Deep Learning with Virtual Agents

How Accented and Synthetic Voices Affect Outcomes

by

Robert Franklin Siegle

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved June 2022 by the
Graduate Supervisory Committee:

Scotty D. Craig, Chair
Nancy J. Cooke
Brian C. Nelson

ARIZONA STATE UNIVERSITY

August 2022

ABSTRACT

The current study investigates accent effects using virtual agents in the context of a multimedia learning environment. In a 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) between-subjects factorial design, the source and accent of the agent's voice were manipulated. Research has shown that an instructor's accent can have an impact on learning outcomes and perceptions of the instructor. However, these outcomes and perceptions have yet to be fully understood in the context of a virtual human instructor. Outcome measures collected included: knowledge retention, knowledge transfer, and cognitive load. Perception measures were collected using the Agent Persona Instrument-Revised, API-R, and a speaker-rating survey. Overall, there were no significant differences between the accented conditions. However, the synthetic condition had significantly lower knowledge retention, knowledge transfer, and mental effort efficiency than the professional voices in the human condition. Participants rated the human recordings higher on speaker-rating and API-R measures. These findings demonstrate the importance of considering the quality of the voice when designing multimedia learning environments.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Learning from an accented instructor can have several impacts on the learning experience. Kavas and Kavas (2008) have outlined the advantages of an accented instructor; e.g. the opportunity to learn from a different cultural perspective and improve communication abilities. However, there have been disadvantages identified in learning outcomes: the increased concentration required to process learning material can affect cognitive load during the learning process (Kavas and Kavas, 2008). This burden on cognitive load has been shown to reduce performance in those that find the accent difficult and are frustrated at any communication barrier (Ahn, 2010). The positive and negative effects surrounding accented instruction have become a pressing matter to address, as the internet provides a global platform for educational material (Blommaert, 2009). Bias against an accent can influence learners' perceptions of the materials' effectiveness, grammatical accuracy, and professionalism (Boucher, et al., 2013). Some researchers have theorized that the globalization of multimedia material will eventually neutralize the impact of accents and identities associated with them (Aneesh, 2015). However, previous research has shown that participants currently have a preference for their own accent, even when accustomed to learning from a foreign accent (Pilus, 2013). The impact accents have on learning is normally a reactionary situation. However, with digitally created instructors and synthetic voices, an accent becomes a design choice. Although accented virtual instructors have not been fully researched, recent studies have shown that changes to their speech patterns can impact learning outcomes (Craig et al., 2019), deep learning outcomes (Davis et al., 2019), and how much the learner trusts their

virtual instructor (Schroeder et al., 2020). This poses the question: *How does the use of accented voices impact learning outcomes when paired with a virtual human?* This study aims to answer this with a study design based on gaps in the literature on virtual agents, social agency theory, and voice effect.

Virtual Humans

Learning from a virtual instructor has been studied extensively to determine the effects of a non-human instructor since Johnson et al. (1997) first placed an animated agent in a knowledge-based learning environment and coined the term pedagogical agent (Johnson et al., 1997). These virtual instructors are commonly referred to as *virtual humans*, a universal term for any human-like animated character in a virtual or multimedia setting (Craig et al., 2015). There are two sub-groupings of virtual humans depending on who is controlling the on-screen entity. The first subgroup is the *avatar*, which is controlled by a human user. Avatars give human users on-screen representation and a visual means of interacting in the virtual world (Bailenson et al., 2008). The second subgroup is the *agent*, which is also referred to as *animated agents*, *pedagogical agents*, or *synthetic humans*. Agents are controlled by the software interface that designed them, with actions either prerecorded or designed for human-agent interaction, to provide information to the human user (Craig et al., 2002).

Since their creation, virtual humans have been used in a variety of educational settings; to provide tutorial dialogue (Lourdeaux et al., 2002), and take on the role of a coach (Shamekhi & Bickmore, 2015), or provide learning companionship (Rickel et al., 1999). One parallel throughout the different research topics has been a desire for more human-like virtual humans. Virtual humans that can display emotional reactions are used

in clinical interfaces to present patients with information (Rizzo et al., 2011). Agents programmed to use communication strategies such as greetings, farewells, and regional phrases are preferred as conversational partners (Kopp, 2006). High-quality voice engines that have a more human-like sound are both preferred by users and improve learning outcomes (Craig & Schroeder, 2017). The agent's more mechanical aspects have also been connected to how human-like and desirable they are for use. Improving the agent's appearance enhances the interaction effects between the virtual human and user, even down to physically smaller details such as the design of eyes (Ruhland et al., 2015). The virtual human's movement also has an impact on interaction outcomes, with more dynamic and life-like agents outperforming those that are stiff, static, or perceived to move robotically (Craig et al., 2015). The research for designing humans that provide social cues originates from Mayer's (2002) work in multimedia presentations. When the learner was primed by social cues in the learning material, social schema would activate for the learner to believe a conversation was taking place. Priming this schema through social cues was found to increase learning performance, and create the *Social Agency Theory* (Mayer, 2002).

Learning with Virtual Agents

Virtual agents are increasingly being used as a teaching tool. Virtual agents have shown to have a positive impact on learning, are cost-effective for being highly reusable, and have proven themselves with a wide variety of learners. Improvements in learning outcomes from the use of a virtual agent have been documented early on in elementary schools (Davis & Antonenko, 2017), found to help students throughout the K-12 and into college (Hew & Cheung, 2010), and shown to help the elderly rebuild lost cognitive skills

(Tao, 2016). In general education, virtual agents have been utilized within classrooms teaching chemistry in virtual laboratories (Morozov et al., 2004) and in activities outside of the classroom, such as providing mentorship in virtual science career fairs (Beck et al., 2018). These virtual agents have been used in both in-person activities and online learning. Online learning continues to become an increasing part of education for many individuals around the world. The aspect of virtual agent inclusion in online learning continues to be the subject of extensive study and review (Craig et al., 2020).

Utilizing social agency theory, virtual agents have also been used outside of general education. Zoll and colleagues (2006) studied the use of virtual agents to prevent bullying in schools. The program "FearNot" used virtual agents to create bullying scenarios in which the learner would spectate. By using virtual agents that appeared to be empathic, Zoll and colleagues were able to teach *Personal and Social education* to students. The virtual agents in the program would model empathy, showing negative expressional behavior towards being bullied while verbalizing their feelings towards each other as the learner observed. (Zoll et al., 2006). Interview training software has capitalized on virtual agents' ability to produce realistic reactions, in which the agent takes the place of the interviewer and provides social cues to the user depending on their performance in the mock interview scenario (Jones et al., 2014). Improvements in interviewing skills were also used to help those with autism and developmental disabilities, giving credence that virtual agents may even be superior in certain circumstances (Burke et al., 2018). Social agency theory has since improved the quality of virtual agents by making the virtual agent more human-like, increasing its benefit to the user (Mayer, 2002).

Mayer and colleagues (2003) examined the impact of voice effects with both accented and synthetic voices. Since then, only virtual agents with synthetic voices have been studied in great detail. Although foreign-accented voice effects were briefly touched on, the complete impact accents can have on virtual agents' effectiveness is still unknown (Mayer et al., 2003).

Social Agency Theory

Mayer's (2002) social agency theory was developed from work with multimedia learning materials. For materials to be classified as multimedia, the expression of information must be present through more than one medium, e.g. words and images. The multimedia principle supports the idea that individuals learn more deeply from words and images presented together, compared to words alone. This research in multimedia learning demonstrated how multiple sources of information could be presented to a learner at one time without causing cognitive overload (Mayer, 2002). The detriment of cognitive overload is documented in Sweller's (1988) research on *Cognitive Load Theory*. Cognitive load theory states that placing a heavy burden on working memory during learning will result in poorer performance and lower learning outcomes. However, any strain on working memory can be reduced by providing multiple sources of the same information (Sweller, 1988). Chandler and Sweller (1991) researched the different ways lessons are presented to learners in a classroom, focusing on examining the connection between cognitive load and learning outcomes. The findings support instructional materials that guide the learner cognitively throughout the material and do not split the learner's attention to different areas at any given point. This research has been used to

improve educators' understanding regarding integrating information into a lesson (Chandler & Sweller, 1991).

Following Chandler and Sweller's findings, research began on best practices for guiding learners cognitively through the material. Interestingly, a live human instructor is not required to guide the learners, as social cues in multimedia material were found to provide the same meaningful effects (Moreno & Mayer, 2000). This finding is likely because people can view a computer as an entity that can socialize (Reeves & Nass, 1996). The combined findings on cognitive load theory, the human ability to perceive social cues from a machine, and the discovery that social cues in multimedia had a meaningful impact formed the guidelines for the multimedia principle and set the stage for Mayer's (2003) social agency theory. Social agency theory states that social cues within instructional material prime the learner to act as though they are interacting with another human being. In the case of listening to a multimedia presentation, the priming creates the feeling of a conversation taking place and the learner responds to the material in a more meaningful way. The learner has a natural familiarity with the dynamics of a conversation and the social cues embedded in the learning material direct the learner's focus to the important information. Drawing attention to specific elements of the instruction and improving learning outcomes (Mayer et al., 2003).

Following social agency theory, virtual humans were created to further enhance the effects caused by social cueing and further increase the priming that the learner was in a conversation with another person (Louwerse et al., 2005) and allow for students to learn more deeply from virtual humans (Moreno, 2001). The additional means to communicate non-verbally and cue the learner through movements and hand gestures

have been shown to improve results over traditional computer-based instruction (Pelachaud, 2009; Moreno, 2001). As design technology advanced, subtle movements such as the agent's eye gaze were used to elicit social cues and further increase learning outcomes (Ruhland et al., 2015). In addition to the virtual humans' visual appearance impacting learning outcomes, the voice of the agent also affects the learning process.

