Situated Learning of Soft Skills with an Interactive Agent in Virtual Reality via Multimodal Feedback

Zhenjie Zhao
The Hong Kong University of Science and Technology
Hong Kong, China
zzhaoao@cse.ust.hk

Xiaojuan Ma
The Hong Kong University of Science and Technology
Hong Kong, China
mxj@cse.ust.hk

ABSTRACT
Nowadays, customer virtual reality (VR) devices can offer high-quality presence experiences in a relatively low cost, which opens opportunities for popularizing situated learning. We investigate situated learning of soft skills, in particular, elevator pitches, through simulating an interactive agent with multimodal feedback in VR. A virtual agent acts as a coach and offers both verbal and nonverbal feedback to users. We developed a prototype system and conducted a quantitative experiment to investigate the potential of situated learning in VR, especially on how users perceive the virtual agent. We summarize design considerations on developing interactive agents with multimodal feedback in VR to facilitate situated learning of soft skills.

Author Keywords
Situated learning; virtual reality; interactive agents; multimodal feedback; soft skills; quantitative study.

CCS Concepts
•**Human-centered computing** → *Virtual reality;*

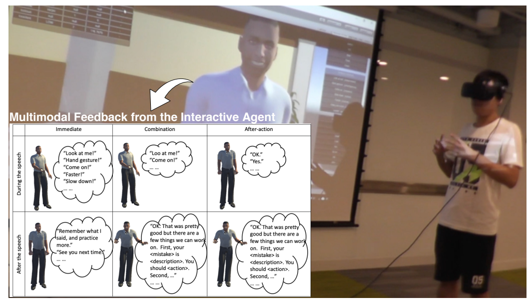

Figure 1. A participant wears a headset to practice elevator pitches with a virtual agent in VR. The virtual agent provides multimodal feedback based on pre-set strategies: immediate, after-action and combination.

INTRODUCTION AND RELATED WORK
The game and smart phone industries make high-quality presence experiences in virtual reality (VR) available with a relatively low cost. We can render realistic scenes and simulate real-world activities in VR, which can potentially help popularize situated learning experiences.

Situated learning emphasizes that knowledge should be learned in a proper context and is dependent on the learning situation [4]. Compared with traditional instructional learning, situated learning places context and how to use knowledge under a specific context in a more important role.

Although previous research works have incorporated situated learning in various domains such as medical training [6], literature education [10], public speech [7] and job interview training [2], they either did not explicitly use VR as a tool, which lacks the presence feeling brought about by VR, or did not consider an interactive agent that simulates real-humans' verbal and nonverbal behaviors, which plays an important role on how we learn soft skills.

We investigate interactive agents with multimodal feedback in VR for situated learning of soft skills. In particular, we use elevator pitches for academical purposes as a case study. An elevator pitch, *i.e.*, delivering a quick persuasive speech to arouse interest in an idea, is an important soft skill that can potentially bring opportunities for jobs and collaborations [3]. We here consider specifically in academical situations, such as introducing personal research interests and projects in a conference, which are important for students to get prepared to advertise their research.

We first conducted semi-structured interviews with domain experts and observed real elevator pitches at various occasions to derive key elements of a good elevator pitch and the general strategies to give learners feedback. These findings guided the design of an interactive agent with multimodal feedback in VR. We conducted a user study with 40 participants on this proof-of-concept system to assess how users perceive the virtual agent for situated learning of elevator pitch skills. Results show that with proper mutlimodal feedback from a virtual agent in VR, users can have a strong sense of being coached. We also summarize several design considerations on developing virtual agents to facilitate situated learning in VR. More information can be found in the project homepage [1].

[1] https://zhaozj89.github.io/InteractiveAgent4SituatedLearning

UIST'20 Adjunct, October 20–23, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7515-3/20/10.
DOI: https://doi.org/10.1145/3379350.3416174

DESIGN PROCESS

We followed a design research process [1], which involves performing initial exploratory qualitative data collection by conducting semi-structure interviews with experts and observing real elevator pitches. Informed by these insights, we investigated how to develop and test the VR system.

Quantify the Performance of an Elevator Pitch

To derive codes for quantifying the performance of an elevator pitch, we interviewed four domain experts from local language and career centers and videotaped 30 videos of typical elevator pitches with transcripts. We quantify the performance of an elevator pitch in a way that it can be measured by existing sensors. Although there may be some kinds of mistakes in an elevator pitch that cannot be identified by our system, the types of issues covered by our system are generally enough for the purpose of exploring users' perception on multimodal feedback from virtual agents in VR. We first discussed as many as possible existing off-the-shelf sensors and then derived specific parameters by analyzing the recorded videos and transcripts. We also verified proposals from existing literatures [9, 2, 8, 5]. After several regular group meetings and discussions, we summarized five main metrics, including eye contact measured by an inertial measurement unit, hand gesture measured by a Kinect, speech rhythm measured by Google speech recognizer, speech volume measured by a microphone, and timing measured by a time clock.

System Design

The system consists of four components: sensing, ranking, feedback, and VR display. With the sensing module, we first detect the learner's mistakes during elevator pitch practice, and then push these mistakes into a queue asynchronously. Afterwards, the ranking module selects a highest priority mistake and sends to the feedback strategy module. With a pre-set feedback strategy and the mistake, we generate corresponding verbal and nonverbal behaviors, and display them in VR. We further design three strategies of giving feedbacks along the dimension of timing. For the immediate strategy, during the speech, it generates feedback whenever users make a mistake. For the after-action strategy, during the speech, it simulates the normal reactions of a person, and only gives feedback after users finish their presentation. The combination strategy hybridizes the immediate and after-action strategies. During the speech, it only gives feedback of main mistakes and summarizes minor ones after the speech. We implemented the system with Unity game engine.

