Designing Virtual Pedagogical Agents and Mentors for Extended Reality

Tiffany D. Do*

University of Central Florida

Abstract

The use of virtual and augmented reality for educational purposes has seen a rapid increase in interest in recent years. Extended reality offers unique affordances to learners, and can enhance learning. Specifically, we are interested in the use of pedagogical agents in extended reality due to their potential to increase student motivation and learning. However, the design of pedagogical agents in extended reality is still a nascent area of study, which can be important in an immersive environment where social cues can be more salient. Pedagogical agent design aspects such as speech, appearance, and modality can prime social cues and affect learning outcomes and instructor perception. In this paper, we propose a project to investigate auditory and visual social cues of pedagogical agents in XR such as speech, ethnicity, and modality.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Mixed / augmented reality; Applied computing—Education—Interactive learning environments

1 Introduction

The increasing commercial availability of virtual reality (VR) and augmented reality (AR) has allowed extended reality (XR) to be used for a variety of applications such as training, therapy, and entertainment. In particular, we are interested in the use of XR for education, which has seen a rapid increase in interest in the last few years [27, 33]. XR has the potential to enhance learning due to the unique affordances it offers learners compared to traditional online learning platforms [2, 27]. The use of XR for education is especially relevant due to the increased adaptation of online learning platforms, which was spurred by the COVID-19 pandemic [22].

We are particularly interested in the design of virtual teachers and mentors, which can prime social cues and affect learning outcomes and instructor perception [29]. A large body of work indicates that including a pedagogical agent in a learning environment can improve student learning and motivation [16, 21, 29, 34]. Furthermore, Ke et al. [17] suggested that virtual agents can improve the sense of presence and student motivation in virtual learning environments in particular. The design of pedagogical agents, which are any anthropomorphic virtual characters designed to support learning, is well-studied in traditional educational media [21]. However, the design of pedagogical agents is still a nascent area of study in XR.

Since XR can create more immersive learning opportunities than traditional learning methods [33], it is important to consider the design of pedagogical agents due to the social cues that they present. Social agency theory [29] refers to the idea that social cues in multimedia instructional messages can prime a social response in learning and influence cognitive processing and learning outcomes. Early work by Moreno et al. [29] indicates that pedagogical agents can promote meaningful learning in multimedia lessons specifically due to social agency. Domagk et al. [10] expanded on social agency theory and found that pedagogical agents with appealing social cues can promote increased transfer performance, but unappealing social cues might even hinder learning. As XR becomes more commonplace, it is important to understand how the qualities of pedagogical agents affect learners in immersive environments, where social cues may be more salient.

The design of pedagogical agents can be especially important in facilitating learning for minority and underrepresented populations (e.g., women in STEM, ethnic-minorities). Kim and Lim [19] reported that minority students who feel less supported in traditional classrooms develop more positive attitudes towards agent-based learning. Furthermore, prior work found preliminary evidence that females in mathematics classes [19] and ethic-minorities [4, 18, 28] are more sensitive to an agent's personal attributes (e.g., age, ethnicity, gender). Previous researchers have suggested that personalizing agents to individual students can positively affect learning [18], and furthering this body of work in the context of XR can provide meaningful insights on designing effective XR learning environments for all types of learners.

Feine et al. [12] presented a taxonomy of social cues for conversational agents, which classified social cues into four major categories (i.e., verbal, visual, auditory, invisible). We focus particularly on visual and auditory social cues of pedagogical agents that can be easily manipulated by a designer. Specifically, we plan to investigate the aspects of synthesized speech, considering previous work found that synthesized speech produces better learning outcomes than recorded human speech [8]. We also plan to investigate agent appearance and computer-mediated communication in the form of closed captions.

2 Related Work

2.1 Pedagogical Agent Design

The study of pedagogical agent design has evolved over the course of several decades. Early work by Moreno et al. [29] suggested that pedagogical agents can improve cognitive processing and learning outcomes due to their inherently social nature (i.e., social agency theory). In particular, Domagk et al. [10] investigated the valence of social cues and reported that appealing social cues can improve transfer performance, but unappealing social cues can hinder learning. A large body of work indicates that human interactions with computers are fundamentally social and that people apply social rules, norms, and expectations to computers [31], which can explain the influence of pedagogical agent design on learning and instructor perception.