Origin and Uses of Voice Effect

Mayer and colleagues' (2003) study of social cueing and social agency theory contained two experiments involving multimedia learning materials. In the first experiment, the audio was manipulated by presenting the material in either a standard American accent or a foreign Russian accent. In the second experiment, the voice was either a human recording or a machine synthesized voice (Bruce from Macintosh G4). Participants in both experiments had improved learning outcomes when the information was presented in a voice the learner was more familiar with, American accents and human recordings respectively. Participants in these familiar conditions also rated the speaker significantly higher in a perception measure. These results demonstrate that the ease with which a voice is understood can impact the cognitive load put on the learner. This was the first documented evidence for the *voice effect* (Mayer et al., 2003). One noteworthy absence in the discovery of the voice effect was the presence of a virtual agent. In a follow-up study, Atkinson and colleagues (2005) created a similar study to Mayer's second (2003) experiment using human and synthesized voices. In Atkinson's study, the voices were paired with a virtual human to further improve social cueing effects by adding a visual presence. In similar findings to Mayer's study, the virtual human with a human voice produced better learning outcomes than the agent with a

synthesized voice (Atkinson et al., 2005). Research has since shown that having both the visual presence of the virtual human and their auditory presence through voice, has a positive interaction effect on improving learning outcomes (Domagk, 2010) and increasing student motivation (Heidig & Clarebout, 2011).

The voice effect continues to be studied extensively with the use of virtual humans to determine best practices in voice design. Virtual humans with more expressive voices have been shown to positively affect general learning outcomes (Bergmann & Kopp, 2009). Davis and colleagues (2019) found that virtual agents with more natural speech patterns have specific positive impacts on deep learning. Voices with high fluctuations of prosody in their speech: intonation, tone, stress, and rhyme improve deep learning outcomes. The high fluctuations cause the voice to be perceived as more human-like, and therefore better able to perform social cueing and reduce cognitive load (Davis et al., 2019). These cueing effects caused by voice are further enhanced when matching the voice to the virtual human's appearance (Mitchell et al., 2011). Virtual humans with voices that provide natural social cues are also rated higher on self-reported measures by learners (Louwerse et al., 2005). Ratings influenced by the voice effect go beyond measures of quality and can impact the level of trust learners have towards the virtual agent teaching them (Craig et al., 2019; Schroeder et al., 2020).

Diversity and Equity Implications

The research on accents, voices, and virtual humans can give insight on how to improve learning outcomes and create a more desirable learning environment. However, the findings on voice effect also have real-world implications for human instructors that are not able to change their voice as virtual humans are able to do. The definition of an

accent is a way in which speech is pronounced, but accents have social ties to location, class, or other ways in which people are grouped (Lippi et al., 1997). Unfortunately, a combination of identifying a person by their accent and stereotyping their identity can allow for individuals to be unfairly judged based on voice alone (Riches & Foddy, 1989).

Workplace Implications

Those with a minority accent have been shown to face discrimination in job interviews. Participants acting as interviewers, showed bias while listening to an audio recording of a potential applicant. The applicant, always having the same name as Victor, differed only in each condition by using a different accent, Midwestern US, French, and Colombian. Without being able to see the applicant, interviewers rated the Midwestern candidate the most favorably, despite all the voice recordings even being generated by the same person (Deprez-Sims & Morris, 2010). This unfortunate discrimination is not an isolated incident when it comes to research involving prejudice in hiring practices. African-American sounding names, in which job applicants with names that sounded distinctly African-American were also viewed as less favorable (Cotton, et al., 2008).

Boucher and colleagues (2013) found that bias in perception of accents can occur with only a regional difference. Controlling for race and country, a Southern United States accent and a Midwestern United States accent were compared. Participants viewed the Midwestern accent as being more grammatically correct, more professional in mannerisms, and the speaker to be more effective overall. This regional accent discrimination shows that there is more nuance to the demographics involved with speech than simply the race of the speaker (Boucher, et al., 2013).

Academic Implications

The workplace is not the only real-world institution where accent can be an influential factor of perceived opinion. Subtirelu (2015) found that in the university setting, accents were a common comment on the website RateMyProfessors.com. Evaluations from students, towards professors with Asian accents, were rated as lower in helpfulness and clarity when compared to their English accented counterparts. A detailed analysis of the comments left by students also showed a significant tendency to mention the accent and understandability of the Asian professors more than the comments of the English accented professors. The comments towards the Asian professor's speaking ability ranged being geared towards more informative: "Her accent is a little hard to understand sometimes, but if you just ask, she'll repeat" and "He is hard to understand at first" to more hostile comments "AWFUL! AWFUL! DO NOT TAKE THIS PROF!! HE BARELY SPEAKS ENGLISH AND IS RUDE TO STUDENTS!" and "Professor Kim has a THICK accent so don't bother asking questions unless you speak Korean." Subtirelu notes that neither type of comment should be mistaken for harmless, and that the professors face a real disadvantage if this or similar student evaluations are used for performance reviews (Subtirelu, 2015).

Kavas and Kavas (2008) preformed an exploratory study of college student' attitudes towards professors with foreign accents. The study outlined potential advantages: e.g. learning from people of different cultural views and improving communication abilities, and disadvantages: e.g. concentration required to learn material and frustration at communication barriers. The study polled students on if the foreign accent of a teacher would affect their ability to learn or cause them to focus more on the

accent than the material, which 39.5% and 38.5% answered yes respectively (Kavas & Kavas, 2008). This poses a potentially significant problem, as in a study examining the effects of accent on cognitive load and achievement, students who stated that they disliked Asian accents exhibited lower assessment performance in knowledge transfer (Ahn, 2010).

McLean (2007) found that the effect accents can have on the learning process does not end with the students' perceptions. Teachers with heavy accents have reported concern that they are not being able to effectively communicate in the classroom due to their accents. Second guessing themselves if looks of confusion from the students are from the material itself not being understood, or if they are not being understood. The participants in McLean's study reported a high confidence level, about their knowledge in the courses they taught. However, the self-confidence was lowered when a language barrier was in play, and the accented instructors felt that their credibility as knowledgeable teachers was called into question more easily (McLean, 2007).

Theses impacts accents have in the learning environment is becoming a more pressing matter as the internet provides a globalized market for educational material (Blommaert, 2008). With the creation of material that is presented across the globe, the likelihood of a student being presented with a foreign, or accent different from their own, is greatly increased. Some have theorized that the globalization of multimedia material "neutralizes" foreign identities and things associated with them such as accents (Aneesh, 2015). Pilus (2013) found that in contradiction to these globalization ideals, there is still a tendency for users to report a preference for their own accent. In a study with Malaysian students who learned English as a second language, the participants were given audio

material in: English with a British accent, English with an American accent, and English with a Malaysian accent. Despite the audio being in English, and the participants reporting a higher admiration for the British accent, the Malaysian accented audio was the preferred learning material. (Pilus, 2013). Although a preference for accents has been reported, the impact accents have on learning with virtual agents has yet to be fully understood. Research showing affects from similar difference in speech sounds and patterns has given this study a basis on what to look for.

Current Study

The findings in Mayer's (2003) original voice effect *Study 1* involving accents have yet to be replicated with the presence of virtual humans, like how Atkinson (2005) replicated *Study 2* with synthetic voices (Atkinson et al., 2005; Mayer et al., 2003). This gap in the literature means it is unknown how virtual humans paired with accented voices impact the dependent variables from Mayer's (2003) *Study 1*; learning retention, learning transfer, cognitive load, and perceptions ratings. The literature has previously indicated that humans are capable of bias against virtual humans based on race, which means a bias against a virtual human based on their accent is supported for investigation (Zipp & Craig, 2019).

Additionally, recent studies by Craig and Schroeder (2017; 2019) reexamined the voice effect by pairing virtual humans with either a modern voice engine (Neospeech) or a classic voice engine (Microsoft speech engine), or a recorded human voice. The modern voice engine significantly outperformed the classic voice engine in learning outcomes, self-reported ratings, and training efficiency. However, the recorded human voice was rated at a similar level as the modern voice engine in self-reported facilitating learning

and credibility measures. The new voice engine was also rated a similar quality to the human voice. It was hypothesized that the low-quality voice engines from a decade ago had a significant impact on previous studies involving synthetic voices. These findings were then replicated and found again without the presence of a virtual human, to further replicate the original study (Craig & Schroeder, 2017; Craig & Schroeder, 2019). However, synthetic voice engines have yet to be paired with accents to determine if there is an interaction effect or if the effects of synthetic voices are accent-dependent.

The gender of the agents will be female for all conditions. Although the gender itself is not a manipulated variable, the impact and cause for this decision requires addressing. The use of a female agent allows the current study to mirror the methods of the previous studies (Atkinson et al., 2005; Craig & Schroeder, 2017; Craig & Schroeder, 2019), and minimize any alternative explanations for differences found in results. However, female agents have been shown to be perceived as less knowledgeable (Baylor & Kim, 2004) and are rated lower in perception-based measures than male agents (Rosenberg et al., 2010).

The goal of the current study was to fill in these literature gaps and identify the impact voice type and voice accent have on learning from a virtual human. Specifically, to understand how learning (retention, transfer), mental effort efficiency, and perceptions (speaker rating, API-R, appearance, and voice ratings) are impacted by voice type and voice accent when paired with the visual presence of a virtual agent. The use of both voice type and voice accent will allow for interaction effects to be studied for the first time. The novelty effect of a foreign-accented machine may allow that condition to outperform all others. However, the novelty effect has not been a significant factor in

recent work with virtual humans. Although the novelty effect remains a possibility for all of the research questions in the current study, it is not included in the predictions below. A 2 (voice type; human, synthetic) x 2 (voice accent; English, Russian) design was used to address the current hypotheses.

Hypotheses

Hypothesis 1, learning retention will be significantly impacted by voice accent with neutral voices having higher learning outcomes, with no significant impacts for voice type or the interaction between accent and type. The current literature has conflicting viewpoints on how learning retention would be impacted. With the presence of the voice effect, human accented voices previously had negative impacts on learning retention (Mayer et al., 2003). However, the voice effect did not produce a significant difference in learning retention when synthetic voices were paired with a virtual human (Atkinson et al., 2005) or when a modern voice engine was used (Craig & Schroeder, 2017; 2019). Because the previous literature has shown voice accent negatively impacts learning retention, and voice type does not impact learning retention, the following prediction was created: Human-English = Synthetic English > Human-Russian = Synthetic Russian.

