Effect of an artificial model’s vocal expressiveness on affective and cognitive learning.

Llaima Eliza González Brouwer
0999377
MSc. Human Technology Interaction
Department of Innovation Sciences
Eindhoven University of Technology

First Supervisor: Uwe Matzat, Second Supervisor: Sofia Fountoukidou,
Third supervisor: Jaap Ham
All supervisors are part of the HTI Group at TU/e, department of innovation sciences.
August 2018.
This study is an investigation of the effect of an artificial model’s (pedagogical agent) vocal expressiveness, when demonstrating the use of a novel software, on affective and cognitive learning. Pedagogical agents, used as behavior models, proved to be beneficial for self-efficacy beliefs and task learning (Fountoukidou et al., 2017). However, the voice, as a nonverbal communication characteristic, can be further explored (Veletsianos, 2009) resulting in a lack of specific guidelines on effectively designing artificial models. We build on previous research on behavior modeling and immediacy derived from teacher-student interaction (Bandura, 1969; Mehrabian, 1987). The question we aim to answer is; what is the effect of an artificial models’ vocal expressiveness on learning? And whether motivation and attention mediate this effect. We first hypothesized that a vocally expressive pedagogical agent, as compared to a vocally monotonous, will have a significant effect on both affective and cognitive learning. Secondly, we hypothesize that state motivation and attention will mediate the effect. A total of 144 participants, in their majority young students of the TU/e, took part of the 2X1 between-subjects study were the artificial agent, Eric, gave a ten-minute video lesson with the same wording and activity but different voice expressiveness; 1). Vocally expressive (N=78) and 2). Vocally monotonous and flat (N=66). Through specific speech parameters (i.e., speech rate, pitch, emotional prosody) the experimental manipulation was achieved. Results: The difference between the two groups showed to be significant for affective learning in all three measured dimensions. For cognitive learning, there are stronger learning outcomes for the first indicator, but not for the second one, leading to partial support for our hypothesis on cognitive learning. State motivation mediates the effect on affective learning but not on cognitive learning.

Furthermore, attention does not mediate any learning outcome. Thus, the study provided some evidence that an artificial model using higher voice expressiveness affects learning outcomes. The agent appeared on screen demonstrating the task (modeling behavior) and was assessed by all participants as ‘likeable’ and significantly more under the vocally expressive condition. Keywords: Pedagogical (artificial) agent, behavior modeling, nonverbal immediacy cue, vocalics, speech parameters, affective, cognitive learning.
effect of an artificial model’s vocal expressiveness on learning.3

Chapter 1: Introduction

Artificial pedagogical agents (PAs) have been implemented into multimedia learning environments as an attempt to introduce more instructional support and motivational elements. Recent findings indicate that using PAs as behavioral models, thereby following Banduras (1969) social learning theory, increase learning outcomes, compared to other teaching methods (Fountoukidou, Ham, Matzat & Midden, 2017). However, that study focused on the agent’s core behavior modeling characteristics (i.e., task demonstration and verbal instruction), while the agent had limited expressiveness regarding speech and facial expressions (i.e., designed to be monotonous and flat). Bandura (1969), suggested that the way a model carries out a task can influence the degree to which the behavior will influence an observer (i.e., model’s non-verbal behavior). Nonetheless, Bandura and subsequent research have not further investigated the models’ nonverbal behavior, resulting in a lack of specific guidelines on effectively designing artificial models. Thus, in our project, we build on these findings and aim to go one step further and examine whether non-verbal behavior, specifically, the vocal expressiveness (vocal variety in pitch and rate, and prosody) of PAs when acting as behavioral models will result in greater behavior effects (i.e., increased learning).

Earlier literature has shown that non-verbal communication plays an essential role in the process of learning. In fact, traditional education research examining teacher expressive style coined the name immediacy; a term that describes the ability of teachers to create a psychological closeness with their students through non-verbal communication (Andersen 1979). Teachers’ immediate non-verbal behaviors, for example, their vocal expressiveness, impact students’ affective learning (i.e., liking of the tutor and the course) as well as cognitive learning (i.e., recall) (Witt, Wheeless & Allen, 2004).

Moreover, in the field of linguistics, we find the subfield of phonetics which studies the sounds of human speech (i.e., acoustic properties, and auditory perception). More precisely, the concept of prosody in phonetics, directly concerns with the emotional aspect embedded in the voice and the perception by a particular recipient (i.e., systematic variations in pitch, loudness, duration, tempo, and rhythm across words, phrases, and sentences and the emotional intention (e.g., humor or sarcasm) (Pisoni & Remez 2008). Ultimately, prosody, used as immediacy cues, convey speakers’ attitudes and feelings affecting the overall
effect of an artificial model’s vocal expressiveness on learning. The communication dynamic (e.g., teacher-student interaction). A viable explanation for the effect of vocal expressiveness is the mechanism of attention and state motivation which may affect cognitive and affective learning. To date, there is minimal work on the effect of PAs vocal expressiveness on learning outcomes and none on the artificial model’s emotional prosody or voice expressiveness (Johnson et al. 2000). Therefore, the use of artificial pedagogical agents provides an opportunity to study the effect of specific forms of speech parameters on learning outcomes, in a behavioral model setting.

**Research Questions**

A. What is the effect of the vocal expressiveness of an artificial behavioral model (Pedagogical Agent) on learners’ cognitive (i.e., recall) and affective (i.e., motivation and perception) learning outcomes?

B. Which are the underlying mechanisms of the effect of an artificial model’s vocal expressiveness on learners’ affective and cognitive learning?
Chapter 2: Literature review

Pedagogical Agents

Pedagogical agents are a branch of software generated animated characters with the purpose of facilitating human learning processes and/or assisting particular tasks in a multimedia learning environment. (pedagogical agent as instructor (Johnson et al. 2000)). The term “agent” refers to presenting a character on the screen (Erickson, 1997). Nowadays, pedagogical agents can be created through available software’s developed to create 3D customized animated characters realistically and with synchronized movements (lips, facial expressions). Examples of these software’s are: Crazy Talk 8, iClone, Poser 5, Go Animate (Pappas, 2017). Because the technical feasibility exists for the creation of realistic virtual characters with real-time communicative capabilities, the nonverbal communication of agents and also robots have become relevant for research and development of artificial intelligence (Vogeley & Bente, 2010).

Baylor & Kim (2005) reviewed the progress on the topic of pedagogical agents and provided recommendations for further research based on at least ten years of investigation. The review indicated that there is broader interest in agents’ social and affective capabilities to support learners as an additional layer to pedagogical agents’ provision of expert guidance. Initially, the focus relied more on design aspects of these characters with the expectation of some motivational benefit through visual presence. Diverse disciplines and researchers are involved in the improvement of pedagogical agents both as instructors and as interactive companions with responsive capabilities (Baylor & Kim, 2005; Veletsianos & Russel, 2014). Some of these disciplines include instructional design, educational technology and psychology, human-computer interaction media communication, and social psychology (Holz, Dragone, & O’Hare, 2009). In our study, concerning pedagogical agents nonverbal capabilities, it is important to contextualize them in Socio-Cognitive theories which are key for human communication (Veletsianos & Russel, 2014).

1 There are also cloud based (online) applications to create virtual characters such as: alteregos.com, livingactors.com.
Because pedagogical agents take the role of teaching, we will combine ideas about behavioral modeling with ideas of the Implicit Communication Theory (Mehrabian, 1968; 1987). The specific term for verbal and nonverbal communication cues is immediacy (Mehrabian, 1968; 1987; Richmon, Gorham & McCroskey, 1987; Allen (1987); Andersen (1979)). Studies on teachers-students immediacy, classify verbal and nonverbal cues that teachers can use to create psychological closeness which impacts on students learning and motivation.

Concretely, for our study we will make use of one class of nonverbal immediacy named vocalics which refers to cues that come from the use of a teachers’ voice (i.e., vocal variety, vocal expressiveness, pronunciation and fillers (Christophel, 1990; Mehrabian, 1987; Richmon, Gorham & McCroskey, 1987; Andersen, 1979). Vocalics are used in the present study, by the pedagogical agent, to communicate towards the learner.

Veletsianos (2009) approached the pedagogical agent’s expressiveness. The study focused on the impact of the characters verbal expressiveness based on a digitally fabricated voice – speech- to-text synthesized - hypothesizing that agent verbal expressiveness would improve the interaction between pedagogical agents and learners, ultimately enhancing learning outcomes. The results indicated that learners who interacted with an expressive agent did score higher on a post-task exam and rated the agent’s ability to interact higher, than learners who communicated with a non-expressive agent. For the dependents variables, participants were asked to rate their communication with the agent in terms of smoothness, naturalness, and effectiveness. Although the results are insightful for the design of pedagogical agents, were designers are advised to embed verbal expressive qualities in agent implementations, Veletsianos (2009) suggests exploring the issue of agent expressiveness beyond the limited notion presented in his paper.

For further research in the field of artificial pedagogical agents, the suggestion is to focus on new delineations of voice expressiveness (e.g., tone and pitch), along with facial expressiveness, and details of agent expressivity in line with message coherency. To date, no work has been done concerning specific characteristics of pedagogical agents vocal expressiveness. We, therefore, build on this proposition by
effect of an artificial model’s vocal expressiveness on learning.7
testing how a specific set of vocalics of a male human voice, foster learning and impact attention as the underlying cognitive mechanism during the learning process.

Social Cognitive Theory

As stated above, it becomes relevant to insert ideas about these pedagogical agents in social cognitive theories when they act as behavior models and understand how they foster learning and if it is comparable to an ordinary teacher-student interaction.

The Social Cognitive Theory of human development (i.e., a.k.a. Social-learning Theory (SLT)) of 1969, argues that people learn through observing and imitating others. This is why the theory refers to modeling the behavior (Bandura, 1971). Concretely, the theory identifies the psychological processes and mechanisms involved in teaching and learning.

Part of the social-cognitive argument stresses the importance of the continuous interaction between three factors; personal factors (i.e., cognition), environment and behaviors referred to as Reciprocal Causation Model (Bandura, 1969). These factors together contribute to the modeling processes of specific behaviors and knowledge. A model’s primary function is to transmit information to the observers through cues for similar behaviors, the strengthening or weakening of learners’ existing constraints and the demonstration of new patterns of behavior. For example, an individual might learn how to replace a kitchen faucet by watching a video of someone modeling this process. Behavior modeling is rooted in an agentic perspective where one subject is influencing the other to adopt a particular behavior and knowledge. The effect of behavior modeling is a powerful method for education in a wide range of behavioral domains and, particularly in technological adoption (Bandura, 1997; Fountoukidou et. al., 2017).
Behavior modeling is rooted in four primary psychological mechanisms involved in learning as shown in Table 1.

<table>
<thead>
<tr>
<th><strong>Mechanism</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Attention</td>
<td>Observing behavioral skills</td>
</tr>
<tr>
<td>2. Memory</td>
<td>Transforming the observed skills into symbolic codes</td>
</tr>
<tr>
<td>3. Motivation</td>
<td>Getting motivated to continue using them</td>
</tr>
<tr>
<td>4. Production</td>
<td>Practicing the skills physically</td>
</tr>
</tbody>
</table>

Here, we focus on attention and motivation as fundamental underlying mechanisms that a teacher or, in our scenario, an artificial pedagogical-agent, can influence because of nonverbal communication cues. In psychology, motivation is defined as an internal state (sometimes described as a need, desire, or want) that serves to activate or energize behavior and give it direction and ultimately relates to the arousal, and persistence of behavior (Kleinginna and Kleinginna, 1981a; Franken, 2006). For the purpose of the present research we focus on state motivation instead of trait motivation or longer term motivation which would help in persistence.