This paradigm has been extended to pedagogical agents in several studies. Pedagogical agent design aspects such as voice, appearance, and gender can influence instructor perception and learning outcomes (e.g., [3, 4, 8, 24]). Lester et al. [20] posited that animated pedagogical agents have persona effects, which describes how the social characteristics of agents influence how much people enjoy interacting with them. Martha and Santoso [21] provided a thorough systematic review on the impact of design aspects for pedagogical agents and suggested that good design elements can influence students to be more involved in learning. For example, Nass et al. [30] found that students rated virtual instructors as more competent and useful when they had the same ethnicity as the student. Baylor and Kim [4] found similar results and reported that students working

*e-mail: tiffanydo@knights.ucf.edu

with agents of the same ethnicity perceived the agent to be more engaging and affable.

2.2 Education in XR

Education in XR has seen increased interest in the last few years due to the growing availability of XR platforms, frameworks, and technology [33]. Scavarelli et al. [33] recently provided a thorough literature review pertaining to social learning spaces in VR and AR and suggested that XR can be a powerful tool for learning. For example, learners can be situated within an Virtual Learning Environment (VLE) that may be difficult or impossible to visit in the real world. A recent study by Baceviciute et al. [2] found that simply situating a learner in a related environment to the material (i.e., learning about cancer in a hospital room) improved knowledge transfer while maintaining the same learning materials, which provides compelling reasons to further investigate education in immersive environments.

Several aspects of XR can be manipulated in order to influence learning. For instance, McMahan and Herrera [26] suggested that aspects of system fidelity can be manipulated in order to enhance cognition. Scavarelli et al. [33] recommended that the effects of socio-cultural aspects in XR learning spaces should be an area of future study. In particular, we are interested in the socio-cultural aspects of pedagogical agent design, which can promote learning in simulation based learning in VR [17]. Furthermore, in a literature review on pedagogical agent design, Martha and Santoso [21] recommended that the visual design aspects of pedagogical agents should be studied within the context AR technology due to the potential interaction effects of virtual environments.

3 Project Overview

In this section, we describe future areas of study pertaining to the design of virtual teachers and mentors for XR. We focus on speech, visual design, and modality of virtual pedagogical agents.

3.1 Synthesized Speech

Agent speech has been an area of notable interest in both human-agent interactions and pedagogical agent design. Seaborn et al. [35] recently provided a thorough review on the effects of agent voice on a wide scope of human-agent interactions. They found that agent voice has a prominent social role in human-agent interactions, which can influence a user's perception of the agent.

The use of machine-synthesized speech, such as text-to-speech (TTS), has seen considerable interest in the educational field (e.g., [1, 5, 8, 25]. Synthesized speech can be automatically generated from text and can be easier and quicker to record than hiring voice actors. However, early work found that TTS had negative effects on learning outcomes and instructor perception [1, 25]. More recent work suggested that this effect should be reconsidered due to the substantial advancements in synthetic speech technology [5, 8] and indicate that modern synthetic speech engines are an effective option for virtual pedagogical agents.

Interestingly, Craig and Schroeder [8] found that a modern synthesized speech engine produced even better learning than human speech, which provide compelling reasons to further investigate the influence of synthesized speech in XR learning platforms. Craig and Schroeder suggested that a pedagogical agent with modern TTS produces better learning than human speech due to the uncanny valley effect. In essence, they suggested that the dissonance between the visual of the virtual human and the human voice may have negatively impacted learning. In this project, we plan to investigate the effects of both synthesized speech accents and synthesized speech quality on agent perception and learning outcomes.

3.1.1 Synthesized speech accents

The influence of an instructor's accent on both instructor perception and learning outcomes has been noted by researchers in the education field. Early work by Gill [13] argued that different accents yield different perceptions of teachers and that teachers with similar accents to the student are perceived more favorably (e.g. more intelligent and dynamic). Additionally, both Gill and Mayer et al. [25] reported that foreign accented human speech produced significantly lower learning outcomes for students compared to native accents. Mayer et al. [25] implied that their results may be attributed to the cognitive load theory [37]—since students must attribute more cognitive resources to understand foreign accented speech, they have less cognitive capacity for meaningful learning.