Hypothesis 2, learning transfer will be significantly impacted by voice accent with neutral voices having higher learning outcomes, with no significant impacts for voice type or the interaction between accent and type. There are slightly different findings in previous literature regarding learning transfer compared to learning retention. Voice effect previously has shown both negative effects for learning transfer with accents (Mayer et al., 2003) and with synthetic voices paired with a virtual human (Atkinson et

al., 2005). The voice effect did not produce these same results for synthetic voices when a modern voice engine was used (Craig & Schroeder, 2017; 2019). As the current study used a modern voice engine, the prediction is that there will not be a significant difference in voice type outcomes. However, because the current literature on voice accent, would predict that the accented voices would produce lower levels of learning transfer, the following prediction was created: Human-English = Synthetic English > Human-Russian = Synthetic Russian.

Hypothesis 3, mental effort efficiency will be significantly impacted by voice accent with neutral voices having higher efficiency outcomes, with no significant impacts for voice type or the interaction between accent and type. The literature regarding mental effort efficiency has consistently found different voice effects to lower the mental effort efficiency of the learners. Overall, the social cueing principles on cognitive load support that a voice the learner is best able to pick up social cues from would have the improved results for the cognitive load (Park, 2015). If there is familiarity with the voice's accent or type, this may improve the ability to process the social cues produced. Accents have also been shown to reduce mental effort efficiency (Mayer et al., 2003). This was also found when using synthetic voices (Atkinson et al., 2005). However, the voice effect did not produce these same results for synthetic voices when a modern voice engine was used (Craig & Schroeder, 2017; 2019). As the current study used a modern voice engine, the prediction is that there will not be a significant difference in voice type outcomes. However, because the current literature on voice accent, would predict that the accented voices would produce lower levels of mental effort efficiency, the following prediction was created: Human-English = Synthetic English > Human-Russian = Synthetic Russian.

Hypothesis 4, perceptual outcomes will be significantly impacted by voice accent with neutral voices having higher ratings, with no significant impacts for voice type or the interaction between accent and type. The different literature regarding voice effect studies has used a variety of perception measures to determine how the voice effect impacts the view of the speaker. The speaker rating survey (Mayer, 2003) and API-R (Schroeder et al., 2017; 2018) are two of these perception measures, which were accompanied by direct questions regarding the virtual human's voice and appearance. The previous finding on perception measures has been consistent with the findings regarding the learning and mental effort outcomes, as learning decreases or mental effort increases the speaker is rated lower (Mayer et al., 2003; Atkinson et al., 2005; Schroeder et al., 2017; 2018). Based on these findings, the following prediction was created for all four of the research questions regarding perception measures: Human-English = Synthetic English > Human-Russian = Synthetic Russian.

CHAPTER 2

METHOD

Design

This study is a between-subjects two (accent) by two (type) factorial design to determine the impact the voice accent, voice type, and the interaction have on learning retention, learning transfer, mental effort efficiency, and perception measures. The accent variable is if the virtual human speaks with either a foreign (Russian) accent or a neutral (Midwestern) accent. The source variable is whether the virtual human speaks with either a pre-recorded human voice (Human) or created through the high-quality text-to-speech engine *Neospeech* (Synthetic). Given the possible combinations in this 2x2 design, the following condition blocks are created: Human-Accented, Human-Neutral, Synthetic-Accented, and Synthetic-Neutral.

Learning was measured by a series of open-ended questions that measured retention and transfer. These questions are from Mayer and Moreno's (1998) study on multimedia presentation and have since been used in virtual agent research (Atkinson et al., 2005; Craig & Schroeder 2017;2019). Cognitive load was measured using Pass' (1992) mental effort scale. Mental effort efficiency for training and testing was calculated using the formula, E = (Z performance – Z mental effort) / sqrt (2).

Perceptions of the virtual human were measured by a series of surveys and scales. The speaker rating survey was used to measure the superiority, attractiveness, dynamism, and overall quality of the virtual human as a speaker (Zahn & Hopper, 1985; Mayer et al., 2003). The Agent Persona Instrument-Revised, *API-R*, was used to measure perceptions of; facilitates learning, credibility, human-like, and engagement (Schroeder et al., 2017;

2018). Lastly, four questions directly asked about the perceptions of the agent's appearance, how natural the agent's voice was, how well the agent's voice facilitated learning, and how easy the agent's voice was to understand.

Participants

For this study, 197 participants were recruited. All participants were recruited from the online Amazon's Mechanical Turk, which was linked to a Qualtrics study for participation to complete. All participants were adults of age 18 years or older, had "normal or corrected-to-normal hearing", and were located within the continental United States of America. Participants read the consent information and consented by proceeding.

Participants' data were removed before analysis for the following reasons. Of the 197 participants that began the study, 66 were removed for either not completing the study or providing no answers to the testing questions. Upon review of the data set, another 50 participants were moved for plagiarism. Each of the participant's short answers was compared against all other participants' answers using the *find* function. Participants that had answered the same to each other verbatim were removed based on suspicion of using a search engine to look up answers to the question. The final data set contained 81 participants that passed all of the above requirements.

Materials

Audio Test

The participant received an audio test at the beginning of the study. The audio test ensured that the participant had *normal or corrected-to-normal levels of hearing* and that their audio equipment was working properly. A prerecorded audio file was played,

containing jazz music at a low volume while a four-digit code is spoken. The participant was required to enter the four-digit code by keyboard to continue with the experiment.

Learning Materials

The learning material was a two-minute video explaining lightning formation. The experimental materials were used in Mayer and Moreno's (1998) multimedia presentation on lightning formation. These materials have been adapted to include the use of virtual humans in previous research (Atkinson et al., 2005). The exact video script of the video was unchanged (Appendix A). Changes from the original video involved the addition of a virtual human and the sources of the voice recordings (Mayer et al., 2003). The virtual human was not a variable in this experiment, remaining the same across all conditions. The consistent variable of the voice across all conditions was the use of a female voice.

There were four voice conditions of a two (accent) by two (type) factorial design. The first independent variable (accent) is if the voice speaking in the video is of a neutral English accent or a Russian accent that is foreign to the user. The Russian accent was selected as the previous research examining virtual humans and accents also used a Russian accent as the for foreign condition (Atkinson et al., 2005; Mayer et al., 2003).

The second independent variable (type) is whether the voice speaking in the video was created synthetically or is a recording of a live human. The synthetic voice recordings are created using the Neospeech engine, which is the same engine that has been used in previous research for the high-quality synthetic voice engine (Craig & Schroeder, 2017; Craig & Schroeder, 2018). The human recordings were obtained from *Voices.com*, in which two professional voice artists were used for the human conditions.

The 2x2 design created the following conditions: Human-Accented, Human-Neutral, Synthetic-Accented, and Synthetic-Neutral (Table 1).

Table 1. Condition Blocks of Accent by Voice Variables

	Accented	Neutral
Human	Human-Accented	Human-Neutral
Synthetic	Synthetic-Accented	Synthetic-Neutral

Assessment Materials

Pretest Questionnaire. The pretest contained eight multiple-choice items. The questions were designed to reflect the scoring criteria of the retention test, which would act as the posttest comparison (Appendix B). The original questionnaire used as a pretest (Mayer & Moreno, 1998) was a 5-point Likert scale containing seven items, in which the participant self-reported their meteorology knowledge. The pretest questionnaire's original purpose was to eliminate participants that already had a high level of knowledge in meteorology, as previous studies have indicated that multimedia instruction produced stronger effects in learners with a low level of experience (Mayer & Gallini, 1990; Mayer & Sims, 1994). However, this previous pretest did not give an accurate assessment of the participants' lightning formation knowledge. To establish a more accurate pretest measure, the current test was implemented.

Demographic Questionnaire and Accent Familiarity. The demographic survey contained four questions. The survey asked the participants to identify their age, race, sex, gender, and educational level. In addition to this demographic survey, the

participants were also asked about their familiarity with different accents. A modified version of Huang et al.'s (2016) scale was used, which previously was used to identify familiarity with accented English speakers from Korean-accented English to Portuguese-accented English (Huang et al., 2016). The participants responded on a 4-point Likert scale; (1) *Not familiar at all*, (2) *Somewhat familiar*, (3) *Moderately familiar*, or (4) *Very familiar*. Of note, the participants were asked to identify their familiarity with Russian-accented English.

Retention Test. The retention test was the same single short answer question used in Mayer and Moreno's (1998) study in which it was created. The instructions for the retention test stated, *Please write down an explanation of how lightning works*. Participants were instructed to give the most complete answer possible with a time limit of four minutes. Scoring was done by two scorers, using a sample of the answers to establish inter-rater reliability using Cohen's Kappa, ensuring that the scorers are crediting participants' answers in the same way. Once reliability was established, the scorers proceeded to score the entire data set, tallying the number of acceptable answers. The eight possible ideas that the participant can receive credit for were: (a) air rises ($\kappa = 1.00$), (b) water condenses ($\kappa = 0.86$), (c) water and crystals fall ($\kappa = 1.00$), (d) wind is dragged downward ($\kappa = 0.83$), (e) negative charges fall to the bottom of the cloud ($\kappa = 1.00$), (f) the leaders meet ($\kappa = 0.86$), (g) negative charges rush down ($\kappa = 1.00$), and (h) positive charges rush up ($\kappa = 0.88$). After the four-minute timer, the participants were automatically moved to the next set of materials.

Transfer Tests. The four transfer tests consisted of four short answer questions used originally in Mayer and Moreno's (1998) study (and scoring is found in Mayer's

2003 social cues). Each test posed a single question with a two-minute time limit: (1) *What could you do to decrease the intensity of lightning?* (2) *Suppose you see clouds in the sky but no lightning, why not?* (3) *What does air temperature have to do with lightning?* and (4) *What causes lightning?* Scoring was done by two scorers, using a sample of the answers to establish inter-rater reliability using Cohen's Kappa, ensuring that the scorers are crediting participants' answers in the same way. Once reliability was established, the scorers proceeded to score the entire data set, tallying the number of acceptable answers. Each of the four tests allowed the participants to score two points each for a total of eight points. For the first question, acceptable answers included (1) statements about removing positively charged particles from the ground ($\kappa = 0.83$) AND (2) placing positively charged particles near the cloud ($\kappa = 1.00$). For the second question, acceptable included (3) stating that the top of the cloud might not be above the freezing level ($\kappa = 0.77$) AND (4) that there are no negatively charged particles in the cloud ($\kappa = 0.87$). For the third question, acceptable answers included (5) stating that the air must be cooler than the surface of the earth ($\kappa = 1.00$) AND (6) stating that the air must be cooler than the surface of the earth ($\kappa = 1.00$). For the fourth question, acceptable answers included (7) stating that there must be a difference of electrical charge within the cloud ($\kappa = 0.90$) AND (8) between the cloud and the ground ($\kappa = 0.88$). This test provided an assessment of the transfer of knowledge as the participant adapted the information given to them to fit the context of each question.