On the other hand, attention to information was a necessary prerequisite for recall, and cognitive learning is directly linked to memory and recall (Frymier, 1994). Attention is defined as the concentration of mental activity that allows us humans to take in a limited portion of the vast stream of information available from both our sensory world (visual, auditory) and our memory (Matlin, 2014). Attention tasks use both bottom-up and top-down processing. Bottom-up relates to concentrating our mental ability because an interesting stimulus captures our attention in the environment. While, top down refers to concentrating mental ability (paying attention) to a specific stimulus. For our study, with a video lesson narrated by the artificial agent we will consider the task as a top-down processing where the participants learning will have to pay selective attention while ignoring other ongoing info (Matlin, 2014).
effect of an artificial model’s vocal expressiveness on learning.9

Therefore, we hypothesize that both attention and motivation play a role while learning during the agent-delivered behavioral model.

Earlier research by Fountoukidou et. al., (2017) discusses the implementation of an artificial agent-delivered behavior modeling as an instructional approach to foster the adoption of innovative technologies and how the agent’s behavior demonstration impacts on motivation and learning. The results of that study confirmed that participants in the agent-delivered behavior modeling showed an increase in the self-efficacy beliefs about the specific system, compared to participants in the two non-modeling treatments (i.e., Agent-delivered instructional narration and No-agent, text-only instruction) this effect remained even after controlling for their general self-efficacy beliefs. Those results are in line with past research (i.e., Gist et al., 1987, Compeau & Higgins 1995a, 1955b) on the effect of behavior modeling on users’ computer self-efficacy compared to non-modeling methods (i.e., lecture training and self-manual). Specifically, participants in the agent-delivered behavior modeling had significantly higher specific software self-efficacy beliefs, as compared to participants in the agent-delivered instructional narration condition and compared to participants in the text-only instruction condition. These findings suggest that an artificial pedagogical agent that acts as a behavioral model improves motivation, computer self-efficacy believes and impact users’ computer skills declarative knowledge (Fountoukidou et. al., 2017). We aim to build on this finding and test whether the vocal expressiveness of the pedagogical agent-delivered behavior modeling will increase specific affective and cognitive learning outcomes by increasing motivation and attention.

Immediacy & Learning

The Implicit communication theory by Mehrabian (1969) states that messages are constantly transmitted through a mixture of verbal and non-verbal communication behaviors known as immediacy. Immediacy, is usually studied and contextualized in the traditional teacher-student interaction, and defined as the extent to which particular cues enhance psychological closeness (i.e., liking or disliking) towards the content and teacher. The verbal aspect in communication weights approximately 20% (Shelly Jones, 2017) and it refers to the content itself and the selection of vocabulary by the teacher. As examples, verbal immediacy include ownership statements (my/our class), inclusive references (we vs. I) and probability
effect of an artificial model’s vocal expressiveness on learning.10

(will v. may) statements, (Rubin, Palmgreen, & Sypher, 1994). On the other hand, and with a higher weight of approximately 80% is the nonverbal component in communication (Mehrabian, 1981). Nonverbal immediacy, refers to the ability of the instructors to convey affective feelings of warmth, closeness, and belonging (Richmond, Gorham, & McCroskey, 1987). Examples of nonverbal immediacy include the use of eye contact, body position, physical proximity, body movement, facial expressions and voice expressiveness (Richmond et. al., 1987; Jordan, 1990; Andersen, 1979). At its foundation, teacher immediacy is based on elements of motivational theory (Gorham & Zakahi, 1988).

Learning.

By definition, learning is the process of acquiring new, or modifying existing, knowledge, behaviors, skills, values, or preferences. For the purpose of this study, in the context of artificial-tutor-student interaction in a multimedia environment, we focus on affective and cognitive learning. Affective learning involves affect (i.e., momentary emotional state) of the person towards the particular topic or lesson and affect towards the teacher. Affect can be traced as motivation, psychological closeness and the attitude that the student feels towards the instruction or teacher and has been shown as the behavioral gateway that leads to motivation and attention which leads to the ultimate goal of cognitive learning (Gorham, 1988). Cognitive learning, based on aspects of Cognitive Learning Theory (Bruner, 1966), implies remembering through memory, building and incorporating schemas within the student's mind.

Several studies (Richmond et. al., 1987; Mehrabian 1978; Christophel, 1990; Gorham, 1988) found positive correlations between a human teacher’s immediacy behaviors and cognitive learning where smiling, vocal expressiveness, a relaxed body position and verbal cues were shown to have a high positive correlation with learning and, therefore, factors influencing students learning processes. Pintrich & De Groot (1990) suggest that teachers need to develop an understanding of appropriate teacher behaviors which could enhance students’ motivation. For the process of cognitive learning one important mechanism is attention as described previously.

Interestingly, a particular research on teachers’, nonverbal immediacy found that both low and high levels of teacher immediacy had negative effects on student state motivation and actual cognitive, affective,
and behavioral learning (Comstock, Rowell & Bowers, 1995). In their study teachers' behavior exhibited varying levels of nonverbal immediacy, the conclusions is that moderately high teacher nonverbal immediacy is more effective in helping students learn than either excessively high or low immediacy. Nevertheless, most studies supports a positive linear relationship between teachers immediacy and students learning (Plax, Kearney, McCroskey & Richmond, 1986; Christensen & Menzel, 1998; Kelley & Gorham, 1988).

Those behaviors taken by teachers that enhance closeness in the interaction with their students and influence instructional outcomes (Andersen and Jensen, 1979) are the use of eye contact, body position, physical proximity, humor, body movement, facial expressions and voice expressiveness (Richmond et. al., 1987; Jordan, 1990; Andersen, 1979).

In the present study, our interest is on the nonverbal cue of voice or what we call speech expressiveness.

The Voice

Our capacity of using the voice and therefore vocally expressing ourselves, is an innate human characteristic. We use our voices in speech, conversation, crying, singing, yelling, laughing, coughing, yawning or whispering. Fundamentally the voice is used for both verbal and nonverbal communication. Verbally we can convey a message through words and phrases while non-verbally it suggests our age-group, gender and emotional states (Kring & Bachorowski, 1999).

Anatomically, the human voice is produced in the larynx, an organ in the top of the neck which houses the vocal cords. Mechanically, whenever air comes out of the lungs, it makes the vocal cords vibrate producing a wave of sound which is controlled by the muscles in the vocal cords. These muscles modulate and adjust the length and tension of the vocal folds to tune the critical parameters of voice which are pitch and tone (Stevens, 2000).

Most people have the ability to modulate the parameters of the voice source consistently and automatically. Interestingly, each person has a unique voice not only because of the anatomical differences between people but also, the manner in which the speech sounds are formed and articulated (Stevens, 2000).
effect of an artificial model’s vocal expressiveness on learning.12

The voice has extensively been explored in the fields of acoustics (i.e., speech-sound generation in the human vocal system), singing voicing, voice projection, intelligibility, linguistics (i.e., phonetics), phonology, psychology interested in sound perception and production voice pedagogy and speech expressiveness (Abercrombie, 1967). The latter is the perspective we approach by this study.

**Speech Parameters**

Speech is the vocalized form of communication used by us humans. Words are created through a phonetic combination of a limited set of vowel and consonant speech sound units called phonemes. These create vocabularies, the syntax that structures them and the semantics behind the structures differ widely across planet Earth, creating thousands of different human languages (Levelt, 1999). Moreover, speech forms as a combination of parameters which include the tone of voice (i.e., pitch), rate (i.e., words per minute) volume (i.e., decibels) and the emotional intention (i.e., prosody).

The pitch parameter is an integral part of the human voice which changes the emitted sound with the rate of vibrations. Therefore, faster rates form higher voices, or higher pitches, while slower rates elicit deeper voices, or lower pitches (Reiman, 2013; Rothenberg, 1967). The classification of vocal ranges based on pitch and tones are commonly defined for singers (e.g., sopranos, tenors, baritones). The definition is based on the fundamental frequency a tone has. By definition in wave mathematics, the fundamental frequency (F0) is the lowest frequency produced by the oscillation of the whole. In this case, the whole of the voice, is distinct from the harmonics of higher frequency (Fant, 2012). The fundamental frequency and formant frequencies are probably the most important concepts in speech synthesis and also in speech processing in general (Lemmetty, 2004). Therefore, the voice can be treated as a speech signal that behaves as a wave of sound and complies to the properties of waves, i.e., speed, amplitude, frequency, wavelength, period (French, 1971). In this sense, the essence of speaking is its continual flexibility, variability, rhythm and adaptability.

The perception of fundamental frequency and corresponding harmonics is commonly known as *voice pitch* and is the parameter and trait that influences aspects of social interaction such as attractiveness and sexual selection (Feinberg, 2008). There is empirical evidence on the link between pitch and voice
effect of an artificial model’s vocal expressiveness on learning.13

attractiveness where on average men prefer high-pitched women's voices to low-pitched women's voices across a set of tested frequencies. On average, women preferred men's voices lowered in pitch, but did not prefer very low men's voices (Feinberg, 2008). The fundamental frequency of a male voice is typically agreed to be 120 Hz (Hsiao, Solomon, Luschei, & Titze, 1994; Baken, 2000).

Pitch and all the speech parameters relate to emotional prosody, characterized as an individual's tone of voice when talking that is conveyed through variations in pitch, loudness, timbre, speech rate, and pauses which is different from linguistic and semantic information (Buchanan, Mirzazade, Specht, Shah, Zilles & Jäncke, 2000). The prosodic attributes of intonation (i.e., pitch), intensity (i.e., loudness), and rhythm (i.e., speech rate) reflect emotional states of the speaker as elements of language that are not encoded in grammar or literal vocabulary such as, question, statement, exclamation, the presence of irony or sarcasm; emphasis, contrast, and focus (Buchanan et. al., 2000; Pisoni & Remez, 2008).

Therefore, speech expressiveness is defined as the combination of these speech parameters which as outcome show the prosodic attributes used by a speaker, resulting in the voice dynamics, the feelings and intentions the voice is exposing. In this study, we will use concrete variations (i.e., manipulations) of the speech parameters resulting in two conditions; a) vocally expressive (i.e., enthusiast prosody) b) vocally unexpressive (i.e., monotonous and flat prosody). So, the nonverbal immediacy cues of the tutors’ voice are embedded in the speech parameters.

The elements of speech used for this study as defined in earlier literature (Mehrabian, 1987; Mizuno & Nakajima, 1998; Buchanan et. al., 2000; Pisoni & Remez, 2008;) are shown in table 2.

<table>
<thead>
<tr>
<th>Vocalics</th>
<th>Prosodic attribute</th>
<th>Description</th>
<th>Measurement Unit</th>
<th>Instrument of Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch (fundamental frequency).</td>
<td>Intonation</td>
<td>Degree of highness or lowness of a tone determined by the vibration of the vocal folds (i.e., the faster vibration per second)</td>
<td>Hz</td>
<td>Praat (Software by Paul Boersma y David)</td>
</tr>
</tbody>
</table>
The effect of an artificial model’s vocal expressiveness on learning.