USER EVALUATION

We conducted a between-subject user study to explore users' perception on situated learning of soft skills with an interactive agent with multimodal feedback in VR. We invited each participant to use our VR system in one of the four feedback conditions: immediate, after-action, combination, and no learning feedback (baseline). A snapshot of the system in use is shown in Figure 1.

We recruited 40 volunteers (17 females, average age 22.8, SD: 2.2) from a local university. We asked participants to deliver an elevator pitch on an academical topic to the virtual agent. Each participant repeated the practice three times. Upon the completion of all three rounds, participants filled out a questionnaire to report if they can feel feedback during speech, help from the virtual agent, being coached, performance improvement, and adaptive behaviors of the virtual agent, followed by an in-depth exit interview.

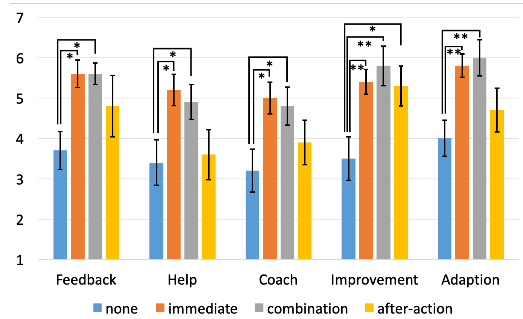

Figure 2. Means and standard errors of feelings on feedback during speech (feedback), help from the virtual agent (help), coaching by the virtual agent (coach), improvement (improvement), and adaptive behaviors of the virtual coach (adaptation) ($*: p < .05$, $**: p < .01$).

Results and Design Considerations

The main result is shown in Figure 2. Generally, we find significant effects on the five measurements. Post-hoc test further shows that the immediate and combination feedback strategies have significantly higher scores than the baseline on the five measurements. We summarize some design considerations:

Provide feedback that matches users' mental model. Otherwise, it may result in negative perception of the virtual agent.

Adapt to users' professional levels. The best experience comes from users being continuously challenged.

Conversational agents that can understand users' utterances and give reasonable feedback are necessary. Most users expect an intelligent agent that can talk like a real human coach.

Provide feedback in the right time. Giving feedback in a wrong time, either too early or too late, will make users feel the virtual agent artificial.

CONCLUSION

Through a design process, we investigated interactive agents with multimodal feedback in VR for situated learning of elevator pitch skills. We designed and developed a proof-of-concept system, and conducted a between-subject user study. Results show that feedback from an interactive agent with multimodal feedback in VR improves users' perception of the virtual agent. In the future, we plan to experiment with a fully functional system built upon insights from this on-going work.

ACKNOWLEDGMENTS

The authors thank Wenjie Yang for helping coordinating the user study, as well as Prof. Huaishu Peng and all reviewers for insightful comments. This work is supported by the Research Grants Council of the Hong Kong Special Administrative Region, China under Grant No.: T44-70716-N.

REFERENCES

[1] Matthew W. Easterday, Daniel G. Rees Lewis, and Elizabeth M. Gerber. 2018. The logic of design research. *Learning: Research and Practice* 4, 2 (2018), 131–160. DOI:http://dx.doi.org/10.1080/23735082.2017.1286367

[2] Mohammed (Ehsan) Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W. Picard. 2013. MACH: My Automated Conversation Coach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 697–706. DOI: http://dx.doi.org/10.1145/2493432.2493502

[3] Richard House, Jessica Livingston, Sarah Summers, and Anneliese Watt. 2016. Elevator Pitches, Crowdfunding, and the Rhetorical Politics of Entrepreneurship. In *2016 IEEE International Professional Communication Conference (IPCC)*. 1–4. DOI: http://dx.doi.org/10.1109/IPCC.2016.7740526

[4] Ann Kovalchick and Kara Dawson. 2003. Education and Technology: An Encyclopedia. *ABC-CLIO* (2003).

[5] Elizabeth Kuhnke. 2012. *Body language for dummies*. John Wiley & Sons.

[6] Huang Hsiu Mei and Liaw Shu Sheng. 2011. Applying Situated Learning in a Virtual Reality System to Enhance Learning Motivation. *International journal of information and education technology* 1, 4 (2011), 298–302.

[7] Jan Schneider, Dirk Börner, Peter van Rosmalen, and Marcus Specht. 2015. Presentation Trainer, Your Public Speaking Multimodal Coach. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*. ACM, New York, NY, USA, 539–546. DOI: http://dx.doi.org/10.1145/2818346.2830603

[8] Stephen M. Smith and David R. Shaffer. 1995. Speed of Speech and Persuasion: Evidence for Multiple Effects. *Personality and Social Psychology Bulletin* 21, 10 (1995), 1051–1060. DOI: http://dx.doi.org/10.1177/01461672952110006

[9] WikiHow Online Document. Retrieved September 19, 2019. How to Measure Decibels. (Retrieved September 19, 2019). https://www.wikihow.com/Measure-Decibels

[10] Rasimah Che Mohd Yusoff, Halimah Badioze Zaman, and Azlina Ahmad. 2010. Design a Situated Learning Environment using Mixed Reality Technology–A Case Study. *World Academy of Science, Engineering and Technology* 47 (2010), 887–892.