Cognitive Load Measurement. Pass' (1992) mental effort scale consists of a single 9-point Likert question: *In solving or studying the preceding problems I invested*.

The participant's response can range from *Very, Very low mental effort* as a 1, to *Very, very high mental effort* as a 9 (Appendix C). The measure was given to the participant three times throughout the study: (1) after the learning materials have been presented, (2) after the retention test is completed, and (3) after the four transfer tests have been completed. The participant's response was used to calculate the efficiency score, *E*, of the participant's training and testing efficiencies. The following equation was used for calculating these efficiency scores (Pass & van Merrienboer, 1993).

$$E = (Z\text{performance} - Z\ \text{Mental Effort}) / \sqrt{2}$$

Speaker Rating Survey. The speaker-rating survey consisted of 15 8-point Likert scale questions used by Mayer and colleagues (2003) and was adapted from the Speech Evaluation Instrument created by Zahn and Hopper (1985). The participants were asked to *indicate a number from 1 to 8 indicating how the speaker sounded along each of the 15 dimensions*. For each dimension, the numbers one and eight were labeled with paired adjectives. One was labeled with a negative quality and 8 were labeled with a positive quality. The survey had three subscales that each contained five of the 15 adjective pairs. The superiority subscale consisted of, *illiterate–literate, unintelligent-intelligent, uneducated-educated, not fluent-fluent,* and *inexperienced-experienced* ($\alpha = 0.86$). The attractiveness subscale consisted of, *unkind–kind, cold-warm, unfriendly-friendly, unpleasant-pleasant,* and *unlikable-likable* ($\alpha = 0.92$). Lastly, the dynamism subscale consisted of, *passive-active, shy-talkative, unaggressive-aggressive, unsure-confident,* and *lazy-energetic* ($\alpha = 0.63$). The five items from each subscale were averaged to find the participant's score for the respective subscale. Then the three subscale scores were averaged to find the overall speaker rating ($\alpha = 0.91$).

Agent Persona Instrument – Revised. The API-R survey is a revised version of Baylor and Ryu's (2003) instrument for gauging the perception of a virtual agent, the revision was created by Schroeder and colleagues (2017; 2018). The measure consisted of twenty-six 5-point Likert scale questions, answering with either 1 – *Strongly Disagree*, 2 – *Disagree*, 3 – *Neither Agree nor Disagree*, 4 – *Agree*, or 5 – *Strongly Agree*. The questions directly mirror the original Baylor and Ryu (2003) API and were used to gauge the participant's opinion of the virtual agent in four subscales. The facilitates learning subscale consisted of 10 items ($\alpha = 0.93$), the credibility consisted of five items ($\alpha = 0.87$), the human-like subscale consisted of five items ($\alpha = 0.92$), and the engagement subscale consisted of five items ($\alpha = 0.93$). Each of the four subscales was scored by averaging the participant's responses within each subscale. The list of questions for each subscale is listed in Appendix A. The 26th question in the API-R was a validation measure, used to determine if the participant was reading the questions and remaining vigilant. The question read as: *To ensure you are paying attention please answer this with 'Agree.'* This question was used as a means of vetting out participants that were not reading the questions, giving cause for removing the data of participants that did not answer the question with the answer *Agree*.

Agent Appearance and Voice Questions. A survey was given to the participants to directly ask them about their perceptions regarding the virtual agent. The measure consisted of four 5-point Likert scale questions, answering with either 1 – *Strongly Disagree*, 2 – *Disagree*, 3 – *Neither Agree nor Disagree*, 4 – *Agree*, or 5 – *Strongly Agree*. The first question measured the visual perception of the virtual agent; *I liked the agent's overall physical appearance.* The other three items given were used to determine

the perception of the agent's voice: *the voice of the agent was natural*, *the voice of the agent facilitated understanding of the message*, and *the agent was easy to understand*. This survey had no subscales and the four items were independently evaluated.

Procedure

The experiment took place exclusively in an online setting. After participants consented to participate in the study, an audio check was implemented to allow participants to set their volume to a desirable level and ensure they could hear the video's narrative. Participants then completed a short demographic survey, followed by an accent familiarity questionnaire, afterwards the participants completed a pretest. After the pretest, the participants were randomized into one of the four experimental conditions and presented with their groups' corresponding educational video on lighting formation. Once the video has concluded, the participant was asked the first cognitive load question, for the effort they used in learning the material.

The participants were then tested on their knowledge learned from the videos, beginning with a knowledge retention test. This test was followed by the second cognitive load question, for the effort used in the knowledge retention test. Afterward, the participants were asked to complete a series of four short knowledge transfer tests. Once the participants had completed the fourth knowledge transfer test, they were given the third and final cognitive load question, for the effort used in the knowledge transfer tests. After the final cognitive load question, participants were given the speaker rating survey and the API-R before concluding the study.

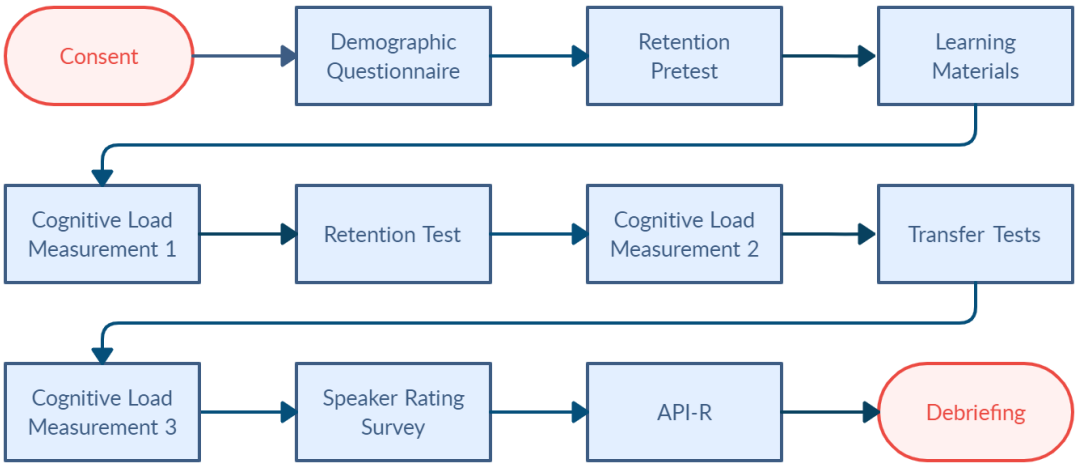

Figure 1. Procedure Flow Chart

CHAPTER 3

RESULTS

A series of 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) ANOVAs were conducted on pretest data with no significant interactions or main effects found for voice type $F(1, 77) = 1.66$, $p = 0.20$, $\eta p2 = 0.02$, voice accent $F(1, 77) = 0.67$, $p = 0.42$, $\eta p2 = 0.01$, or the interaction between voice type and voice accent $F(1, 77) = 0.00$, $p = 0.98$, $\eta p2 = 0.00$. Leven's test of equality of error variances failed to reject the hull hypothesis that there is homogeneity of variances $F(3, 77) = 1.75$, $p = 0.17$.

A series of 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) ANOVAs were conducted on accent familiarity data with no significant interactions or main effects found for voice type, $F(1, 77) = 0.03$, $p = 0.86$, $\eta p2 < 0.01$, voice accent, $F(1, 76) = 0.37$, $p = 0.55$, $\eta p2 = 0.01$, or the interaction between voice type and voice accent, $F(1, 76) = 0.09$, $p = 0.77$, $\eta p2 < 0.01$. Leven's test of equality of error variances failed to reject the hull hypothesis that there is homogeneity of variances $F(3, 77) = 1.20$, $p = 0.32$.

A series of 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) x 2 (gender of participant: female, male) ANOVAs were conducted on all of the dependent variables to determine if participant gender had a significant impact on the results. The gender of the participant was found to have no main effects, or interactions when examined with the voice type or voice accent. The only result trending towards significance is the effect of participant gender on learning retention, $F(1, 77) = 3.39$, $p = 0.07$, $\eta p2 = 0.07$, and the interaction between participant gender and voice accent, $F(1, 77) = 3.19$, $p = 0.08$, $\eta p2 = 0.04$. These non-statistically significant results are indicated

as research on gender match has previously shown significant results (Rosenberg-Kima et al., 2010; Baylor & Kim, 2004).

Learning Retention

A series of 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) ANOVAs were conducted on the participants' learning data to determine differences in learning retention, voice type, and voice accent.

There was a significant effect for voice type, $F(1, 77) = 13.49$, $p < 0.01$, $\eta p2 = 0.15$. There was not a significant effect for voice accent, $F(1, 77) = 0.76$, $p = 0.39$, $\eta p2 = 0.01$, or the interaction between voice type and voice accent, $F(1, 77) = 0.15$, $p = 0.70$, $\eta p2 < 0.01$.

In the main effect found for voice type, the human condition outperformed the synthetic condition in the learning retention test data. This shows voice type influenced retention-based learning outcomes.

Table 2
Learning Retention

Condition		Retention Test		
		N	Mean	SD
Russian Accent	Synthetic	22	2.09	1.95
	Human	19	3.63	2.50
	Total	41	2.80	2.33
English Accent	Synthetic	22	2.32	1.99
	Human	18	4.22	1.96
	Total	40	3.18	2.17
Total	Synthetic	44	2.20	1.95
	Human	37	3.92	2.24
Total Average		81	2.99	2.24

Learning Transfer

A mixed 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) ANOVA was conducted on the participants' learning data to determine differences in learning transfer, voice type, and voice accent.

There was a significant effect for voice type, $F(1, 77) = 10.94$, $p < 0.01$, $\eta p2 = 0.12$. There was not a significant effect for voice accent, $F(1, 77) < 0.01$, $p = 0.97$, $\eta p2 = 0.00$, or the interaction between voice type and voice accent, $F(1, 77) = 0.62$, $p = 0.44$, $\eta p2 < 0.01$.