<table>
<thead>
<tr>
<th>Speech rate</th>
<th>Rhythm and articulation (fluency, emphasis, stress, pauses)</th>
<th>Word count on Microsoft Office. Total amount of words divided by the duration of each video in minutes.</th>
</tr>
</thead>
</table>
| Duration and tempo across words, phrases and sentences leading to fluency. For our study we need the correct articulation of words. Conversational speech generally falls between 120 wpm at the slow end, to 160 - 200 wpm in the fast range (Susan Dugdale, 2018). Slow speech is usually regarded as less than 110 wpm. | [WPM] Word per minute which can also be calculated to Syllables per second [SPS] | Note. For this study, volume or loudness was not manipulated and is maintained in an ideal range for both experimental conditions [60-75 dBA].

Note. For this study, volume or loudness was not manipulated and is maintained in an ideal range for both experimental conditions [60-75 dBA].
effect of an artificial model’s vocal expressiveness on learning.

Here, we isolate the pedagogical agent’s speech expressiveness parameters (i.e., pitch, rate and consequently emotional prosody) and measure the effect on affective and cognitive learning outcomes. We hypothesize that moderately high vocal expressiveness used by the artificial modeler instructor will lead to significantly higher learning outcomes than low vocal expressiveness.

**Hypotheses:**

i. Participants who are confronted with a virtual pedagogical agent who uses enhanced vocalics and vocal variety (variety in pitch and pace) during a learning tutorial will show significantly higher affective learning when compared to participants who are confronted with a virtual pedagogical agent who uses a moderately low degree of vocalics and vocal variety.

ii. Participants who are confronted with a virtual pedagogical agent who uses enhanced vocalics and vocal variety (variety in pitch and pace); will show significantly higher cognitive learning (measured in a recall test) as compared to participants who are provided with a virtual pedagogical agent who uses a moderately low degree of vocalics and vocal variety.

iii. Participants level of state motivation and attention to the artificial modeling task instructions will mediate the effect of the type of vocal expressiveness (expressive versus monotonous) on affective learning.

iv. Participants level of state motivation and attention to the artificial modeling task instructions will mediate the effect of the type of vocal expressiveness (expressive versus monotonous) on cognitive learning.

Comentado [1]: Grammar changed
effect of an artificial model’s vocal expressiveness on learning.16

Chapter 3: Method

This chapter describes the experimental setup and method used for the study; this includes a description of the population and design, procedure, materials, the selected measurements for the dependent variables, their scale reliabilities and the description of experimental control variables.

Population and Design

The target population for this study consisted of people registered in a local database (JFC), and most of them are part of the Eindhoven University of Technology (TU/e), either students or employees. The only restriction for being part of the study was not having participated in previous Gaze The Web studies (Fountounkidou et. al., 2017) to avoid any pre-knowledge on the topic under evaluation. No further restrictions were involved. The study employed a between-subjects design, with the participants being randomly assigned to one of the two experimental conditions and therefore two groups:

1. Vocally expressive agent-delivered behavior modeling
2. Vocally unexpressive (i.e., monotonous and flat) agent-delivered behavior modeling.

The study’s dependent variables are affective and cognitive learning. Overall, the duration of the study was approximately 30 minutes, for which participants received a 5€ compensation for their participation and a 7€ if they were external to TU/e.

Procedure

Participants were welcomed in the central hall of the lab building. The first step is registering as a participant in the “artificial tutoring” experiment in the official TU/e ARCHI database. Secondly, each participant was asked to read and sign an informed consent form, stating the general purpose of the research, the compensation, risks and ultimately their willingness to participate in the study (see Appendix C). Thirdly, participants were seated in a cabinet with a PC a large screen and headphones plugged in. On the screen, the welcome message of the survey. Once they pressed ‘next’ after the welcome screen, the instructional video was ready to be started.
effect of an artificial model’s vocal expressiveness on learning.17

Participants were instructed to start the video themselves (YouTube embedded video), enlarge it to full screen and high quality, while the volume pre-set at ~50% for every PC. In both versions of the video a virtual agent provided a short tutorial on how to use a new interface for a web browser, called Gaze the Web (GTW). The video shows the agent-delivered-behavior in two screens; on the right side the virtual agent named Eric explaining and demonstrating what appears on the left side of the screen which is the GTW interface and functionalities.

For a visual example see Materials; figures 1 and 2 which show sample screenshots for the procedure. Each version of the ~10-minute video lesson contained identical wording and demonstrations by the agent because the manipulation was on the voice expressiveness only.

Subsequently at the end of the instructional video, participants’ were requested to answer the online questionnaire, as also a multiple choice test. After this, they were debriefed, thanked for their contribution and compensated.
effect of an artificial model’s vocal expressiveness on learning.

Materials

Content.

The content of the instruction pertained to a novel eye tracking software, called GazeTheWeb (GTW). GTW, illustrated in Figure 2, is a web-browser, developed to be controlled just with the use of the eyes movement, using an eye-tracking hardware (for more information see Kumar, Menges & Staab, 2017). To access the video lessons refer to Appendix C.

Pedagogical Agent.

The 3D animated pedagogical agent, named Eric, who teaches the instruction was created using the Software CrazyTalk 8 (https://www.reallusion.com/crazytalk/). The agent was designed to resemble participants’ characteristics in terms of appearance, according to the guidelines derived from the earlier literature (Rosenberg-Kima, Baylor, Plant & Doer, 2008). Since the majority of the participants are students at a Dutch University, the agent was designed to be young (<30 years-old), a real appearance, be attractive (as manipulated by the agent’s facial features) and “cool” (as manipulated by the agent’s clothing and hairstyle). Refer to figure 1 and 2 or Appendix C for the full video lesson.

Figure 1: Eric, the Artificial Agent guiding the video lesson.
effect of an artificial model’s vocal expressiveness on learning. 19

Figure 2: On the right hand side is Eric the artificial agent teaching and demonstrating what is to be seen on the left hand side within Gaze the web environment.

Voice.

The manipulation of speech parameters, as the nonverbal immediacy cue for the experimental conditions, were accomplished through systematic use of the nonverbal immediacy cues derived from immediacy research and theory (Andersen 1979; Gorham, 1988; Jordan 1990; Mehrabian, 1968,1969,1981; Richmond et. al., 1987). We controlled for the objective acoustic speech parameters; pitch and rate (Software Praat, 2018), refer to Table 2:Parameters of speech in Chapter 2.

The male voice of Eric, our pedagogical agent, was recorded by an actor following a script (Appendix A) in a home-studio and saved as a MP3 file. The length of the tutorials reached approximately 10 minutes therefore, we controlled for tiredness of the actor by recording 10 parts of ~1 minute each for both experimental condition, resulting in 20 recorded parts in total.

The actor was chosen because of the neutral English accent and clear pronunciation. The same actor performed the two versions of the instruction following a script. To control for pitch and speech rate variations we constantly measured the minimum and maximum pitch for each ~1 minute audio part providing a mean pitch or fundamental frequency. This was visualized on the spot during the recordings through Audacity, the software used for the voice recordings (Audacity, 2018). Finally, once the recordings
effect of an artificial model’s vocal expressiveness on learning.20

were completed, we used Praat software to explore the speech variations in depth. The speaking fundamental frequency (F0) for a male adults is agreed to be 120 Hz (Hsiao et. al., 1994, Hollien & Shipp, 1972) therefore, the pitch boundary we set to divide between expressive and monotonous was 120 Hz.

Summary of Speech parameters per condition,

1) Vocally Expressive: pitch >120 hz, speech rate ~ 133 WPM (words per minute), prosody; enthusiast.
2) Vocally Unexpressive (i.e., Monotonous): pitch <120 hz, speech rate ~119 WPM (words per minute), prosody; neutral).
effect of an artificial model’s vocal expressiveness on learning.21

Pitch Analysis.

Pitch as defined in the literature review (Chapter 2), is generally measured as the fundamental frequency of the sound wave. To accurately measure this for both of our experimental conditions we took the average pitch between ten recordings for each condition, and analyzed the fragment of the wave in terms of their pitch. According to the Voice Academy (2018) males often speak at 65 to 260 Hertz. The fundamental frequency for male speaking is agreed to be 120 Hz and, is therefore the boundary number that separates the experimental conditions between expressive and monotonous based on spoken language processing studies (Hsiao, 1994; Hollien & Shipp, 1972; Mizuno & Nakajima, 1998).

Table 3:
Pitch Analysis (Praat output)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Minimum Pitch</th>
<th>Maximum Pitch</th>
<th>Mean Pitch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expressive</td>
<td>134 Hz</td>
<td>386 Hz</td>
<td>~260 Hz</td>
</tr>
<tr>
<td>Monotonous</td>
<td>87 Hz</td>
<td>143 Hz</td>
<td>~115 Hz</td>
</tr>
</tbody>
</table>

Note. One minute average sample values for pitch based on ten audio extracts, one for each explained section by the artificial agent (Refer to Appendix A). Analysis performed by Praat Software and STATA.

Note that for the pitch analysis, Praat software requires a maximum of 10 seconds of the audio file to get; maximum pitch, minimum pitch and the mean value of pitch in the selected audio extract. For our study, we decided to take ten samples of 10 seconds for each condition. The samples were randomly chosen, however, for both conditions the same 10 seconds of audio file to maintain consistency (i.e., ten first seconds of each audio extract). For each section of the script that the voice actor followed we created one audio file (i.e., refer to sections on Appendix A, to access the audio files refer to appendix A.1). To perform the average calculations of pitch we used STATA/IC 14, adding all the maximum pitch values for each section of the script and divided it by ten, the same procedure was applied for both experimental conditions.
effect of an artificial model’s vocal expressiveness on learning.

An example of the followed procedure to achieve the average values on table 3 through Praat Software refer to the graphical output below. The graph draws the audio wave for a ~10 [s] audio extract for both conditions respectively. (i.e., audio section number 2, refer to section 2 in Appendix A)

Maximum Pitch Analysis

Expressive,

![Figure 3.1. Maximum Pitch graphical analysis. Expressive_2 audio cut: 227.98 Hz (maximum pitch in SELECTION), Mean Pitch 128.79 Hz (mean pitch in SELECTION).](image)

Monotonous,

![Figure 3.2. Monotonous audio cut: 140.3468 Hz (maximum pitch in SELECTION), 105.2138 Hz (mean pitch in SELECTION).](image)
effect of an artificial model’s vocal expressiveness on learning.23

Minimum Pitch Analysis

Expressive,

Figure 4.1. Minimum graphical pitch analysis. Expressive_2 audio cut: 80.76 Hz (minimum pitch in SELECTION), Mean Pitch 128.79065656268975 Hz (mean pitch in SELECTION).

Monotonous,

Figure 4.2. Minimum Pitch analysis. Monotonous_2 audio cut 87.74 Hz (minimum pitch in SELECTION), 105.2138 Hz (mean pitch in SELECTION).
effect of an artificial model’s vocal expressiveness on learning.24

**Speech rate Analysis.**

Speech rate is the term given to the speed at which you speak. It's calculated by the number of words spoken in a minute. A normal number of words per minute (wpm) can vary hugely. Studies show speech rate alters depending on the speaker's culture, geographical location, subject matter, gender, emotional state, fluency, profession or audience (Ngiam, Charumilind & Zhenghao, 2017). Some agreed guidelines are that conversational speech generally falls between 120 wpm at the slow end, to 160 - 200 wpm in the fast range (Susan Dugdale, 2018). Slow speech is usually regarded as less than 110 wpm.