Furthermore, users typically prefer an agent with a similar accent to their own within human-agent interactions. In a broad literature review pertaining to the effects of agent voice on human-agent interactions, Seaborn et al. [35] reported that in many studies, agents with accents that match the user's were more positively perceived. This effect persisted across multiple regional accents (e.g., United States English, New Zealand English) and languages.

However, there is little work regarding the influence of a pedagogical agent's speech accent, especially within XR applications. Advancements in TTS have allowed for the advent of realistic regional synthetic accents (e.g., American English, British English, Indian English), which can prime social cues and affect learning outcomes and instructor perception. For example, students may learn better with a pedagogical agent that has a similar synthesized accent as them. As online education platforms grow in popularity around the world [9], understanding the effects of these social cues can help developers decide whether personalizing a pedagogical agent to a student's demographics could improve learning. We plan to investigate a pedagogical agent's synthesized accent while controlling for learner ethnicity. We hypothesize that students would learn better with an agent that has an accent similar to their own.

3.1.2 Synthesized speech quality

The advent of neural network TTS, which uses neural networks as a postfiltering step [6], has allowed for more natural sounding synthetic speech. Neural TTS has gained commercial popularity in recent years. Microsoft released their commercially available neural TTS voices in 2018 [15], while Amazon introduced neural TTS voices in AWS Polly in 2019 [36]. The rise in popularity of neural TTS may be attributed to its potential to improve user experience. For example, Microsoft claimed that their neural TTS voices reduce listening fatigue and sound more natural than standard synthetic voices [15]. In this project, we plan to investigate compare neural TTS to concatenative TTS and recorded human speech. We hypothesize that the reduced listening fatigue of neural TTS may improve learning by reducing cognitive load, while also maintaining the benefits of using TTS over human speech.

3.2 Agent Ethnicity

The perceived ethnicity of an agent may influence a learner's perception of the agent. Early studies using on-screen agents found that students working with an agent that had a similar ethnic appearance as their own perceived the agent more positively [3, 4, 28, 30]. Furthermore, this effect was more prominent for minority students, such as students of color [3, 28].

The visual quality of virtual humans has greatly improved since these seminal studies were conducted in the early 2000's. It is important to investigate whether a high-fidelity virtual human's perceived ethnicity impacts learner perception in order to understand whether ethnicity serves as a salient social cue with improved visuals. Finally, this effect may change within immersive virtual learning environments where a virtual agent's appearance may play a more prominent role in learning. In this project, we plan to investigate an agent's perceived ethnicity while controlling for learner ethnicity within immersive virtual learning environments. We hypothesize that students will prefer agents that have a similar ethnic appearance as their own.

3.3 Agent Modality

Prior work regarding different agent modalities has been mixed. In a literature review, Martha et al. [21] suggested that on-screen pedagogical agents communicate better via spoken text than voice or text only. However, both Craig et al. [7] and Mayer [23] found that providing redundant information in multiple forms (e.g., spoken-text along with printed text) negatively affected learning when learning from an animated agent. It is unclear whether including subtitles for narrated text would improve learning. Additionally, subtitles may particularly help ESL (English as a Second Language) students, considering prior work found that subtitles improve listening comprehension [14] and learning outcomes [11] for ESL learners.

Furthermore, reading educational text in VR requires more cognitive effort and time than reading text in normal settings [2,32], which presents a unique problem in VR learning environments. For instance, reading in VR can pose an issue due to the cognitive-load theory [37]. An increased cognitive load during a learning experience can cause students to have less cognitive capacity for meaningful learning, which negatively affects their learning outcomes.

While implementing subtitles or closed captions for pedagogical agents in VR can be a fairly simple task, it is important to understand the effects of modality on learning. For example, learners may devote more cognitive effort to reading the closed captions and have worse learning outcomes as a result. In our project, we plan to investigate the effects of including subtitles for a pedagogical agent in an immersive virtual environment on learning outcomes and learning efficiency. We hypothesize that including subtitles for a pedagogical agent in VR will have worse learning outcomes than not including subtitles.

4 Conclusion

Virtual and augmented reality can provide unique learning opportunities and aid in learning. Pedagogical agents within these immersive virtual learning environments can improve the sense of presence and increase student motivation [17]. However, the design of pedagogical agents can also present salient social cues that can affect learning outcomes as well as a student's perception of the instructor [21]. The effects of these social cues often depend on a learner's personal attributes, such as their ethnicity or gender [4,28]. Furthermore, it is suggested that underrepresented learners are more sensitive to an agent's personal attributes [19].