In the main effect found for voice type, the human condition outperformed the synthetic condition in the learning transfer test data. This shows voice type influenced transfer-based learning outcomes.

Table 3

Learning Transfer

Condition		Transfer Tests		
		N	Mean	SD
Russian Accent	Synthetic	22	1.64	1.05
	Human	19	2.47	1.74
	Total	41	2.02	1.46
English Accent	Synthetic	22	1.36	1.40
	Human	18	2.72	1.74
	Total	40	1.96	1.69
Total	Synthetic	44	1.50	1.23
	Human	37	2.59	1.72
Total Average		81	2.00	1.57

Mental Effort Measures

A series of 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) ANOVAs were conducted on the participants' mental effort data to determine differences in mental effort efficiency, voice type, and voice accent. Mental effort training efficiency was calculated by using the performance of the retention test and the mental effort used

in watching the learning materials video. The mental effort training efficiency was also analyzed for the participants testing efficiency on the retention test and transfer tests.

In the standardized mental effort training efficiency, there was a significant effect for voice type, $F(1, 77) = 11.96$, $p < 0.01$, $\eta p2 = 0.13$. There was not a significant effect for voice accent, $F(1, 77) = 0.38$, $p = 0.54$, $\eta p2 = 0.01$, or the interaction between voice type and voice accent, $F(1, 77) = 0.50$, $p = 0.48$, $\eta p2 = 0.01$.

In the standardized mental effort efficiency for the retention test, there was a significant effect for voice type, $F(1, 77) = 9.36$, $p < 0.01$, $\eta p2 = 0.11$. There was not a significant effect for voice accent, $F(1, 77) = 0.15$, $p = 0.70$, $\eta p2 < 0.01$, or the interaction between voice type and voice accent, $F(1, 77) = 1.81$, $p = 0.18$, $\eta p2 = 0.02$.

In the standardized mental effort efficiency for the transfer tests, there was a significant effect for voice type, $F(1, 77) = 9.52$, $p < 0.01$, $\eta p2 = 0.11$. There was not a significant effect for voice accent, $F(1, 77) = 0.11$, $p = 0.75$, $\eta p2 < 0.01$, or the interaction between voice type and voice accent, $F(1, 77) = 0.24$, $p = 0.63$, $\eta p2 < 0.01$.

In the main effect found for voice type, the human condition showed improvements for both mental effort training and testing over the synthetic condition. This was the case in mental effort training efficiency, mental effort testing efficiency on the retention test, and mental effort testing efficiency on the transfer tests. This shows voice type influenced mental effort efficiency outcomes.

Table 4

Training And Testing Mental Efforts

Condition			Video		Retention Test		Transfer Test	
		N	Mean	SD	Mean	SD	Mean	SD
Russian Accent	Synthetic	22	-0.45	0.82	-0.43	0.85	-0.30	0.66
	Human	19	0.38	1.11	0.41	1.12	0.41	1.17
	Total	41	-0.06	1.04	-0.04	1.06	0.03	0.99
English Accent	Synthetic	22	-0.18	0.79	-0.10	0.71	-0.27	0.66
	Human	18	0.37	0.86	0.22	0.65	0.25	1.04
	Total	40	0.06	0.86	0.04	0.70	-0.03	0.88
Total	Synthetic	44	-0.32	0.80	-0.27	0.79	-0.28	0.66
	Human	37	0.38	0.98	0.32	0.92	0.34	1.10
Total Average		81	0.00	0.95	0.00	0.89	0.00	0.93

Speaker Rating Survey

A series of 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) ANOVAs were conducted on the participants' speaker rating data to determine differences in speaker rating, voice type, and voice accent. A separate ANOVA was conducted on each of the three subscales of the speaker rating survey, superiority, attractiveness, and dynamism. Lastly, an ANOVA was conducted on the mean scores of the three subscales to obtain the overall speaker rating.

There was a significant effect in the superiority rating for voice type, $F(1, 77) = 12.59$, $p < 0.01$, $\eta p2 = 0.14$. There was not a significant effect in the superiority rating for voice accent, $F(1, 77) = 1.99$, $p = 0.16$, $\eta p2 = 0.03$, or the interaction between voice type and voice accent, $F(1, 77) = 0.48$, $p = 0.49$, $\eta p2 = 0.01$.

There was a significant effect in the attractiveness rating for voice type, $F(1, 77) = 13.42$, $p < 0.01$, $\eta p2 = 0.15$. There was not a significant effect in the attractiveness rating

for voice accent, $F(1, 77) = 0.34$, $p = 0.56$, $\eta p2 < 0.01$, or the interaction between voice type and voice accent, $F(1, 77) < 0.01$, $p = 0.97$, $\eta p2 = 0.00$.

There was a significant effect in the dynamism rating for voice type, $F(1, 77) = 7.91$, $p < 0.01$, $\eta p2 = 0.09$. There was not a significant effect in the dynamism rating for voice accent, $F(1, 77) = 0.02$, $p = 0.90$, $\eta p2 = 0.00$, or the interaction between voice type and voice accent, $F(1, 77) = 1.08$, $p = 0.30$, $\eta p2 = 0.01$.

There was a significant effect in the overall speaker rating for voice type, $F(1, 77) = 16.11$, $p < 0.01$, $\eta p2 = 0.17$. There was not a significant effect in the overall speaker rating for voice accent, $F(1, 77) = 0.61$, $p = 0.44$, $\eta p2 = 0.01$, or the interaction between voice type and voice accent, $F(1, 77) = 0.37$, $p = 0.55$, $\eta p2 = 0.01$.

In the main effects found for voice type, the human condition outperformed the synthetic condition in ratings of superiority, attractiveness, dynamism, and overall quality. This shows voice type independently influenced how the participants perceived the virtual human. The virtual humans with human voices were rated higher on the three subscales and overall rating.

Table 5

Speaker Rating Survey

Condition			Superiority		Attractiveness		Dynamism		Overall	
		N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Russian	Synthetic	22	5.68	1.28	5.03	1.74	5.17	1.02	5.29	1.19
	Human	19	6.42	0.86	6.12	0.86	5.56	0.68	6.03	0.64
	Total	41	6.02	1.15	5.53	1.49	5.35	0.89	5.64	1.03
English	Synthetic	22	5.87	1.46	5.21	1.31	4.92	1.29	5.33	1.15
	Human	18	6.96	0.76	6.28	1.11	5.76	0.69	6.33	0.66
	Total	40	6.36	1.30	5.69	1.33	5.30	1.13	5.78	1.08
Total	Synthetic	44	5.77	1.36	5.12	1.53	5.05	1.16	5.31	1.15
	Human	37	6.68	0.85	6.19	0.98	5.65	0.68	6.18	0.66
Total Average		81	6.19	1.23	5.61	1.41	5.32	1.01	5.71	1.05

Agent Persona Inventory - Revised

A series of 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) ANOVAs were conducted on the participants' API-R data to determine differences in API-R, voice type, and voice accent. A separate ANOVA was conducted on each of the four subscales of the API-R; facilitates learning, credibility, human-like, and engagement.

There was a significant effect in facilitates learning for voice type, $F(1, 77) = 12.59$, $p < 0.01$, $\eta p2 = 0.14$. There was not a significant effect in the facilitates learning for voice accent, $F(1, 77) = 1.99$, $p = 0.16$, $\eta p2 = 0.03$, or interaction between voice type and voice accent, $F(1, 77) = 0.48$, $p = 0.49$, $\eta p2 = 0.01$.

There was a significant effect in credibility for voice type, $F(1, 77) = 12.59$, $p < 0.01$, $\eta p2 = 0.14$. There was not a significant effect in credibility for voice accent, $F(1, 77) = 1.99$, $p = 0.16$, $\eta p2 = 0.03$, or the interaction between voice type and voice accent, $F(1, 77) = 0.48$, $p = 0.49$, $\eta p2 = 0.01$.

There was a significant effect in human-like for voice type, $F(1, 77) = 12.59$, $p < 0.01$, $\eta p2 = 0.14$. There was not a significant effect in human-like for voice accent, $F(1, 77) = 1.99$, $p = 0.16$, $\eta p2 = 0.03$, or the interaction between voice type and voice accent, $F(1, 77) = 0.48$, $p = 0.49$, $\eta p2 = 0.01$.

There was a significant effect in engagement for voice type, $F(1, 77) = 12.59$, $p < 0.01$, $\eta p2 = 0.14$. There was not a significant effect in engagement for voice accent, $F(1, 77) = 1.99$, $p = 0.16$, $\eta p2 = 0.03$, or the interaction between voice type and voice accent, $F(1, 77) = 0.48$, $p = 0.49$, $\eta p2 = 0.01$.

In the main effects found for voice type, the human condition outperformed the synthetic condition in ratings of facilitates learning, credibility, human-like, and engagement. This shows voice type independently influenced how the participants perceived the virtual human. The virtual humans with human voices were rated higher on the four subscales.

Table 6

Agent Persona Inventory-Revised Subscales

Condition			Facilitates Learning		Credibility		Human-Like		Engagement	
		N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Russian	Synthetic	22	3.50	0.88	3.71	0.82	3.06	1.27	3.17	1.27
	Human	19	4.00	0.76	4.11	0.54	3.79	0.64	3.74	0.86
	Total	41	3.73	0.85	3.89	0.72	3.40	1.08	3.44	1.23
English	Synthetic	22	3.69	0.56	3.82	0.59	2.83	1.02	3.06	0.99
	Human	18	4.09	0.52	4.34	0.47	3.52	0.87	3.67	0.98
	Total	40	3.87	0.57	4.06	0.60	3.14	1.01	3.34	1.02
Total	Synthetic	44	3.60	0.73	3.76	0.71	2.95	1.14	3.12	1.23
	Human	37	4.04	0.65	4.22	0.52	3.66	0.77	3.71	0.91
Total Average		81	3.80	0.72	3.97	0.66	3.27	1.05	3.39	1.07

Agent Appearance and Voice Questions

A series of 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) ANOVA was conducted on the participants' data of the virtual human's appearance to determine differences in appearance perceptions, voice type, and voice accent. Additionally, a series of 2 (voice type: human, synthetic) x 2 (voice accent: English, Russian) ANOVAs were conducted on the participants' voice data to determine differences in virtual human voice perceptions, voice type, and voice accent. A separate ANOVA was conducted on each of the three voice questions.