Table 4:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Speech duration (minutes)</th>
<th>Number of words</th>
<th>Words per minute (wpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Expressive</td>
<td>9:30</td>
<td>1240</td>
<td>133</td>
</tr>
<tr>
<td>2. Monotonous</td>
<td>10:20</td>
<td>1240</td>
<td>119</td>
</tr>
</tbody>
</table>

Speech rate analysis

*Note.* Words per minute were calculated through word count tool on Microsoft Office divided by the total duration of the video lesson.

As shown in table 4, the values of speech rate were within the range of conversational speech for both conditions. The difference among expressive and monotonous of fourteen words is, in fact, not too high. For the manipulation, we wanted to achieve a somewhat regular conversational rate for a video lesson using moderately high vocal nonverbal immediacy in the expressive condition and moderately low vocal nonverbal immediacy in the monotonous condition.
effect of an artificial model’s vocal expressiveness on learning.

**Questionnaire.**

The design and implementation of the questions were achieved using Limesurvey (limesurvey.com). The platform allowed the deployment of the full experiment; a welcome screen, the randomization for the two experimental conditions and the groups of questions based on scales defined in previous literature both perception (i.e., affective) and recall (i.e., cognitive). Lastly, demographic questions on gender, education, and level of computer use were asked. Refer to Appendix B for the full questionnaire.
effect of an artificial model’s vocal expressiveness on learning.26

**Measurements**

The two main dependent variables for the first hypothesis are students’ affective learning and cognitive learning. The questionnaire had two sections. The first set of thirty-seven questions aimed at measuring individuals’ perception of voice nonverbal immediacy, attention, motivation and affective learning while the following twenty-eight questions aimed at recall as the cognitive learning measure. For the first part on affective learning, motivation and attention five main variables were identified as shown in table 5 (to find the specific questions refer to appendix A by the question codes exposed on table 5). To compose the explanatory variables respondents answered the questions in a differential semantic-type of scale ranging from 1 being negative to 7 being positive. All answers were based on participants’ perception and feelings towards the aforementioned variables. An exception are the indicators of cognitive learning (see below).
effect of an artificial model’s vocal expressiveness on learning.27

### TABLE 5
AVERAGE MEASURES OF NONVERBAL IMMEDIACY AND AFFECTIVE LEARNING

<table>
<thead>
<tr>
<th>Scale</th>
<th>Measure</th>
<th>Reference Scale</th>
<th>Items (Appendix B)</th>
<th>Number of items in scale</th>
<th>Scale Reliability (correlation r, Cronbach α, n=144)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b. Vocal expressiveness</td>
<td>Nonverbal immediacy cues of vocalics (NVI ( Mehrabian (1981), Richmond, et. al., (1987), LeFebvre et. al., (2014).)</td>
<td>B1[item5;6;7]</td>
<td>3</td>
<td>α = 0.83</td>
</tr>
<tr>
<td></td>
<td>c. Vocal Attractiveness</td>
<td>Teacher voice assessment (Servilha, et. al., (2015).)</td>
<td>C1[item1;2;3] +</td>
<td>4</td>
<td>α = 0.82</td>
</tr>
<tr>
<td>2. Artificial Agent</td>
<td>a. Anthropomorphism</td>
<td>Godspeed questionnaire (Bartneck, C., Croft, E., Kulic, D., 2008)</td>
<td>D1[items 1;2;3;4;5]</td>
<td>5</td>
<td>α = 0.83</td>
</tr>
<tr>
<td></td>
<td>b. Animacy</td>
<td></td>
<td>D1[items 5;6;7] +</td>
<td>5</td>
<td>α = 0.79</td>
</tr>
<tr>
<td></td>
<td>c. Likeability</td>
<td></td>
<td>E1[items 3;4;5;6]</td>
<td>4</td>
<td>α = 0.93</td>
</tr>
<tr>
<td>3. Affective Learning</td>
<td>a. Liking of the Content</td>
<td>Semantic differential-Scale (Witt &amp; Wheeless (2001); Andersen, 1979)</td>
<td>F1[items 1;2;3]</td>
<td>3</td>
<td>α = 0.86</td>
</tr>
<tr>
<td></td>
<td>b. Liking of the Teacher</td>
<td></td>
<td>F2[items 1;2;3]</td>
<td>3</td>
<td>α = 0.88</td>
</tr>
<tr>
<td></td>
<td>c. Likelihood of future similar tutoring by the same teacher</td>
<td></td>
<td>G1[1;2;4]</td>
<td>3</td>
<td>α = 0.90</td>
</tr>
<tr>
<td>4. Motivation</td>
<td>State Motivation</td>
<td>Trait motivation scale (Christopher, 1990)</td>
<td>H1 items [1-9]</td>
<td>9</td>
<td>α = 0.91</td>
</tr>
<tr>
<td>5. Attention</td>
<td>Attention</td>
<td>Attention (Fountoukidou et. al., 2017 ; Yi &amp; Davis, 2003)</td>
<td>I1 items [1-4]</td>
<td>4</td>
<td>α = 0.86</td>
</tr>
</tbody>
</table>
effect of an artificial model’s vocal expressiveness on learning.28

**Voice Immediacy.**

The instrument used for the answers measuring voice immediacy (i.e., as nonverbal immediacy cue) was a bipolar semantic differential-type scale. The purpose for constructing these voice immediacy variables was to check whether the experimental manipulation worked correctly. Participants were asked to assess ‘the instructor Eric’ on specific attributes of the voice (e.g., vocal variety, speech rate, friendliness) from 1 to 7 (e.g., The instructor Eric Uses a monotonous voice | 1|2|3|4|5|6|7) [Uses vocal variety]. This scale provides composite values that act as additional variables for exploring the model explaining overall learning through the nonverbal immediacy cue of vocalics.

The reliability and internal consistency is tested in terms of either the correlation or Cronbach’s reliability coefficient between single items that assess similar traits (Refer to table 5 for scientific references).

Specifically,

a. **Vocal Variety**, initially it was assessed by three questions (i.e., refer to table 5 and Appendix B), the first item was excluded of the scale because the value showed low correlation among the other two items ($r < 0.3$) thus, it was not a reliable measure for the scale. Also, by our criteria - when revising the question - it was not entirely focused on the variety of pitch but rather the perception of high or low pitch. Therefore, only two items (i.e., B1 [2, 3]) measured vocal variety (number of items: 2, Correlation $r = 0.47$).

b. **Vocal expressiveness**, assessed by three questions (i.e., refer to table 5 and Appendix B) that composed the items for this scale (number of items: 3, Cronbach’s $\alpha = 0.83$).

c. **Vocal attractiveness**, assessed by four questions (i.e., refer to table 5 and Appendix B) that composed the items for this scale (number of items: 4, Cronbach’s $\alpha = 0.83$).
effect of an artificial model’s vocal expressiveness on learning.

**Artificial Agent.**

The “Godspeed” questionnaire (Bartneck, Croft, Kulic, 2008) was used to measure three key concepts of Human-Computer interaction, namely, animacy, anthropomorphism, and likability. This questionnaire was administered in a 7-point semantic differential, scale. We constructed reliable measures of anthropomorphism (Cronbach’s α = .83), animacy (Cronbach’s α = .79), and likeability (Cronbach’s α = .93) by averaging participants’ answers to each set of questions for n=144. Although the perception towards the pedagogical agent was not part of the hypotheses of the present study, it is a relevant indicator for the context of the video lesson where an artificial agent is the tutor.

**Affective Learning.**

Affective learning was measured with the semantic differential-type scale developed by Andersen (1979) and Scott and Wheels (1975). The scale for affective learning implied asking participants to answer three groups of questions regarding liking towards the content, liking towards the tutor and their likelihood of following this tutor on similar content if available. They lead to three types of affective learning. The instrument used for the answers was bipolar semantic differential-type scale. (e.g., “I feel the content of the video is : bad/good, worthless/valuable, negative/positive). We constructed reliable measures of content affective learning (Cronbach’s α=0.86) , teacher affective learning (Cronbach’s α=0.88 ) and their likelihood of future similar tutoring² (Cronbach’s α = 0.90).

All measures consist of the arithmetic mean of the values of the items.

² For the likelihood of following future similar tutoring, items G1 [1 and 3] (Appendix B ) were measuring the same trait so, we decided to drop one of these items when constructing the scale, namely item number 3.
effect of an artificial model’s vocal expressiveness on learning.30

**Perceived learning.**

In order to measure participants perceived learning a "learning loss" score was then computed by subtracting the score on the first 1 to 7 scale from the score on the second 1 to 7 scale, indicating the students’ overall perceived learning score. The specific questions were; G3. *How much do you think you could have learned from this video had you had this ideal teacher?* – G2. *How much did you learn during the video lesson?* (Witt et. al. 1990; Richmond, Gorham, & McCroskey, 1987). For question codes refer to Appendix B. Perceived learning through the ‘learning loss’ indicator was measured in order to have an indicator of individuals’ perception of learning and examine differences between the experimental conditions.

**Motivation.**

State motivation was measured through nine items, on a differential semantic scale from 1 to 7, adapted from Christophel (1990) which construct the state motivation scale (Cronbach’s α=.91). The aim was to assess participants’ state motivation, thus their motivation at the precise moment.

**Attention.**

To measure participants attention we used four items (i.e., refer to group II; items 1 to 4 on the questionnaire Appendix B) based on previous research by Fountoukidou et al., (2017) and Yi & Davis (2003). The reliability of the scale was high with Cronbach’s α=0.86. The average value between these four first questions were used for the mediation analysis in the results chapter. Additionally, we asked two extra self-constructed questions regarding divided attention with the goal of controlling for how much of their perceived attention was given to both the video lesson and explanation or only to the demonstration. These questions were: *a) I paid more attention to the GTW software activity than to the explanation given by Eric* and, *b) During the instructional video I paid equal attention to both visuals and the instructor’s example* (i.e., refer to group II items 5 and 6 on Appendix B ).
effect of an artificial model’s vocal expressiveness on learning.

Cognitive learning.

Immediate recall was measured as an index of cognitive learning using a modified cloze procedure based on Taylor (1956) and specific multiple choice questions related to the content of the video. The cloze procedure is an accepted means of assessing both the readability of a variety of written materials and the reading ability of a variety of subjects. The essence of the cloze procedure is the random deletion of words from a text, the subjects then being required to replace those words. The more words a subject replaces exactly, the greater his reading ability (Robinson, 2006). In our study, we modified this cloze procedure to an open-ended section where the participants must recall exact words, synonyms or the basic concept related to what they heard the pedagogical agent say and fill the empty gap in the text.

The test consisted of two parts; a) nine open-ended questions that ask for information about the functionality of GTW and b) eighteen multiple-choice questions. Each item had a value of either zero (0) for incorrect answers or one (1) for correct answers. Thus, providing a scale ranging from a minimum of 0 to a maximum of 9 when open-ended questions and 0 to 18 when multiple choice questions.

a. Open ended questions: The first nine questions were open questions designed to make the participant recall synonyms, concepts or literal extracts of the video lesson watched.