It is important to investigate how agent design aspects affect learners in order to understand how to best utilize pedagogical agents within XR learning platforms. In particular, we plan to investigate synthesized speech accent, synthesized speech quality, agent ethnicity, and modality. With this project, we hope to gain insights on designing virtual pedagogical agents for XR that aide students in learning.

Acknowledgements

This material is based on work supported by the National Science Foundation under Grant No. 2021607 – "CAREER: Leveraging the Virtualness of Virtual Reality for More-Effective Training."

References

[1] R. K. Atkinson, R. E. Mayer, and M. M. Merrill. Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, 30(1):117–139, 2005. doi: 10.1016/j.cedpsych.2004.07.001

[2] S. Baceviciute, T. Terkildsen, and G. Makransky. Remediating learning from non-immersive to immersive media: Using EEG to investigate the effects of environmental embeddedness on reading in Virtual Reality. *Computers and Education*, 164(December 2020):104122, 2021. doi: 10.1016/j.compedu.2020.104122

[3] A. Baylor. Effects of Images and Animation on Agent Persona. *Journal of Educational Computing Research*, 28(4):373–394, 2003.

[4] A. L. Baylor and Y. Kim. The role of gender and ethnicity in pedagogical agent perception. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education. Association for the Advancement of Computing in Education (AACE)*, 2003.

[5] E. K. Chiou, N. L. Schroeder, and S. D. Craig. How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers and Education*, 146, 2020. doi: 10.1016/j.compedu.2019.103756

[6] M. Coto-Jiménez and J. Goddard-Close. LSTM Deep Neural Networks Postfiltering for Enhancing Synthetic Voices. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(1):1–24, 2018. doi: 10.1142/S021800141860008X

[7] S. D. Craig, B. Gholson, and D. M. Driscoll. Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94(2):428–434, 2002. doi: 10.1037/0022-0663.94.2.428

[8] S. D. Craig and N. L. Schroeder. Reconsidering the voice effect when learning from a virtual human. *Computers and Education*, 114:193–205, 2017. doi: 10.1016/j.compedu.2017.07.003

[9] S. Dhawan. Online Learning: A Panacea in the Time of COVID-19 Crisis. *Journal of Educational Technology Systems*, 49(1):5–22, 2020. doi: 10.1177/0047239520934018

[10] S. Domagk. Do pedagogical agents facilitate learner motivation and learning outcomes?: The role of the appeal of agent's appearance and voice. *Journal of Media Psychology*, 22(2):84–97, 2010. doi: 10.1027/1864-1105/a000011

[11] K. C. S. Dumlao, R. C. Alfonso, E. S. Paguirigan, and G. S. Subia. The Efficacy of Bimodal Subtitling in Improving the Listening Comprehension of English as a Second Language (ESL) Learners. *International Journal of English Literature and Social Sciences*, 5(6):2196–2201, 2020. doi: 10.22161/ijels.56.54

[12] J. Feine, U. Gnewuch, S. Morana, and A. Maedche. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human Computer Studies*, 132(July):138–161, 2019. doi: 10.1016/j.ijhcs.2019.07.009

[13] M. M. Gill. Accent and Stereotypes: Their Effect on Perceptions of Teachers and Lecture Comprehension. *Journal of Applied Communication Research*, 22(4):348–361, 1994. doi: 10.1080/00909889409365409

[14] A. Hayati and F. Mohmedi. The effect of films with and without subtitles on listening comprehension of efl learners. *British Journal of Educational Technology*, 42(1):181–192, 2011.

[15] X. Huang. Microsoft's new neural text-to-speech service helps machines speak like people, 2018.