There was a significant effect in *I liked the agent's overall physical appearance* for voice type, $F(1, 77) = 19.21$, $p < 0.01$, $\eta p2 = 0.20$. There was not a significant effect on credibility for voice accent, $F(1, 77) = 0.14$, $p = 0.71$, $\eta p2 < 0.01$, or the interaction between voice type and voice accent, $F(1, 77) = 0.00$, $p = 0.98$, $\eta p2 = 0.00$.

There was a significant effect in *voice of the agent was natural* for voice type, $F(1, 77) = 19.21$, $p < 0.01$, $\eta p2 = 0.20$. There was not a significant effect on credibility for voice accent , $F(1, 77) = 0.14$, $p = 0.71$, $\eta p2 < 0.01$, or the interaction between voice type and voice accent, $F(1, 77) = 0.00$, $p = 0.98$, $\eta p2 = 0.00$.

There was a significant effect in *voice of the agent facilitated understanding* for voice type, $F(1, 77) = 7.25$, $p < 0.01$, $\eta p2 = 0.09$. There was not a significant effect on human-like for voice accent, $F(1, 77) = 2.93$, $p = 0.09$, $\eta p2 = 0.04$, or the interaction between voice type and voice accent, $F(1, 77) = 0.01$, $p = 0.93$, $\eta p2 = 0.00$.

There was a significant effect in *agent was easy to understand* for voice type, $F(1, 77) = 11.22$, $p < 0.01$, $\eta p2 = 0.13$. There was not a significant effect on engagement for

voice accent, $F(1, 77) = 2.62$, $p = 0.11$, $\eta p2 = 0.03$, or the interaction between voice type and voice accent, $F(1, 77) = 1.18$, $p = 0.28$, $\eta p2 = 0.02$.

In the main effects found for voice type, the human condition outperformed the synthetic condition in ratings of facilitates learning, credibility, human-like, and engagement. This shows voice type independently influenced how the participants perceived the virtual human. The virtual humans with human voices were rated higher on the four subscales.

Table 7

Agent Appearance and Voice Questions

Condition			Agent's Appearance		Voice was Natural		Voice Facilitated Understanding		Voice Easy to Understand	
		N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Russian	Synthetic	22	3.09	0.97	2.73	1.39	3.00	1.45	3.14	1.42
	Human	19	3.83	0.79	3.84	0.60	3.63	0.90	4.11	0.94
	Total	41	3.42	0.96	3.24	1.22	3.29	1.25	3.59	1.30
English	Synthetic	22	3.50	1.19	2.82	1.40	3.41	0.85	3.73	0.77
	Human	18	4.05	0.71	3.94	0.87	4.00	0.59	4.22	0.43
	Total	40	3.76	1.02	3.32	1.31	3.68	0.80	3.95	0.68
Total	Synthetic	44	3.30	1.09	2.77	1.38	3.20	1.19	3.43	1.17
	Human	37	3.95	0.74	3.89	0.74	3.81	0.78	4.16	0.73
Total Average		81	3.59	1.00	3.28	1.26	3.48	1.06	3.77	1.05

CHAPTER 4

DISCUSSION

Voice type had main effects for all the dependent variables measures apart from the control variables, pretest, and accent familiarity. Across the board, the recorded human voice outperformed the synthetic voice. These findings create interesting discussion questions to be addressed as the predictions based on previous literature prompted the following predictions:

Human-English = Synthetic English > Human-Russian = Synthetic Russian

However, the results of this study concluded with these results for all seven research questions, Human-English = Human-Russian > Synthetic English = Synthetic Russian.

Voice Effect and Learning Outcomes

Research question one (How do voice type and voice accent impact *learning retention*?) and **two** (How do voice type and voice accent impact *learning transfer*?) both postulated that for learning outcomes participants instructed by an accented voice would have reduced outcomes due to the voice effect. This hypothesis was not supported, learning outcomes participants instructed by a human voice scored significantly higher in both the retention and transfer tests.

The results of this study with accented virtual humans provide significant findings of voices with human sources outperforming those with synthetic voices on learning outcomes. This is in direct contrast with the most recent literature involving synthetic voices which showed modern voice engines could perform at the same level as human voices (Craig & Schroeder 2017;2019). The learning outcomes of this study are also inconsistent with previous research findings involving voice accents. The original study

on which voice effect was authored, showed significant learning detriment from instructional voices that spoke in a foreign accent (Mayer et al., 2003). The results of this study do not support the voice effect for accents, with no significant main effects being found for voice accents. The same pattern is found for questions regarding mental effort efficiency. **Research question three** was based on previous findings that an accented voice would be detrimental to mental effort (Mayer et al., 2003), and that a modern voice engine would allow the synthetic voice to perform on par with that of a human recording (Craig & Schroeder 2017;2019). This hypothesis was also not supported as the participants instructed by the human voice had significantly better mental effort training and testing efficiency than the participants instructed by the synthetic human voice. As was the case with the outcomes on retention and transfer, no significant effect was found for the difference between foreign-accented and English-accented voices for mental effort efficiency.

Voice Effect and Perception Outcomes

Research questions four through seven all dealt with perception measures of the speaker, which again the literature supported that the accented voices producing a stronger voice effect would suffer and be rated lower (Mayer et al., 2003). However, the predicted outcomes again did not support the hypothesis that the English-accented voices would be rated higher than the foreign-accented voices. There were no significant effects found for voice accent in any of the perception measures. Instead, the virtual human speaking with a recorded human voice was rated higher than the virtual human speaking in a synthetic voice. This voice type finding was true in the speaker rating survey, the API-R, and direct questions regarding the speaker's voice and appearance. These voice-

type findings went against the previous literature that showed modern voice engines being rated on par with human voices (Craig & Schroeder 2017;2019).

Voice Type

The current study's results on voice type appear to be in direct contrast with the most recent research that reexamined the voice effect with synthetic voice engines. This research has shown evidence that virtual humans paired with synthetic voice engines performed on the same level as human voice recordings (Craig & Schroeder, 2017). Despite the apparent contrasts between the current findings and previous research, the results of this experiment fall in line with a continuing trend within the field. This becomes understandable from the perspective of the *quality of the voice*. The previous works involving voice effects and synthetic voices were reexamined over a decade later, to find that a higher quality voice engine had closed the gap between synthesized voices and their human recording counterparts (Atkinson et al., 2005; Craig & Schroeder 2017;2019). These updated findings came from using a higher quality source of synthesized voices (Neospeech over a Classic voice engine).

The current study is showing a difference with human recordings outperforming because the recorded voices were obtained from professional voice actresses. The previous research used human volunteers with no professional speaking or voice acting training (Craig & Schroeder 2017;2019). The previous research did find that human recordings were on par with a synthetic voice engine. However, the current study's results put a new light on these findings, *the average human recording is on par with a high-quality voice engine*. This should in no way detract from the considerable advancements that voice engines have made or deter anyone from improving upon these

engines. Currently, this study shows that a high-quality (professional) human recording outperforms the high-quality synthetic voice engine. From this perspective, a *quality of voice effect* can be seen that connects the results of this study with the previous research, that a higher quality voice will produce superior outcomes.

Voice Accent

This study's results in the voice accent condition are interesting as they appear to go against previous research at first glance. All previous research based on Mayer's *voice effect* work has repeatedly shown evidence that virtual humans paired with accented voices result in lower learning outcomes and perceptions of the speaker (Mayer et al., 2003). There are multiple explanations as to why this was the case.

The first is that the current study had low power due to participant attrition, with 81 participants in a 2x2 between subject's design. Examining the non-significant means tells an interesting story, with learning results that would look like this: Human-English > Human-Russian > Synthetic-Russian > Synthetic-English. If the study had proper power, the discussion could have been saying that accents create a voice effect for human-recorded voices. However, either this same voice effect is not present in synthetic voices or a novelty effect for the Russian voice is stronger than the impact of the voice effect in synthetic voices.

Second, with globalization increasingly allowing interaction and integration among different people the Russian accent may no longer be as *foreign* as it once was. However, this explanation would require increased power to be a properly supported explanation and for the means of the current findings to become more balanced with this increased power.

The third and final explanation is that the previous studies may have had a quality difference between their foreign-accented and neutral-accented conditions. As the voice type condition showed, there was a *quality of voice effect* when examining the human-synthetic voices, and it is possible that professional voice actresses were on the same level of quality as each other regardless of the accent in which they spoke. Additionally, since both synthetic voices came from the same engine the quality of those voices are of similar quality despite any accent as well. There is the possibility that the voices used in previous voice effect studies involving accents did not have comparable quality between them. If this is the case, then the voice effect would remain consistent as a quality measure across all previous studies.

Limitations

Participant recruitment was conducted over Amazon Mechanical Turk, *MTurk.* This decision allowed for testing the voice accent and voice type variables in the same real-world conditions that distance learning occurs. This decision was not without trade-off, control of the participants' testing condition was unable to be accounted for which could add variance into the analysis. Criticism towards online studies, MTurk in particular, have been lobbied against the practice for providing unreliable data. Examinations of online data collection have shown that there are advantages and disadvantages to using MTurk for research (Goodman et al., 2012; Landers & Behrend, 2015; Paolacci et al., 2010; Peer et al., 2014). Despite these criticisms, research has shown that data samples from MTurk and traditional student samples have no significant differences in performance (Mason & Suri, 2012). Of particular note for the current study, Casler et al. (2013) found that multimedia presentations have no performance

difference between MTurk participants and traditionally collected participants (Casler et al., 2013).

To minimize any potential issues, best practices were implemented to ensure that the data would remain reliable. The risk of reduced attentional effort from participants was lessened by requiring participants have a Human Intelligence Task (HIT) approval rating of 95% or higher (Paolacci & Chandler, 2014). Previous studies have shown that with this practice, MTurk participants were found to have higher attentiveness to instructions than traiditonal subject pool participants (Hauser & Schwarz, 2016). Participants were only able to successfully submit their responses by reaching the end of the study, which provided them with the correct code. The code itself was changed with every eight participants to prevent the code from being leaked online. The data was also screened post-collection to eliminate all participants that plagiarized their answers from an online source.