Questions from [K1-M2] (e.g., L1. The GazeTheWeb cursor has the shape of ……………………..), refer to Appendix B. Two researchers reviewed the answers given by participants and agreed on the criteria to assess them as correct (i.e., = 1) or incorrect (i.e., = 0). The inter-rater level of agreement between the researchers was 89%. Note. that one observation - under the monotonous condition - was eliminated because it was a non-valid answer remaining a total of 143 observations.

b. Eighteen questions were multiple choice format (MC) with a single correct answer. The instructions, emphasized that if a participant did not recall the answer please to answer the alternative “I don’t know” in each MC question. Questions from [N1-S3], refer to appendix B.
Chapter 4: Results

In the results chapter, several statistical analysis using STATA /IC 14 were ran over the full data set we collected during the two week experiment sessions\(^3\) (N=144). Two sections are described; 1) Descriptive results and, 2) Hypothesis testing through the mean comparison on the dependent variables of the two groups (two tailed t-test) and the mediation path analysis based on Baron & Kenny (1986).

Descriptive Results

A brief demographic overview (i.e., Questions from [T1-W1], refer to Appendix B) indicated that of the total of hundred-forty-four (n=144) participants who completed the study, fifty-five were females (38%) and eighty-nine of them males (62%). Participants were mainly University students between the ages of 19-26 while only a few of them (<10%) were older. Ninety-two participants (63%) are University educated (i.e., BSc, MSC, Ph.D.), forty-five (31.2%) of them are high school educated (i.e., applied sciences HAVO, Gymnasium, VWO) and seven (4.86%) either chose “do not want to tell” or chose the option “something else”. Hundred eighteen participants (82%) reported to spend more than twelve hours using a computer per week and sixteen (11%) between ten and twelve hours, only ten participants (7%) spend less time using a computer per week.

There were two sample groups randomly assigned to either one or the other condition; 1). Seventy-eight (N=78) received the vocally expressive treatment versus, 2). Sixty-six (N=66) of the participants received the vocally monotonous treatment. Assumptions of normality and homogeneity of variances were met for both groups. Moreover, the clarity of pronunciation of the artificial agent (i.e., B1 item 4 in Appendix B) was assessed by both groups as relatively high on the 1 to 7 scale, a simple t-test revealed no significant difference, between the two groups with \(M_1=6.11 (SD=1.27, n=78), M_2=5.77 (SD=1.21, n=66)\) and \(t(142)=1.64, p=0.1, 95\%CI[5.4-6.4]\). These results show that the pronunciation was clear for the two experimental conditions confirming that the manipulation in the experiment went as expected.

\(^3\) At IPO building Eindhoven University of Technology, between June 4\(^{th}\) – June 15\(^{th}\) 2018.
effect of an artificial model’s vocal expressiveness on learning.33

**Voice Immediacy.**

Independent two-tailed tests measured the mean differences on the vocal expressiveness parameters.

Perceived **vocal variety** was $M_1=4.07$ ($SD=1.14; n=78$) for the expressive condition and $M_2=2.47$ ($SD=0.91; n=66$) for the monotonous condition compared to an overall mean of $\bar{X}=3.34$ ($SD=1.3, n=144$).

To investigate whether the two groups differed in average vocal variety assessment we conducted a t-test with vocal variety as the dependent variable. The effect of vocal variety by the group was significant $t(142)=9.11, p<0.001, 95\% CI[2.42-4.32]$, showing that participants assessed vocal variety higher under the vocally expressive condition.

Estimated **vocal expressiveness** was tested similarly. For the expressive condition, it was $M_1=4.18$ ($SD=1.3, n=78$) while on the monotonous condition $M_2=2.83$ ($SD=0.99, n=66$) compared to an overall $\bar{X}=3.56$ ($SD=1.3, n=144$). The t-test revealed a significant difference between the means of the two groups; $t(142)=6.87, p<0.001, 95\% CI[2.59-4.48]$ showing that vocal expressiveness was assessed higher under the expressive condition.

Lastly **vocal attractiveness** showed $M_1=4.89$ ($SD=1.01, n=78$) and $M_2=3.73$ ($SD=0.96, n=66$) respectively compared to an overall $\bar{X}=4.36$ ($SD = 1.14, n = 144$). The t-test provided significant differences between the two groups $t(142)=6.9, p<0.001, 95\% CI[3.4-5.12]$. Thus, there was indeed a clear difference in the assessment towards agent’s vocal attractiveness.

Thus, all speech parameters confirm the correct experimental manipulation.

**Artificial Agent.**

Because we assessed how participants rated the artificial pedagogical agent in terms of animacy, anthropomorphism and likeability we explore these results overall and by group. As shown in table 6 , participants (n=144), perceived Eric - the artificial agent -, as having an average level of anthropomorphism $M_{anthro}=3.43$ ($SD=1.12$), as being fairly animated $M_{animacy}=3.57$ ($SD=1.02$), and, as highly likeable $M_{likeability}=5.04$ ($SD=1.1$) in a 1 to 7 scale (i.e., Refer to D1, E1 questions in Appendix B ).
The effect of an artificial model’s vocal expressiveness on learning.

Furthermore, all three measures were significant when comparing the mean of the two groups (i.e., 1) expressive v/s 2) monotonous) meaning that the treatment of our experiment must have impacted on the perception of the agent. Normality and homogeneity of variances were all met.

Table 6
Mean Artificial Agent assessment scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total sample (n=144)</th>
<th>Expressive (n=78)</th>
<th>Monotonous (n=66)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M1</td>
</tr>
<tr>
<td>Anthropomorphism</td>
<td>3.43</td>
<td>1.12</td>
<td>3.68**</td>
</tr>
<tr>
<td>Animacy</td>
<td>3.57</td>
<td>1.02</td>
<td>3.80**</td>
</tr>
<tr>
<td>Likeability</td>
<td>5.04</td>
<td>1.1</td>
<td>5.4**</td>
</tr>
</tbody>
</table>

Although Eric looked exactly the same in both conditions and was demonstrating the same task the voice manipulation created a difference in the perception towards him.
effect of an artificial model’s vocal expressiveness on learning.35

Learning loss.

The measure of learning loss (i.e., \( G2/G3 \), Appendix B) was also significantly different for both groups where in the expressive condition people in average had significantly less learning loss. Results of the two-tailed test show that \( M=0.41 \) (\( SD=1.03 \), \( n=78 \)) and \( M2=0.83 \) (\( SD=1.03 \), \( n=66 \)) \( t(142)=-2.3, p < 0.05 \). These results can further confirm – as previous research on teacher-student immediacy (Witt & Wheless, 2001) – that vocalics are an important aspect that an instructor can use when aiming at facilitating learning of their students and, impacts on the perception of learning loss even for a short video lesson.

Motivation

First, for a better understanding of our results in the present study, we had to recode the scale for state motivation. The original scale measured, in a 1 to 7 scale, that lower values were higher state motivation and, conversely, higher values were lower state motivation\(^4\). Therefore we generated a new variable called ‘positive state motivation’ (i.e., [7 - state motivation]). Refer to group H1 in of the questionnaire on Appendix B.

Overall positive state motivation was \( M=3.27 \) (\( SD=1.2 \), \( n=144 \)) and the distribution behaved normally. The two-tailed test showed significant differences between the mean of the two groups, \( M1=3.57 \) (\( SD=1.21 \), \( n=78 \)), \( M2=2.9 \) (\( SD=1.1 \), \( n=66 \)), \( t(142)=-3.44, p<0.01 \). These results suggest that state motivation was higher for participants under the expressive condition.

\(^4\) Three of the items, namely item4,6 and 9 (refer to H1 in Appendix B) were measuring, in a 1 to 7 scale, that lower values were lower motivation and higher values higher motivation. We corrected for this mistake first by recoding these variables as all the others and after this we calculated the positive state motivation scale (i.e., 7 - state motivation values).
effect of an artificial model’s vocal expressiveness on learning.36

Attention

Overall attention distribution proved to behave normally with $a M = 4.8$ ($SD = 1.2, n = 144$). The two-tailed test show significant differences between the groups, $M1 = 5$ ($SD = 1.2, n = 78$), $M2 = 4.56$ ($SD = 1.3, n = 66$) and $t(142) = -2.15, p < 0.03$.

Both items assessing divided attention (i.e., Refer to II item 5 and 6 on Appendix B) were not significantly different for both groups and did not contribute to any essential conclusion for the present study, however, overall their mean values were $M_{demonstration} = 5$ ($SD = 1.6, n = 144$) and $M_{voice + demonstration} = 3.68$ ($SD = 1.84, n = 144$). So, we can infer that, on a 1 to 7 scale, all participants assessed having paid more attention to the demonstration provided by the artificial agent than an equal level of attention to both the demonstration and the explanation given by Eric. Apparently the behavior modeling aspect of the video lesson was perceived as more relevant.

Hypotheses Testing

Because our study had only one manipulation which was, the speech parameters on the voice of the artificial agent, we tested the first hypothesis through comparisons of means between the two sample groups (i.e., t-test) and, analyzed the difference among them on affective and cognitive learning. After worth we examine if motivation and attention are mediators.

In line with the first hypothesis on affective learning, the findings revealed a significant difference between the means of the two independent samples (i.e., condition 1 and 2). For content and teacher affective learning and, also for the likelihood of taking a similar tutoring (Refer to table 7). Normality and equality of variances properties were met (i.e., histogram and skewness).
effect of an artificial model’s vocal expressiveness on learning.37

Table 7
Summary of affective learning scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total sample (n=144)</th>
<th>Expressive (n=78)</th>
<th>Monotonous (n=66)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M1</td>
</tr>
<tr>
<td>Liking of the Content</td>
<td>5.36</td>
<td>1.04</td>
<td>5.57**</td>
</tr>
<tr>
<td>Liking of the Teacher</td>
<td>4.95</td>
<td>1.22</td>
<td>5.36**</td>
</tr>
<tr>
<td>Likelihood of similar</td>
<td>3.91</td>
<td>1.6</td>
<td>4.4**</td>
</tr>
<tr>
<td>tutoring</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. All estimates are in unstandardized units (7-point scale, ranging from 1 to 7). **p<0.01.

As shown in table 7, the mean scores for participants under the monotonous condition are lower for all the variables that measured affective learning and, significantly so when performing a two-tailed test comparing the experimental conditions (i.e., 1. expressive and 2. monotonous). Therefore, we can conclude that the experimental treatment increased affective learning in all three dimensions; content, teacher, and likelihood of future similar tutoring.

For the second hypothesis on cognitive learning, however, we found significant differences between the two groups on the mean of the total score for the open-ended questions, however, no significant difference between the two groups for the multiple choice questions. Results are exposed in table 8. Normality and equality of variances properties were met.
effect of an artificial model’s vocal expressiveness on learning.

Table 8
Mean Cognitive Learning Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Condition</th>
<th>Total sample (n=144)</th>
<th>Expressive (n=78)</th>
<th>Monotonous</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>LL</td>
<td>UL</td>
</tr>
<tr>
<td>Cognitive Learning</td>
<td></td>
<td>3.9</td>
<td>1.6</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Multiple choice</td>
<td></td>
<td>7.5</td>
<td>1.2</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

Note. LL: lower limit, UL: upper limit. One observation under monotonous condition for cognitive learning open-ended questions was lost due to an empty answer. *p<0.05.

For the multiple choice part of the test results suggest there are no differences between the two groups. The scale for multiple choice ranged from scores between [0-18], the mean for both groups was close to 10, which indicates an average of ~50% of recall, independent of which experimental condition participants were in.

The scale for the open-ended questions ranged from scores between [0-9], the mean for both groups was between [7-8], where for the expressive condition it was one point higher and significantly so; $M=7.97$ ($SD=2.41$, $n=78$), $t(142)=2.26$, $p<0.02$. Thus, people under the vocally expressive experimental condition did, recall better what the artificial agent said.