[16] W. L. Johnson and J. C. Lester. Face-to-Face Interaction with Pedagogical Agents, Twenty Years Later. *International Journal of Artificial Intelligence in Education*, 26(1):25–36, 2016. doi: 10.1007/s40593-015-0065-9

[17] F. Ke, Z. Dai, C.-P. Dai, M. Pachman, R. S. Chaulagain, and X. Yuan. Designing Virtual Agents for Simulation-Based Learning in Virtual Reality. In *Cognitive and Affective Perspectives on Immersive Technology in Education*. 2020. doi: 10.4018/978-1-7998-3250-8.ch008

[18] Y. Kim. The Role of Agent Age and Gender for Middle-Grade Girls. *Computers in the Schools*, 33(2):59–70, 2016. doi: 10.1080/07380569.2016.1143753

[19] Y. Kim and J. H. Lim. Gendered Socialization with an Embodied Agent: Creating a Social and Affable Mathematics Learning Environment for Middle-Grade Females. *Journal of Educational Psychology*, 105(4):1164–1174, 2013. doi: 10.1037/a0031027

[20] J. C. Lester, S. G. Towns, C. B. Callaway, J. L. Voerman, and P. J. Fitzgerald. Deictic and Emotive Communication in Animated Pedagogical Agents. In *Embodied Conversational Agents*, pp. 123–154. MIT press, Cambridge, MA, 2000. doi: 10.7551/mitpress/2697.003.0007

[21] A. S. D. Martha and H. B. Santoso. The design and impact of the pedagogical agent: A systematic literature review. *Journal of Educators Online*, 16(1), 2019. doi: 10.9743/jeo.2019.16.1.8

[22] C. T. Martín, C. Acal, M. E. Honrani, and Á. C. M. Estrada. Impact on the virtual learning environment due to covid-19. *Sustainability (Switzerland)*, 13(2):1–16, 2021. doi: 10.3390/su13020582

[23] R. E. Mayer and R. Moreno. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2):312–320, 1998. doi: 10.1037/0022-0663.90.2.312

[24] R. E. Mayer and C. Pilegard. Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In *The Cambridge Handbook of Multimedia Learning, Second Edition*, pp. 316–344. 2014. doi: 10.1017/CBO9781139547369.016

[25] R. E. Mayer, K. Sobko, and P. D. Mautone. Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, 95(2):419–425, 2003. doi: 10.1037/0022-0663.95.2.419

[26] R. P. McMahan and N. S. Herrera. AFFECT: Altered-fidelity framework for enhancing cognition and training. *Frontiers in ICT*, 3:1–15, 2016. doi: 10.3389/fict.2016.00029

[27] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis. Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers and Education*, 70:29–40, 2014. doi: 10.1016/j.compedu.2013.07.033

[28] R. Moreno and T. Flowerday. Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology*, 31(2):186–207, 2006. doi: 10.1016/j.cedpsych.2005.05.002

[29] R. Moreno, R. E. Mayer, H. A. Spires, and J. C. Lester. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19(2):177–213, 2001. doi: 10.1207/S1532690XCI1902_02

[30] C. Nass, K. Isbister, and E.-J. Lee. Truth is beauty: Researching embodied conversational agents. *Embodied Conversational Agents*, pp. 374–402, 01 2000.

[31] C. Nass, J. Steuer, and E. R. Tauber. Computers are Social Actors. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 72–78, 1994. doi: 10.1109/VSMM.2014.7136659

[32] P. L. P. Rau, J. Zheng, Z. Guo, and J. Li. Speed reading on virtual reality and augmented reality. *Computers and Education*, 125(April):240–245, 2018. doi: 10.1016/j.compedu.2018.06.016

[33] A. Scavarelli, A. Arya, and R. J. Teather. Virtual reality and augmented reality in social learning spaces: a literature review. *Virtual Reality*, 25(1):257–277, 2021. doi: 10.1007/s10055-020-00444-8

[34] N. L. Schroeder, O. O. Adesope, and R. B. Gilbert. How effective are pedagogical agents for learning? a meta-analytic review. *Journal of Educational Computing Research*, 49(1):1–39, 2013. doi: 10.2190/EC.49.1.a

[35] K. Seaborn, N. P. Miyake, P. Pennefather, and M. Otake-Matsuura. Voice in Human–Agent Interaction. *ACM Computing Surveys*, 54(4):1–43, 2021. doi: 10.1145/3386867

[36] J. Simon. Amazon Polly Introduces Neural Text-To-Speech and Newscaster Style, 2019.

[37] J. Sweller, P. Ayres, and S. Kalyuga. *Cognitive Load Theory*. Springer Science+Business Media, New York, 2011. doi: 10.1007/978-1-4419-8126-4_2