CHAPTER 5

CONCLUSION

Once the *quality of voice effect* explanation is accepted, this explains the current significant and non-significant accent findings. For the non-significant voice accent, the human recordings were both at a professional level and the synthetic voices were created by the same modern engine, creating no difference in quality across accents for outcomes to deviate. Regarding the significant findings of voice type, the human recordings were of a higher quality than the synthetic voice engines, which created a difference in quality to produce the significant findings. This shows that there is still a *professional voice barrier* the synthetic voice engines have not been able to cross. If multimedia materials are being created, the findings of this study support that the use of professionals to record human voices would yield superior results to that of a synthetic engine. However, access to these professionals is costly and not always a practical option. Smaller projects may not have a budget for hiring these professionals or the content of the learning materials may change frequently to render the expensive recording outdated. For these instances, the findings of this study also support that the use of high-quality synthetic voice engines would be appropriate to use if the options for a human-recording are that of a non-professional.

The findings in the accent condition also support that these learning materials could be used without worry of differing accents hindering learning outcomes. However, the power of this study should be considered for these findings. The low number of participants makes it currently unknown if the non-significant accent findings are a Type II error. Additional studies are in progress to correct this power issue, and a further

understanding of how accents impact learning from virtual humans will soon be provided.

REFERENCES

Ahn, J. (2010). *The effect of accents on cognitive load and achievement: The relationship between students' accent perception and accented voice instructions in students' achievement* (Doctoral dissertation, Ohio University).

Aneesh, A. (2015). *Neutral accent: How language, labor, and life become global*. Duke University Press.

Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology, 30*(1), 117–139. https://doi-org.ezproxy1.lib.asu.edu/10.1016/j.cedpsych.2004.07.001

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences, 17*(1), 102–141.

Baylor, A., & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In *International Conference on Intelligent Tutoring Systems*, 592-603.

Baylor, A., & Ryu, J. (2003). The API (Agent Persona Instrument) for assessing pedagogical agent persona. In EdMedia+ Innovate Learning (pp. 448-451). *Association for the Advancement of Computing in Education* (AACE).

Beck, S., Carr, K., Davis, D. M., Nordhagen, J. N., & Nye, B. D. (2018). Virtual Mentors in a Real STEM Fair: Experiences, Challenges, and Opportunities. In *Third International Workshop on Intelligent Mentoring Systems*.

Bergmann, K., & Kopp, S. (2009). Increasing the expressiveness of virtual agents: autonomous generation of speech and gesture for spatial description tasks. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1* (pp. 361-368). International Foundation for Autonomous Agents and Multiagent Systems.

Blommaert, J. (2009). A market of accents. *Language policy, 8*(3), 243-259.

Boucher, C. J., Hammock, G. S., McLaughlin, S. D., & Henry, K. N. (2013). Perceptions of competency as a function of accent. *Psi Chi Journal of Psychological Research, 18*(1), 27-32.

Burke, S. L., Bresnahan, T., Tan Li, Epner, K., Rizzo, A., Partin, M., Ahlness, R. M., & Trimmer, M. (2018). Using Virtual Interactive Training Agents (ViTA) with Adults with Autism and Other Developmental Disabilities. *Journal of Autism & Developmental Disorders, 48*(3), 905–912. https://doi-org.ezproxy1.lib.asu.edu/10.1007/s10803-017-3374-z

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in human behavior*, *29*(6), 2156-2160.

Cotton, J. L., O'neill, B. S., & Griffin, A. (2008). The "name game": Affective and hiring reactions to first names. *Journal of Managerial Psychology*.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293-332.

Craig, S. D., Chiou, E. K., & Schroeder, N. L. (2019). The Impact of Virtual Human Voice on Learner Trust. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 63, No. 1, pp. 2272-2276). Sage CA: Los Angeles, CA: SAGE Publications.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, *94*(2), 428

Craig, S. D., & Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers & Education*, *114*, 193–205. https://doi-org.ezproxy1.lib.asu.edu/10.1016/j.compedu.2017.07.003

Craig, S. D., & Schroeder, N. L. (2018). Text-to-Speech software and learning: Investigating the relevancy of the voice effect. *Journal of Educational Computing Research*, *57*(6), 1534-1548.

Craig, S. D., Schroeder, N. L., Roscoe, R. D., Cooke, N. J., Prewitt, D., Siyuan, L., ... & Clark, A. (2020). *Science of Learning and Readiness State of the Art Report*. Arizona State University Tempe United States.

Craig, S. D., Twyford, J., Irigoyen, N., & Zipp, S. A. (2015). A test of spatial contiguity for virtual human's gestures in multimedia learning environments. *Journal of Educational Computing Research*, *53*(1), 3–14. https://doi-org.ezproxy1.lib.asu.edu/10.1177/0735633115585927

Davis, R., & Antonenko, P. (2017). Effects of pedagogical agent gestures on social acceptance and learning: Virtual real relationships in an elementary foreign language classroom. *Journal of Interactive Learning Research*, *28*(4), 459-480.

Davis, R. O., Vincent, J., & Park, T. (2019). Reconsidering the Voice Principle with Non-native Language Speakers. *Computers & Education, 140*, 103605.

Deprez-Sims, A., & Morris, S. B. (2010). Accents in the workplace: Their effects during a job interview. *International Journal of Psychology*, *45*(6), 417–426. https://doi-org.ezproxy1.lib.asu.edu/10.1080/00207594.2010.499950

Domagk, S. (2010). Do Pedagogical Agents Facilitate Learner Motivation and Learning Outcomes? Journal of Media Psychology: *Theories, Methods, and Applications, 22*(2), 84–97.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213-224.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior research methods*, *48*(1), 400-407.

Huang, B., Alegre, A., & Eisenberg, A. (2016). A cross-linguistic investigation of the effect of raters' accent familiarity on speaking assessment. *Language Assessment Quarterly*, *13*(1), 25-41.

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, *6*(1), 27–54.

Hew, K. F., & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British journal of educational technology*, *41*(1), 33-55.

Jeno, L. M., Vandvik, V., Eliassen, S., & Grytnes, J. A. (2019). Testing the novelty effect of an m-learning tool on internalization and achievement: A Self-Determination Theory approach. *Computers & Education, 128*, 398-413.

Johnson, W. L., & Rickel, J. (1997). Steve: An animated pedagogical agent for procedural training in virtual environments. *ACM SIGART Bulletin, 8*(1-4), 16-21.

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial intelligence in education, 11*(1), 47-78.

Jones, H. E., Sabouret, N., Damian, I., Baur, T., André, E., Porayska-Pomsta, K., & Rizzo, P. (2014). Interpreting social cues to generate credible affective reactions of virtual job interviewers. *arXiv preprint arXiv:1402.5039*.

Kavas, A., & Kavas, A. (2008). An exploratory study of undergraduate college students' perceptions and attitudes toward foreign accented faculty. *College Student Journal*, *42*(3), 879-891.

Kopp, S. (2006). How people talk to a virtual human-conversations from a real-world application. *How People Talk to Computers, Robots, and Other Artificial Communication Partners*, 101.

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology*, *8*(2), 142-164.

Lippi, R., Donati, S., Lippi-Green, R., & Donati, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press.

Lourdeaux, D., Fuchs, P., Burkhardt, J. M., & Bernard, F. (2002). Relevance of an intelligent tutorial agent for virtual reality training systems. *International Journal of Continuing Engineering Education and Life Long Learning*, *12*(1-4), 214-229.

Louwerse, M. M., Graesser, A. C., Lu, S., & Mitchell, H. H. (2005). Social cues in animated conversational agents. Applied Cognitive Psychology: *The Official Journal of the Society for Applied Research in Memory and Cognition*, *19*(6), 693-704.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, *44*(1), 1-23.

Mayer, R. E. (2002). Multimedia learning. In *Psychology of Learning and Motivation* (Vol. 41, pp. 85-139). Academic Press.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of educational psychology*, *90*(2), 312.

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, *95*(2), 419. https://doi-org.ezproxy1.lib.asu.edu/10.1037/0022-0663.95.2.419

Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, *2*(1), 10-12.

Moreno, R., & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology*, *92*, 724–733.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). 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.

Morozov, M., Tanakov, A., Gerasimov, A., Bystrov, D., & Cvirco, E. (2004, August). Virtual chemistry laboratory for school education. In *IEEE International Conference on Advanced Learning Technologies*, 2004. Proceedings. (pp. 605-608). IEEE.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*, 429e434.

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer's (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 27e42). New York, NY: Cambridge University Press.

Paas, F., van Merrienboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419e430.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63e71.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current directions in psychological science, 23*(3), 184-188.

Park, S. (2015). The Effects of Social Cue Principles on Cognitive Load, Situational Interest, Motivation, and Achievement in Pedagogical Agent Multimedia Learning. *Journal of Educational Technology & Society, 18*(4).

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior research methods, 46*(4), 1023-1031.

Pelachaud, C. (2009). Studies on gesture expressivity for a virtual agent. *Speech Communication, 51*(7), 630-639.

Pilus, Z. (2013). Exploring ESL learners' attitudes towards English accents. *World Applied Sciences Journal, 21*(21), 21.

Reeves, B., & Nass, C. (1996). *The Media Equation*. New York: Cambridge University Press.

Rizzo, A., Lange, B., Buckwalter, J. G., Forbell, E., Kim, J., Sagae, K., ... & Kenny, P. (2011). SimCoach: an intelligent virtual human system for providing healthcare information and support.

Rosenberg-Kima, R. B., Plant, E. A., Doerr, C. E., & Baylor, A. L. (2010). The influence of computer-based model's race and gender on female students' attitudes and beliefs towards engineering. *Journal of Engineering Education, 99*(1), 35-44.

Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., ... & McDonnell, R. (2015, September). A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. In *Computer Graphics Forum* (Vol. 34, No. 6, pp. 299-326).

Riches, P, & Foddy, M. (1989). Ethnic accent as a status cue. *Social Psychology Quarterly, 52,* 197-206. doi: 10.2307/2786714

Rickel, J., & Johnson, W. L. (1999). Virtual humans for team training in virtual reality. In *Proceedings of the ninth international conference on artificial intelligence in education* (Vol. 578, p. 585).