Because results for cognitive learning are conflicting, we can interpret that for the open-ended part people had to recall the words or synonyms they heard from the agent invoking their auditory memory more so than their visual memory, and the treatment of a vocally expressive agent might have impacted on how well they could retrieve the information.
effect of an artificial model’s vocal expressiveness on learning.

**Mediation analysis.**

We performed a two path mediation analysis based on Baron & Kenny (1986) and Cohen (1988) and further examination on the proposed mediators of state motivation and attention. All mediating variables were standardized for a better interpretation of the model except for the variable on the experimental treatment. The analysis of direct and indirect effects was performed using the experimental treatment of the artificial model (agent’s) voice expressiveness and, their effect on the dependent variables; affective learning and cognitive learning, third and fourth hypothesis respectively. The studied independent variable was the treatment they were in: 1) expressive (i.e., Group = 1), 2) monotonous (Group = 0). First, we look at affective learning path analysis.

5 Coded in STATA/IC 14 as Group=1 if expressive condition and Group=0 if monotonous condition.
effect of an artificial model’s vocal expressiveness on learning.40

I. Affective Learning

1. Content Affective Learning. The analysis by linear regression showed that the treatment of a vocally expressive agent (i.e., variable ‘Group’ either monotonous (coded Group=0) or expressive (coded Group=1)) was a significant predictor of content affective learning \(t(142)=2.69, p<0.01, R^2=0.04\) (i.e., refer to patch c in figure 3). Thus, only a 4% of the variance can be explained by the voice expressiveness treatment. The mediation stages are explored in the following sections for each proposed mediator: state motivation and attention (i.e., refer to figure 3).

Figure 3. Mediation analyses of participants content affective learning due to the treatment of behavior modeling participants were in; agent vocally expressive (group=1) or monotonous (group=0). All estimates are in standardized units (7-point scale, ranging from 1 to 7). \(a1b1, a2b2\) based on sobel test and 1000 bootstrap samples (bias corrected). *p<.05, ** p < 0.01 the dotted lines and numbers without* are not significant p>0.05.

1.1 State Motivation, the mediation for state motivation is significant, the indirect paths (i.e., \(a1b1\) in figure 3) as well as each stage of the mediation (i.e., \(a1\) and \(b1\)
effect of an artificial model’s vocal expressiveness on learning.41 respectively) thus, there is some evidence on content affective learning being mediated by state motivation due to the experimental treatment. Moreover, the explained variances are tested through regressing each mediation path to investigate the explained variance explained by each mediation stage. The regression for state motivation as the independent variable and ‘group’ (experimental treatment) as the dependent variable, \( t(142) = 7.07, p < 0.01, R^2 = 0.3 \), exposes that a ~30% of the variance is explained by state motivation. The second mediation stage (path b1) explains a 13% of the variance on liking towards the content. 

1.1.2 Attention, the mediation for attention is non-significant meaning that the indirect paths (i.e., a2b2 in figure 3) are non-significant. However, each stage of mediation (i.e., a2 and b2) are significant by themselves suggesting that attention is important both for the treatment and content affective learning. However, we cannot conclude there is mediation by attention. Overall, attention explains a 26% of the variance on content affective learning with \( t(142) = 7.93, p < 0.01, R^2 = 0.26 \).

The direct effect when the mediators are included (\( c' = 0.071 \)) demonstrates the remaining effect when we controlled for attention and state motivation and is non-significant suggesting complete mediation through state motivation. The combined effects of indirect paths is significant which hints on that the mediators together can explain part of the effect of vocal expressiveness on content affective learning (i.e., \( c = c' + \sum abf \)).
effect of an artificial model’s vocal expressiveness on learning.42

1.2 Teacher Affective Learning, the analysis by linear regression showed that the treatment of a vocally expressive agent (i.e., variable ‘Group’ either monotonous (coded Group=0) or expressive (coded Group=1)) was a significant predictor of teacher affective learning $t(142)=4.62, p<0.01, R^2=0.13$ (i.e., refer to patch c in figure 4). Thus, an 13% of the variance can be explained by the voice expressiveness treatment on teacher affective learning. The mediation stages are explored in the following sections for each proposed mediator; state motivation and attention.

Figure 4. Mediation analyses of participants teacher affective learning due to the treatment of behavior modeling participants were in; agent vocally expressive (group=1) or monotonous (group=0). All estimates are in standardized units (7-point scale, ranging from 1 to 7). $a_{1b1}$, $a_{2b2}$ based on Sobel test and 1000 bootstrap samples (bias corrected).*p<.05, ** p < 0.01 the dotted line and numbers without* are not significant p>0.05.

Comentado [4]: I changed some coef. In the diagram
1.2.1 State motivation:

The first stage of mediation ($a_1 = .277$) means that under the expressive condition participants get a higher state motivation and, significantly so (i.e., refer to $a_1$ in figure 4) with $p=0.000$. The second stage of mediation ($b_1 = .432$) shows that state motivation impacts on teacher affective learning, and is significant (i.e., refer to path $b_1$ in figure 4). The total indirect path (i.e., $a_1b_1$) is also significant thus, we can conclude that there is evidence for mediation by state motivation on teacher affective learning and, our third hypothesis is thereby supported (i.e., path $a_1$ and $b_1$ in figure 4). To test the proportion of variance explained by each path of the mediation we performed a regression for each path based on . The overall effect of state motivation on teacher affective learning is $t(142)=8.51, p<0.01, R^2=0.337$, thus a ~34% of the variance on teacher affective learning can be explained by state motivation. And, the regression for state motivation with 'group' (experimental treatment) $t(142)=3.44, p<0.001, R^2=0.07$ thus a 7% of the variance of state motivation is explained by the vocal expressiveness of the agent.

1.2.2 Attention:

The first stage of mediation ($a_2=0.177$) means that under the expressive condition participants get higher level of attention scores and, significantly so (i.e., refer to path $a_2$ in figure 4). The second stage of mediation ($b_2=0.137$) indicates that attention impacts teacher affective learning, however, not significantly (i.e., path $b_2$ figure 3). The indirect path through attention (i.e., $a_2b_2$ figure 4) is non-significant. We can therefore conclude that attention is not mediating the effect on teacher affective learning. Thus, our fourth hypothesis that attention mediates teacher affective learning is not supported for this indicator of affective learning. The direct effect ($c' = 0.217$) demonstrates the remaining effect when we controlled for attention and state motivation and it is significant hinting on total mediation. The combined
effect of an artificial model’s vocal expressiveness on learning.

The effects of indirect paths is also significant which hints on that the mediators together can explain part of the effect of vocal expressiveness on teacher affective learning (i.e., $c = c' + \sum a_{ij}b_j$) or partial mediation thus, state motivation and attention accounts for some, but not all, of the relationship between the experimental treatment of voice expressiveness and liking towards the teacher.

Moreover, overall ‘attention’ (iv) shows as a significant predictor for teacher affective learning (dv) when regressing only these two variables and, explains ~20% of the variance ($t(142)=6.06$, $p<0.01$, $R^2=0.205$. Therefore, although attention is not mediating the effect it is a predictor for liking towards the teacher.
effect of an artificial model’s vocal expressiveness on learning.

1.3 **Likelihood of similar tutoring**, the analysis by linear regression showed that the treatment of a vocally expressive agent (i.e., variable ‘Group’ either monotonous (coded Group=0) or expressive (coded Group=1)) was a significant predictor of the likelihood of similar tutoring as measure for affective learning \( t(142)=4.13, p<0.01, R^2=0.107 \) (i.e., refer to patch c in figure 4). Thus, ~11% of the variance can be explained by the voice expressiveness treatment. The mediation stages are explored in the following sections for each proposed mediator; state motivation and attention.

![Diagram of mediation analysis](https://example.com/diagram.png)

*Figure 5. Mediation analyses of participants likelihood of future similar tutoring due to the treatment of behavior modeling participants were in: agent vocally expressive (group=1) or monotonous (group=0). All estimates are in standardized units (7-point scale, ranging from 1 to 7). \( a1b1, a2b2 \) based on sobel test and 1000 bootstrap samples (bias corrected). *p<.05, ** p < 0.01 the dotted line and numbers without* are not significant p>0.05.*
1.3.1 **State Motivation**, as shown in figure 5, there is evidence for mediation through state motivation (i.e., refer to path a1, b1 and a1b1) because of the experimental treatment of voice expressiveness on the likelihood of taking similar tutoring in the future. Overall, when regressing state motivation (iv) with the likelihood of future tutoring (dv) a ~48% of the variance is explained by state motivation with $t(142)=11.55$, $p<0.01$, $R^2=0.484$. When regressing state motivation (iv) with Group (dv) we get a ~7% variance of the vocal expressiveness treatment is explained by state motivation.

1.3.2 **Attention**, as shown in figure 5, there is no evidence for mediation through attention (i.e., refer to path a2b2) although attention seems to affect each path independently (i.e., refer to path a2 and b2 respectively). Overall the direct path (i.e., refer to path $c'$) when the mediators are included is significant and we can conclude that the proposed mediators indeed mediate the effect on the likelihood of similar tutoring to some extent. Overall state motivation explains a ~29% of the variance on the likelihood of following similar tutoring in the future with $t(142)=7.60$, $p<0.01$, $R^2=0.289$.

In summary, the result of the mediation analysis suggests that state motivation does mediate to some extent the effect on affective learning in all three dimensions; content, teacher and their likelihood of taking similar tutoring in the future. On the contrary, there is no evidence for attention mediating the effect. Therefore our third hypothesis is supported with respect to the hypothesized mediation by motivation, but not by attention.
effect of an artificial model’s vocal expressiveness on learning.47

2. Cognitive Learning

Because for cognitive learning we only had significant results for the open-ended questions, as explained in the hypotheses testing section above, we will only perform a mediation analysis on the output of these set of questions and therefore drop the multiple choice section. Similarly as for affective learning we will perform mediation for the two proposed mediators. The analysis by linear regression showed that the treatment of vocal expressiveness (i.e., variable ‘Group’ either monotonous (=0) or expressive (=1)) used by the artificial agent was a significant predictor of cognitive learning open-ended questions but it does not explain much of the total variance only a 3%, $t(141)=2.26$, $p<0.02$, $R^2 = 0.03$. In other words, the effect of the predictor ‘Group’ on dependent variable ‘cognitive learning open-ended’ is significant (refer to c in figure 6).

Figure 6. Mediation analyses in participants level of cognitive learning open-ended questions due to the agent’s vocal expressiveness treatment of behavior modeling. All estimates are standardized units (7-point scale, ranging from 1 to 7). $a1b1$, $a2b2$ based on 1000 bootstrap samples (bias corrected). *$p<0.01$, **$p<.05$, dotted lines and numbers without* are not significant $p>0.05$. 
effect of an artificial model’s vocal expressiveness on learning.48

2.2 **State Motivation:** There is a significant effect in the first path from group to state motivation (refer to path a1 in figure 8), meaning there is some effect of the vocal expressiveness treatment on state motivation but not for the second indirect path (refer to path b1). We can therefore not conclude that there is mediation through state motivation on cognitive learning. Moreover, overall state motivation does not explain the variance on cognitive learning with $t(141)=0.96, p=0.340$, $R^2 = 0.006$.

