classroom (Lin and Lan 2015). It facilitates better understanding and retention of formulaic expressions (Taguchi, Li, and Tang 2017), persuasive talk, awareness of audience, and collaborative communication (Yamazaki 2018) through digital immersion in authentic speaking situations (Čok 2016). At the same time, most existing technologies utilize PC-sized displays, and seldom leverage Artificial Intelligence (AI) (Divekar et al. 2018c).

Against that backdrop, RMP presents a human scale, immersive language learning environment for an experience that revolves around intelligent interactions with AI Agents. Additionally, RMP leverages gamification to engage second language learners in an environment where they can "self-govern" their experience, offering them activity, autonomy, and interaction (Dubbels 2011) with other students and human-scale virtual agents simultaneously.

An initial usability study was conducted in November, 2017 with an earlier prototype. In that study, researchers collected data from 16 students: 8 male, 8 female, Ages 18-22; first languages: 1 Taishanese, 1 Cantonese, 2 Spanish, and 12 English. A questionnaire given during the study yielded the following key qualitative findings: 16/16 students liked being able to use gestures (pointing) to disambiguate food menu items and found it helpful; 14/16 students liked that the system was able to repeat itself on command and found it helpful. Complete findings from the usability study appear in Divekar et al. (2018a).

Technical Details

The demo integrates AI, immersive systems, and gamification for intelligent interactions (e.g. experience through multi-modal dialogue) which emerge from a network of granular AI tasks. We define "agent" as components that can do one or more tasks (e.g. transcribing speech to text). Some of these are commercially available solutions while others are developed internally. The following text gives a summary of the core technologies utilized.

- **Distributed architecture.** The application EXECUTOR handles all system responses at the top-most architectural level. As input, it ingests data from multiple message queue streams. On the most granular level, an array of system agents populate streams with data packets of formatted input from individual sensors. Mid-layer agents then digest those raw data streams to provide higher level inferred data packets for the EXECUTOR. Agents are distributed over both an array of physical devices within the language learning environment and online through external web services.
- **Intelligent systems.** The foremost data stream of RMP is the speech pipeline. Here, student utterances are recorded, transcribed, and labeled with intents detected from the transcribed text. The gesture stream, enabled by skeletal tracking devices and custom gesture recognition software,

provides input about what gestures are made by users—i.e. pointing (Zhao et al. 2018).

- **Multimodal inference.** Agents inform the EXECUTOR through multiple modalities and inferred data streams:
  1. Recorded audio
  2. Transcribed speech from audio
  3. Detected gestures from skeletal tracking
  4. Context inferred from combinations thereof

  These inputs enable system-level decisions based on both individual modalities and contextually inferred data from combined modalities. Individual modality interactions include pitch analysis of utterances [1], verbal conversations with virtual agents [2], and gesture recognition [3]—pointing at objects in the world to select them and continuous detection of movement sequences during tai chi exercises. Interactions inferred from combined modalities include the interpretation of deictic utterances like "take me over there" [2] combined with pointing gesture recognition [3] to teleport students to the inferred location [4], and pointing [3] at a menu item to specify "I want this" [2] when ordering food [4].
- **Multimodal presentation.** Synthesized speech, ambient audio, SFX, and immersive game visuals present system responses on the front end to complete a multimodal communication loop. A detailed description of how these systems facilitate disambiguation and communication through the use of multimodal input and output appears in Divekar et al. (2018b).

Immersive language learning

- **Multi-round dialogue.** Students engage with AI agents in multiple contexts—restaurant dining, navigating street market purchases, or practicing tai chi—returning to agents for multiple rounds of unique dialogue.
- **Educational outcome.** The demo provides an engaging simulation of international language immersion with authentic dialogue opportunities to enable students to expand their vocabulary and practice their pronunciation of Mandarin Chinese.

Gamification

- **Role play.** The demo presents a role-playing scenario for a student on a trip to China through 2 scavenger hunt activities, 3 navigable scenes, 8 embodied agents with approximately 78 turns of multi-modal dialogue.
- **Modular characters.** AI agents instance a modular character rig with options for 4 hairstyles, 3 upper outfits, 3 lower outfits, and 3 accessories for a total of 108 possible unique character models. Additional options (e.g. colors), enabled the rapid creation of multiple unique avatars.
- **Performance capture.** Animations were recorded with inertial measurement unit (IMU)–based technology.

Conclusion

Human-scale immersion and intelligent interactions with multiple AI agents place the Rensselaer Mandarin Project at the forefront of advanced research in cognitive and immersive classrooms for cyber-enabled language learning. In preparation for use in a 6-week classroom experience during Summer 2019, plans for future development of this technology include: support for AI-enabled individual student learning assistant, student profiles, many-to-many–student-to-agent interactions, additional environments, lip-synced characters, and Hololens integration.

Acknowledgments

Game components were developed by a team of RPI student researchers during Summer 2018. This research was supported by Cognitive and Immersive Systems Laboratory—a research collaboration between Rensselaer Polytechnic Institute and IBM through IBM's AI Horizon Network.

References

Canto, S., and Ondarra, K. J. 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):21–53.

Čok, T. 2016. ICT-supported language learning tools for Chinese as a foreign language: a content review. *Revija za Elementarno Izobraževanje* 9(3):103–120.

Divekar, R. R.; Drozdal, J.; et al. 2018a. Interaction challenges in AI equipped environments built to teach foreign languages through dialogue and task-completion. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, 597–609. New York, NY, USA: ACM.

Divekar, R. R.; Peveler, M.; et al. 2018b. CIRA—an architecture for building configurable immersive smart-rooms. In *Proceedings of SAI Intelligent Systems Conference*, 76–95. Springer.

Divekar, R. R.; Zhou, Y.; et al. 2018c. Building human-scale intelligent immersive spaces for foreign language learning. *iLRN 2018 Montana* 94.

Dubbels, B. 2011. Designing learning activities for sustained engagement: Four social learning theories coded and folded into principals for instructional design through phenomenological interview and discourse analysis. In *Discoveries in gaming and computer-mediated simulations: New interdisciplinary applications*. IGI Global. 189–216.

Güzel, S., and Aydin, S. 2016. The effect of Second Life on speaking achievement. *Online Submission* 6(4):236–245.

Lin, T.-J., and Lan, Y.-J. 2015. Language learning in virtual reality environments: Past, present, and future. *Journal of Educational Technology & Society* 18(4):486–497.

Taguchi, N.; Li, Q.; and Tang, X. 2017. Learning Chinese formulaic expressions in a scenario-based interactive environment. *Foreign Language Annals* 50(4):641–660.

Yamazaki, K. 2018. Computer-assisted learning of communication (calc): A case study of japanese learning in a 3d virtual world. *ReCALL* 30(2):214–231.

Zhao, R.; Wang, K.; et al. 2018. An immersive system with multi-modal human-computer interaction. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, 517–524. IEEE.