Schroeder, N. L., Chiou, E. K., & Craig, S. D. (2020). Trust influences perceptions of virtual humans, but not necessarily learning. *Computers & Education, 160*, 104039.

Schroeder, N. L., Romine, W. L., & Craig, S. D. (2017). Measuring pedagogical agent persona and the influence of agent persona on learning. *Computers & Education, 109*, 176-186.

Schroeder, N. L., Yang, F., Banerjee, T., Romine, W. L., & Craig, S. D. (2018). The influence of learners' perceptions of virtual humans on learning transfer. *Computers & Education, 126*, 170-182.

Shamekhi, A., & Bickmore, T. (2015). Breathe with me: A virtual meditation coach. In *International Conference on Intelligent Virtual Agents* (pp. 279-282).

Subtirelu, N. C. (2015). “She does have an accent but…”: Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com. *Language in Society, 44*(1), 35-62.

Sweller, J., Cognitive load during problem solving: Effects on learning, *Cognitive Science, 12,* 257-285 (1988).

Tao, P. (2016). An augmented teachable agent for elderly-silver agent. In *SCSE Student Reports,* pp. 60.

Zahn, C. J., & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology, 4*, 113–122

Zoll, C., Enz, S., Schaub, H., Aylett, R., & Paiva, A. (2006, April). Fighting bullying with the help of autonomous agents in a virtual school environment. In *7th International Conference on Cognitive Modelling*.

Zipp, S. A., & Craig, S. D. (2019). The impact of a user's biases on interactions with virtual humans and learning during virtual emergency management training. *Educational Technology Research and Development, 67*(6), 1385-1404.

APPENDIX A

LIGHTNING MATERIALS SCRIPT – LIGHTNING FORMATION

Cool moist air moves over a warmer surface and becomes heated.

Warmed moist air near the earth's surface rises rapidly.

As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud.

The cloud's top extends above the freezing level, so the upper portion of the cloud is composed of tiny ice crystals.

Within the cloud, the rising and falling air currents cause electrical charges to build. The charge results from the collision of the cloud's rising water droplets against heavier, falling pieces of ice.

The negatively charged particles fall to the bottom of the cloud, and most of the positively charged particles rise to the top.

Eventually, the water droplets and ice crystals become too large to be suspended by updrafts.

As raindrops and ice crystals fall through the cloud, they drag some of the air in the cloud downward, producing downdrafts.

When downdrafts strike the ground, they spread out in all directions, producing the gusts of cool wind people feel just before the start of the rain.

A stepped leader of negative charges moves downward in a series of steps. It nears the ground.

A positively charged leader travels up from such objects as trees and buildings. The two leaders generally meet about 165 feet above the ground.

Negatively charged particles then rush from the cloud to the ground along the path created by the leaders. It is not very bright.

As the leader stroke nears the ground, it induces an opposite charge, so positively charged particles from the ground rush upward along the same path.

This upward motion of the current is the return stroke. It produces the bright light that people notice as a flash of lightning.

APPENDIX B

PRETEST QUESTIONNAIRE

Please mark next to the items that apply to you:

1. What causes air to rise from the earth's surface in a thunderstorm?
    a. The sun's rays heat the air and cause it to rise.
    b. Warmer air currents are introduced and heat the existing air.
    c. Cool moist air moves over a warmer surface and becomes heated.
    d. Air high up is cooled and sinks below the air on the earth's surface pushing it upwards.
2. What happens when water vapor condenses?
    a. The water vapor is warmed by the air and rises further.
    b. The water immediately calls back to the earth's surface.
    c. Electrical charges are built up in the water vapor.
    d. The water vapor condenses into water droplets and forms a cloud.
3. Why do water droplets and ice crystals fall?
    a. The water droplets and ice crystals become too large to be suspended by updrafts.
    b. The water droplets and ice crystals become too cold and are forced out of the cloud by rising warm air.
    c. Air currents push the water droplets and ice crystals out of the cloud in which they are suspended.
    d. The droplets and ice crystals become too warm and become too volatile to stay part of the cloud.
4. What occurs when raindrops and ice crystals fall through the cloud?
    a. The raindrops and ice crystals create a conduit for the lightning to flow through.
    b. The friction of the raindrops and ice crystals moving past each other creates electrical charges.
    c. The vacuum created by their absence causes updrafts to fill the space.
    d. They drag some of the air in the cloud downwards, creating downdrafts.
5. What is built up in the bottom of the cloud before a flash of lightning?
    a. A collection of positively charged particles.
    b. A collection of negatively charged particles.
    c. A collection of both positively charged particles and negatively charged particles.
    d. A rapidly changing mixture of positively charged particles and negatively charged particles.
6. What moves downwards to the ground from the cloud?
    a. A leader of negative charges.
    b. A positively charged leader.
    c. A naturally charged leader.
    d. Nothing, both leaders come from the ground.
7. What moves upwards from the ground to the cloud?
    a. A positively charged leader.

b. A leader of negative charges.
c. A neutrally charged leader.
d. Nothing, both leaders come from the cloud.

8. What is happens to the two leaders?
    a. The first leader hits the cloud, followed by the second leader leaving the ground and hitting the ground.
    b. The two leaders go past each other simultaneously.
    c. The first leader hits the ground, followed by the second leader leaving the ground and hitting the cloud.
    d. The two leaders meet in the air.

APPENDIX C

ACCENT FAMILIARITY SCALE

How would you best describe your familiarity with the following non-native accents? For each accent below, please indicate your familiarity with the 1-4 scale below.

Select "1" for: Not familiar at all. I am not able to tell if the person speaks with this particular accent.

Select "2" for: Somewhat familiar. I can sometimes tell whether the person speaks with this particular accent.

Select "3" for: Moderately familiar. I can often tell whether the person speaks with this particular accent.

Select "4" for: Very familiar. I can always tell whether the person speaks with this particular accent.

1. Belarusian-accented English
2. Bulgarian-accented English
3. Czech-Slovak-accented English
4. Lechitic-accented English
5. Macedonian-accented English
6. Polish-accented English
7. Russian-accented English
8. Rusyn-accented English
9. Serbo-Croation-accented English
10. Slovene-accented English
11. Sobian-accented English
12. Ukrainian-accented English.

APPENDIX D

MENTAL EFFORT SCALE

In solving or studying the preceding problems I invested:

1. Very, very low mental effort
2. Very low mental effort
3. Low mental effort
4. Rather low mental effort
5. Neither low nor high mental effort
6. Rather high mental effort
7. High mental effort
8. Very high mental effort
9. Very, very high mental effort

60

APPENDIX E

SPEAKER RATING SURVEY

Superiority									
Illiterate	1	2	3	4	5	6	7	8	Literate
Unintelligent	1	2	3	4	5	6	7	8	Intelligent
Uneducated	1	2	3	4	5	6	7	8	Educated
Not fluent	1	2	3	4	5	6	7	8	Fluent
Inexperience	1	2	3	4	5	6	7	8	Experience
Attractiveness									
Unkind	1	2	3	4	5	6	7	8	Kind
Cold	1	2	3	4	5	6	7	8	Warm
Unfriendly	1	2	3	4	5	6	7	8	Friendly
Unpleasant	1	2	3	4	5	6	7	8	Pleasant
Unlikable	1	2	3	4	5	6	7	8	Likeable
Dynamism									
Passive	1	2	3	4	5	6	7	8	Active
Shy	1	2	3	4	5	6	7	8	Talkative
Unaggressive	1	2	3	4	5	6	7	8	Aggressive
Unsure	1	2	3	4	5	6	7	8	Confident
Lazy	1	2	3	4	5	6	7	8	Energetic

APPENDIX F

API-R SURVEY QUESTIONS

Here are a few questions pertaining to the agent used during the training session. Please select the number that best represents your attitude towards the agent by selecting one of the following options.

Facilitates Learning (subscale)

1. The agent led me to think more deeply about the presentation.
2. The agent made the instruction interesting.
3. The agent encouraged me to think about what I was learning.
4. The agent kept my attention.
5. The agent communicated the main ideas clearly.
6. The agent helped me to concentrate on the presentation.
7. The agent helped me focus on the relevant information.
8. The agent helped me learn the material.
9. The agent was good at teaching.
10. The agent was easy to learn from.

Credibility (subscale)

11. The agent seemed knowledgeable.
12. The agent seemed intelligent.
13. The agent was useful.
14. The agent was helpful.
15. The agent was an effective teacher.

Human-Like (subscale)

16. The agent had a personality.
17. The agent's emotion was natural.
18. The agent was human-like.
19. The agent's movement was natural.
20. The agent showed emotion.

Validation Question

21. To ensure you are paying attention please answer this with "Agree"

Engagement (subscale)

22. The agent was engaging.
23. The agent was enthusiastic.
24. The agent was entertaining.
25. The agent was motivating.
26. The agent was easy to connect with.

APPENDIX G

ADDITIONAL AGENT QUESTIONS

Appearance question

1. I liked the agent's overall physical appearance.

Voice questions

2. The voice of the agent was natural.
3. The voice of the agent facilitated understanding of the message.
4. The agent was easy to understand.

APPENDIX H

IRB APPROVAL

EXEMPTION GRANTED

Scotty Craig
IAFSE-PS: Human Systems Engineering (HSE)
480/727-1006 Scotty.Craig@asu.edu

Dear Scotty Craig:

On 7/26/2021 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Deep learning with virtual agents: How accented and synthetic voices effect outcomes
Investigator:	Scotty Craig
IRB ID:	STUDY00014223
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Accent Familiarity Questionnaire, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Accents Consent Form.pdf, Category: Consent Form; • Agent Perception Survey, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Cognitive Load Measurement, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Demographic Questionnaire.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • IRB Protocol Accents (2).docx, Category: IRB Protocol; • Learning Materials Script.pdf, Category: Recruitment materials/advertisements /verbal scripts/phone scripts;

	• MTurk screenshot.pdf, Category: Recruitment Materials; • Pretest, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Retention Test.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Screenshot of learning video.pdf, Category: Other; • Speaker Rating Survey.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Transfer Tests.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions);

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2) Tests, surveys, interviews, or observation on 7/26/2021.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required. Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

Sincerely,

IRB Administrator

cc: Robert Siegle
Scotty Craig
Robert Siegle