2.3 **Attention:** While there is a significant effect of the treatment on attention (refer to path a2 in figure 6) there is not a significant effect of attention on cognitive learning. What can be explained by the mediators (i.e., refer to $\Sigma a_ib_j$) is non-significant, therefore there is no mediation. Moreover, attention does not significantly explain the variance on cognitive learning (i.e., open-ended questions), $t(141)=1.90$, $p=0.06$, $R^2 = 0.0249$.

In summary, the mediation analysis for cognitive learning shows no mediation by attention nor state motivation on cognitive learning, providing thus no support for the fourth hypothesis.

Overall, the mediation analyses show an effect of state motivation on affective learning but not on cognitive learning while attention did not mediate the effects on both affective and cognitive learning.
Chapter 5: Discussion

The principal goal of this study was to determine the effect of the vocal expressiveness of an artificial model on learning, specifically, affective (i.e., liking towards content and tutor) and cognitive (i.e., recall) learning. The secondary goal was to understand whether state motivation and attention mediated this effect. We performed an empirical study with hundred-forty-four participants that are part of a database from the Department of Innovation Sciences of the Technical University of Eindhoven in the Netherlands, and the majority were students below the age of thirty. The experiment consisted of a ten-minute video lesson where the artificial pedagogical agent explained and demonstrated the functionalities of a novel software called Gaze The web, which is a browser that can be controlled solely with a person’s eye movements. We based our hypothesis on earlier literature on immediacy (Mehrabian, 1987; Gorham, 1988, Andersen, 1979), behavior modeling (Albert Bandura (1969) and artificial pedagogical agents (Heidig & Clarebout, 2009; Veletsianos, 2009; Kim & Baylor, 2016; Fountoukidou et. al., (2017)). The main idea is that a vocally expressive artificial model, compared to a vocally monotonous artificial model, would significantly increase both affective and cognitive learning and that state motivation and attention mediated this effect.

This investigation demonstrated that voice expressiveness impacts the overall affective learning of participants in the three measured dimensions; liking towards the content, liking towards the teacher and, their likelihood of following similar tutoring in the future. These results suggest that through a stronger use of speech parameters by the pedagogical agent, the learners get more motivated at the moment (i.e., state motivation) and, therefore, have higher affective learning compared to learners confronted with an agent who has a low use of speech parameters were the voice was monotonous and flat. Thus, there is evidence for mediation through state motivation on affective learning or, stated differently, on immediate emotional reactions towards learning. However, attention did not mediate this effect although we know that attention and focus is critical in the process of learning.
effect of an artificial model’s vocal expressiveness on learning.

For cognitive learning, there is some evidence to conclude that the sound of voice has an effect. In our case, the analyses using the indicator relying on the open-ended questions showed a significant difference between the two groups (i.e., vocally expressive vs vocally monotonous). For the other indicator of cognitive learning that relies on multiple choice questions, however, all participants scored relatively low and similar independent of the experimental condition - about a ~50% of recall – and, there were no significant differences between the groups. These conflicting results are maybe because the multiple choice section of the test was detailed on Gaze The Web procedures and icons where participants had trouble remembering exactly how to proceed, and this impacted on all participants. In reality, when people learn through behavior modeling, they would be able to ‘produce’ the behavior, using trial and error, and not only recalling what was shown which makes a difference for cognitive learning. In turn, for the open-ended section participants had to retrieve what they heard the agent say, where our manipulation had an effect, leading to higher scores on recall. Neither state motivation or attention mediated the effect on cognitive learning.

The investigation can conclude that the perception of animacy, anthropomorphism, and likeability towards the artificial agent and the vocal expressiveness, had an impact on affective learning. Moreover, the impression towards the realness and likeability of the agent was significantly higher when the agent was vocally expressive. There is some evidence for cognitive learning when retrieving extracts of what they heard the agent said. Thus the study hints at the importance of an expressive voice when teaching to increase or support education.

The limitations of this study rely on the cognitive test of the task where learners could not reproduce what was shown by the agent (i.e., behavioral modeling) and this might impacted the exact measure for cognitive learning, we only measured their recall where in reality people would be able to reproduce and show behaviorally what they learned. Also, a more diverse group of participants regarding age and education, results could have been more explicit especially for cognitive learning. Diversity matters because multimedia learning environments have a broad societal reach, with no strict territorial or demographic limitations, hence, a diverse group of participants would reflect reality more accurately. Also, this study
effect of an artificial model’s vocal expressiveness on learning.

was only half an hour whereas real knowledge and education implies a longer-term process where voice expressiveness and all immediacy cues a teacher or agent uses become more relevant and critical. Additionally, the voice belongs to a ‘real’ man and, nowadays, investigations are more in the line of digitally fabricated voices based on speech synthesis and artificial intelligence as the artificial production of human speech (Nilsson, 2014; Sokol, & Flach, 2018; Mizuno, (1998); Cimen, Yuan Sumner, Coros & Guay, 2018; Allen, Hunnicutt, Dennis, 1987). Digitally fabricated voices are being researched, developed and implemented by IT industry leaders for the creation of new assistive technologies as for example, Alexa by Amazon or Siri by Apple, who can respond to spoken commands. Intelligent virtual characters are relevant in the present (Cole et al., 2018; Zibrek, Kokkinara, & McDonnell, 2018) and large interdisciplinary conferences are held to present research on modeling, developing and evaluating intelligent virtual agents (IVAs) with a focus on communicative abilities and social behavior (ACM International Conference, 2018). For future studies, it could be of importance to consider these results for synthetic voices, based on artificial intelligence, in the context of multimedia learning environments. Additionally, further investigate how the design of the virtual character overall affects learning.

Although, intuitively the voice of a teacher or instructor is essential for attention and motivation it was interesting to test this empirically and demonstrate this effect on learners. The central learning here is that voice expressiveness used by any teacher, and specifically, an artificial tutor's voice is highly relevant for state motivation of the learners and their perception towards the agent which can have an effect on how people relate to what and how they are learning. Surely, behavior modeling is a crucial aspect when learning a procedural task but, the way it is explained and also presented influences and can change the learning outcome.
effect of an artificial model’s vocal expressiveness on learning.52

Acknowledgements

This study was funded by the Innovation Sciences department for the Human Technology Interaction group at TU/e.

I would like to acknowledge supervisors, family, and friends who were, directly and indirectly, involved in the fulfillment of this project, for all of you my deepest thank you.

To my Supervisors, Uwe Matzat: ever since I started the HTI master in 2016, I was captivated by the field of Sociology and how the theories and frameworks related to the actual human technology context we are living. The classes you teach (i.e., HTI in the social context, Social media, and life online and Network Society) and, tutoring through the master thesis process helped me learn and opened a new lane for me. Also, your constant support, push, accuracy and wittiness largely contributed to the result. Sofia Fountoukidou: thank you for your daily support and contributions to this study during the process which at times was rough. You showed me how to do proper scientific research, and dive deep into details. Jaap Ham, I enjoyed and learned from the courses you teach on social psychology which also were interesting for me as I had never learned on psychology before. Additionally, thank you for the support provided during the thesis process, especially, when Uwe was absent for a few weeks.

To my Family, Mienek Brouwer and Raphael Jenner, were my rock during my time living in Eindhoven supporting me emotionally all the way through. To my parents and brother, Jose Gonzalez, Jani Brouwer and Raul, because of you, this master of Science happened since you always pushed me to go further academically and use my opportunities wisely. Finally, to my partner Luis Hinrichs who day to day was supporting me, when in the distance by heart and technology. I love you all. The friends I made at TU/e and outside, were influential in my day to day life making my overall experience pleasant.
effect of an artificial model’s vocal expressiveness on learning.

References


effect of an artificial model’s vocal expressiveness on learning.54


effect of an artificial model’s vocal expressiveness on learning.


effect of an artificial model’s vocal expressiveness on learning.

Appendix

<table>
<thead>
<tr>
<th>Section</th>
<th>A. Gaze The web video lesson Script</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hello!, this is Eric and in this tutorial we will cover the basics of GazeTheWeb. This is a new web browser that you can control using only your eyes.</td>
</tr>
<tr>
<td></td>
<td>In this video, I will introduce to you some of the basic features of Gaze the Web and I will demonstrate to you! How to conduct a web search and the basic functionalities you can use when navigating the internet with this browser that uses only your gaze.</td>
</tr>
<tr>
<td></td>
<td>In the remainder of the video there will be two separated screens. On the right side, you will see me explaining and performing the web search using my gaze, while on the left side the related GazeTheWeb functions will be displayed.</td>
</tr>
<tr>
<td></td>
<td>For now, this video shows only me explaining and demonstrating things to you on how to conduct a web search.</td>
</tr>
<tr>
<td></td>
<td>So!, let’s get started…</td>
</tr>
<tr>
<td>2</td>
<td>GazeTheWeb is a web browser that captures your eye movements and therefore it allows you to browse the internet just with the use of your eyes. For capturing your gaze we use a calibrated eye tracker.</td>
</tr>
<tr>
<td></td>
<td>The GazeTheWeb cursor looks like an eye. This eye icon, is located below the GazeTheWeb logo that looks like a capital letter T, indicating the position of my own gaze on the computer’s display.</td>
</tr>
<tr>
<td></td>
<td>The color of the eye cursor is white, so as to have a subtle contrast with the background. This is to prevent your eyes from getting tired or from getting strained.</td>
</tr>
<tr>
<td></td>
<td>I hope that this is visible to you, as I am showing you...</td>
</tr>
<tr>
<td>3</td>
<td>GazeTheWeb consists of three panels:</td>
</tr>
<tr>
<td></td>
<td>1. On the left hand side, there is a web panel for the browser menu. There, you find buttons, such as the tab overview, the going back and forward button and the settings button.</td>
</tr>
<tr>
<td></td>
<td>The pause button on the upper left side can be clicked at any moment to stop any interaction with the browser, as I am demonstrating to you now.</td>
</tr>
<tr>
<td></td>
<td>2. On the right side there is a tab panel to interact with a particular webpage, for example the button for selecting a link.</td>
</tr>
<tr>
<td></td>
<td>3. In the center, there is an input field, where you proceed with your search query.</td>
</tr>
<tr>
<td></td>
<td>To begin a web search, I focus my eyes on the middle of the screen on the big T until it gradually changes color from orange to blue. Once the T button gets blue, the GazeTheWeb virtual keyboard appears on-screen. Here I can type a search query.</td>
</tr>
<tr>
<td></td>
<td>To type a search term, look at the respective keyboard letters, one by one. Like this… For this demonstration I will type in Stephen Hawking, as I am interested in his biography.</td>
</tr>
</tbody>
</table>
effect of an artificial model’s vocal expressiveness on learning.
effect of an artificial model’s vocal expressiveness on learning.

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>On the right side outside the frame there is a clock button which takes me to the history of what I have done so far within the GazeTheWeb environment. <strong>as you can see now.</strong> Now, inside the frame, there are four buttons; a pencil, a star, a chatbox and a bin. Each of these buttons has a purpose.</td>
</tr>
<tr>
<td>● To bookmark a tab I can press on the star icon.</td>
</tr>
<tr>
<td>● To reload a tab I can press on the chat box button</td>
</tr>
<tr>
<td>● To remove a tab I can press on the bin icon.</td>
</tr>
<tr>
<td>Now, bookmarks are accessible by clicking the pencil icon on the right side, where I can find all the sites I have bookmarked so far by clicking on the agenda icon, and scroll through them <strong>as I am demonstrating.</strong></td>
</tr>
<tr>
<td>When I click on the star-shaped button in this blue keyboard setting <strong>as you can see,</strong> the site will be saved as a bookmark.</td>
</tr>
<tr>
<td>Also, I can erase bookmarks by clicking on the bin that appears on each bookmark on the left side, for example <strong>I will now</strong> erase the duck bookmark.</td>
</tr>
<tr>
<td>10. Well, this is what I wanted to <strong>show</strong> you in this tutorial. <strong>Thank you very much</strong> for watching it and I hope that you have learned something useful about this new web experience.</td>
</tr>
<tr>
<td><strong>See you in the next tutorial!</strong></td>
</tr>
<tr>
<td>Watch the video lessons on the youtube links below:</td>
</tr>
<tr>
<td><strong>Expressive:</strong> <a href="https://www.youtube.com/watch?v=8EJa5bgD2hs&amp;t=399s">https://www.youtube.com/watch?v=8EJa5bgD2hs&amp;t=399s</a></td>
</tr>
<tr>
<td><strong>Monotonous:</strong> <a href="https://youtu.be/OuwCP7ut8Y">https://youtu.be/OuwCP7ut8Y</a></td>
</tr>
</tbody>
</table>

A1. Final audio cuts can be found in the following link:

[https://drive.google.com/drive/folders/1ugA3AmO4f_MI0qkxtw0vOQbURRs6zPbW?usp=sharing](https://drive.google.com/drive/folders/1ugA3AmO4f_MI0qkxtw0vOQbURRs6zPbW?usp=sharing)
effect of an artificial model’s vocal expressiveness on learning.

B. Questionnaire:

Welcome, in this study, we make you familiar with a new web browser called GazeTheWeb.

You will learn about the functionalities to perform a web search via an instructional video. The instructional video is guided by Eric, an artificial tutor, also known as a pedagogical agent. On the next page, the instructional video will start. After having watched the video, we will ask you to answer several questions about what you have learned and about the instruction itself.

Please make sure that the YouTube video is in full screen and check whether it is set in HD quality (1080p). During the YouTube video viewing do not press any buttons (i.e., pause, fast-forward, etc.). Moreover, only press next once the video is finished. If you face any problems call the experimenter.

A1. \{rand\(1,2\)\}

Below you see a list of traits. Please indicate how you would describe the voice of Eric, the artificial agent who guided you through the video instruction on these traits, on a scale from 1 to 7. *Note.* On Limesurvey all differential semantic bipolar scales were shown with one statement on the left and the other on the right, only when exporting as pdf file it was shown side to side as below.

B1. *The instructor Eric...*
effect of an artificial model’s vocal expressiveness on learning.

<table>
<thead>
<tr>
<th>C1. The instructor Eric...</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has an unappealing/messaging voice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has an appealing/engaging voice</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Has an unfriendly voice</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Has a friendly voice</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Has an ugly voice</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Has a pretty voice</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C2. To what extent did you like Eric’s voice...</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all/Very much</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
effect of an artificial model's vocal expressiveness on learning.61

- Please indicate how you would describe Eric, the artificial agent who guided you through the video lesson, on the lined units below on a scale from 1 to 7.

D1. The Instructor Eric was...

1. Faking Natural
2. Macho-like/Human-like
3. Unconscious/Conscious
4. Artificial/Life-like
5. Moving rigidly/Moving elegantly
6. Dead/Alive
7. Mechanical/Organic

E1. The Instructor Eric was...

1. Incredibly Interactive
2. Apathetic/Responsive
3. Unfriendly/Friendly
4. Unkind/Kind
5. Unpleasant/Pleasant
6. Awful/Nice
effect of an artificial model’s vocal expressiveness on learning.

F1. I feel the content of the video is:

Bad/Good  [scale]
Wordless/Valuable  [scale]
Negative/Positive  [scale]

F2. I feel that the teacher I had during the instructional video is:

Bad/Good  [scale]
Wordless/Valuable  [scale]
Negative/Positive  [scale]

G1. The likelihood of actually following this specific teacher on YouTube for watching other instructional tutorials of related content (assuming they were available) is:

Unlikely/Likely  [scale]
Impossible/Possible  [scale]
Improbable/Probable  [scale]
Would not/Would  [scale]

G2. How much did you learn during the video lesson?

Not at all/Very much  [scale]

G3. To answer the following question, first try to imagine your ideal teacher.

How much do you think you could have learned from this video had you had this ideal teacher?

Not at all/Very much  [scale]
The following items are shown here to see how you feel about the instrumental model that you just watched. Please, select the number several options which best represents your feeling. Note that in some cases the more positive score is “1” while in other cases it is “7”.

III. During the video lesson I felt:

- Motivated/Unmotivated
- Interested/Uninterested
- Involved/Uninvolved
- Not stimulated/Stimulated
- Interested/Uninterested
- Excited/Not excited
- Absorbed/Not absorbed
- Enthusiastic/Enthusiastic

In the following, please choose for each proposition the answer that best represents your feelings from 1=strongly disagree to 7=strongly agree.

11. During the video lesson:

- I paid close attention to the instructional video: [ ]
- I was able to concentrate on the video: [ ]
- The video held my attention: [ ]
- I was absorbed by the presented software activity: [ ]

11.1. I paid more attention to the GTW software activity than to the explanation given by Eric: [ ]
11.2. During the instructional video I paid equal attention to both visuals and to the instructor’s explanation: [ ]

Thank you for answering the questions. In the next part we would like to assess your knowledge of GTWTheWeb after having watched the instructional video.

Please go next to continue with the test.
effect of an artificial model’s vocal expressiveness on learning.
effect of an artificial model's vocal expressiveness on learning.
effect of an artificial model’s vocal expressiveness on learning.

N2. The left-hand side panel of GazeTheWeb contains...
- Buttons to apply a query to the search engine or directly start the search
- The GazeTheWeb logo and the input field
- Buttons to interact with a webpage (e.g., link selection, automatic scrolling button)
- Buttons such as the tab overview, the go back and the forward button
  I don’t know

N3. Which of the following is NOT an icon of GazeTheWeb?
- A “T” icon
- A finger-point icon
- A diamond icon
- A “T” icon
  I don’t know

N4. To begin a web search on GazeTheWeb, you should focus your eyes on the T button until it changes color...
- From blue to orange
- From orange to blue
- From orange to brown
- From brown to orange
  I don’t know

For each of the following questions choose the one answer that is correct. Please do not guess but choose the option “I don’t know” in case you can’t remember the correct answer.

O1. The backspace icon is located...
  - on the left-hand side of the virtual keyboard
  - on the right-hand side of the virtual keyboard
  - in the middle of the virtual keyboard
  - at the bottom of the virtual keyboard
  I don’t know

O2. What is the difference between icons with the arrows on the right-hand side and the left-hand side of the keyboard?
The arrows on the right-hand side move the cursor to the previous/next letter while the arrows on the left-hand side move the cursor to the previous/next word
The arrows on the right-hand side move the cursor to the previous letter while the arrows on the left-hand side move the cursor to the previous word
The arrows on the right-hand side move the cursor only to the previous letter while the arrows on the left-hand side move the cursor only to the next letter
The arrows on the right-hand side move the cursor only to the previous word while the arrows on the left-hand side move the cursor only to the next word
  I don’t know

O3. Which icon activates the “automatic scrolling” function?
- A left-arrows icon
- A square containing a number
- A diamond icon
  None of the previous. Automatic scrolling is activated by default
  I don’t know
effect of an artificial model’s vocal expressiveness on learning.

For each of the following questions choose the one answer that is correct. Please do not guess but choose the option ‘I don’t know’ in case you can’t remember the correct answer.

P1. What happens when the “hyperlink navigation” function is activated?
   - The eye icon changes to a finger-point button
   - A new tab is open for every link presented on-screen
   - You can copy the URL of the desired link
   - The webpage starts zooming in
   - I don’t know

P2. How do you deactivate the zoom icon once it is selected?
   - It is deactivated automatically once it is used
   - By focusing the eyes on the same icon
   - By moving back to the homepage
   - By moving forward to the next page
   - I don’t know

P3. The finger-point button...
   - Activates the keyboard
   - Activates the copy-paste function
   - Allows to pass from a sub-page to another
   - Allows hyperlink navigation
   - I don’t know

Q1. The text selection icon contains:
   - A pencil
   - The capital letter A
   - The letters ABC
   - The capital letter A
   - The numbers 123
   - I don’t know

Q2. Immediately after the text selection button has been clicked, a message appears on screen asking...
   - ...to look at the end point of the text selection
   - ...to copy text to clipboard
   - ...to look at the starting point of the text selection
   - ...to look both at the starting and end point of the text selection
   - I don’t know
effect of an artificial model’s vocal expressiveness on learning.

R1. The clock button:
  - is used to bookmark a page
  - shows the tabs that are currently open
  - is used to cancel the action and go back to the navigation panel
  - shows the history of all the actions performed within the GazeTheWeb environment
  - I don’t know

R2. How would you access bookmarks that have been already saved?
  - By focusing on the pencil button and then on the star-shaped button
  - By focusing on the pencil button and then on the agenda button
  - By focusing on the star-shaped button and then on the pencil button
  - By focusing on the agenda button and then on the pencil button
  - I don’t know

R3. Which of the GazeTheWeb icons below would you use to start the search directly after you have typed your search term on the keyboard?

S1. To create a bookmark, which of the gaza the web icons would you use?
  - A pencil
  - A chart
  - A star
  - A clock
  - I don’t know

S2. To refresh a tab, which of the gazette web icons would you use?
  - A chart
  - A pencil
  - A star
  - A clock
  - I don’t know

S3. Which of the GazeTheWeb icons below would you use to apply the search term on the search engine after you have typed it on the keyboard?

I don’t know
effect of an artificial model’s vocal expressiveness on learning.

Next, please answer some questions about your use of software technologies in general.

T1. On average, how many hours per week do you spend on using a computer?

- Less than 3
- 4 - 6
- 7 - 9
- 10 - 12
- More than 12

U1. Assistive technology use:

Some people utilize Assistive software technologies, such as a screen magnifier, screen reader, word prediction, spell checking, voice recognition, as well as assistive hardware technologies, like a joystick or an alternative mouse, when using their computer. Do you make frequent use of any assistive technologies, when using your computer?

- Yes
- No

Finally, please answer some questions about yourself.

V1. Please indicate your gender:

- Male
- Female
- No answer

Finally, please answer some questions about yourself.

W1. What is the highest degree or level of school you have completed? If currently enrolled, highest degree received.

- Pre-school (Basisonderwijs)
- Vocational high school (VMBO, MAVO of MAVO onderbouw)
- Applied science high school (HAVO of vwo/havo)
- High school (VWO, gymnasium of vwo/gymnasium)
- Vocational school (MBO, MTS)
- University of applied science (HBO, HTS)
- University Bsc (WOBsc)
- University MSc (WO MSc)
- PhD
- Something else
- Don’t want to tell

This is the end of the experiment. Unfortunately, the GazeTheWeb software is still under development and therefore, you will not be able to try it out yourself. Our goal was to identify your most genuine impressions of the GazeTheWeb and the training provided, for further optimization. You can leave the room and report to the experiment leader.
effect of an artificial model’s vocal expressiveness on learning.

C. Video Lesson by the artificial model on Youtube.

1. Expressive Eric: https://www.youtube.com/watch?v=8EJa5bgD2hs&t=89s