The effect of an artificial agent's vocal expressiveness on immediacy and learning

Received: 3 September 2020 | Revised: 18 July 2021 | Accepted: 2 October 2021

DOI: 10.1111/jcal.12632

ARTICLE

Journal of Computer Assisted Learning WILEY

The effect of an artificial agent's vocal expressiveness on immediacy and learning

Sofia Fountoukidou | Uwe Matzat | Jaap Ham | Cees Midden

Human-Technology Interaction, Eindhoven University of Technology, Eindhoven, Netherlands

Correspondence
Sofia Fountoukidou, Eindhoven University of Technology, Human-Technology Interaction, P.O. BOX 513, 560 MB, Eindhoven, Netherlands.
Email: s.fountou@tue.nl

Abstract

Background: Though pedagogical artificial agents are expected to play a crucial role in the years to come, earlier studies provide inconsistent results regarding their effect on learning. This might be because their potential for exhibiting subtle nonverbal behaviours we know from human teachers has been untapped. What is more, there is little evidence of the processes underlying the effect of nonverbal behaviours of teachers (either human or artificial) on learning, so as to better guide their practical application.

Objectives: The aims of the current research were threefold: firstly, to examine the effect of an artificial agent's vocal expressiveness on non-verbal immediacy (teachers' ability to increase psychological closeness through nonverbal communication). Secondly, to test whether an artificial agent showing strong vocal expressiveness will enhance affective and cognitive learning (perceived and actual), as compared to an artificial agent that shows weak vocal expressiveness. Thirdly, to examine whether the underlying mechanisms of motivation and attention explain the effect of immediacy (and thereby also of vocal expressiveness) on the two learning outcomes.

Methods: The study used a between-participants design, with the participants being randomly assigned to one of the two experimental conditions: artificial modelling with strong expressiveness and artificial modelling with weak vocal expressiveness.

Results and conclusions: Results showed that an artificial agent with strong vocal expressiveness increased affective and perceived cognitive learning. Partial support was found for actual cognitive learning. What is more, our findings revealed that vocal expressiveness is related to affective and perceived cognitive learning because it promotes nonverbal immediacy. Finally, results provided evidence of motivation as a mediator of the path from immediacy to affective learning.

Major takeaways: The current findings verify the important role of nonverbal immediacy found in traditional educational settings. However, showing that these results also apply to artificial teachers is essential, given that the educational landscape is changing and reshaping by artificial intelligence. Thus, taking into consideration the role of vocal expressiveness in the development of artificial teachers or voice assistants like Alexa, Siri, and Google assistant, as a way to enhance immediacy and affective experience of learners is of imperative value, since they will be found more and more in our societies.

KEYWORDS
artificial agent, learning, modeling, vocal expressiveness

1 | INTRODUCTION

Pedagogical artificial agents are animated characters that are embedded in virtual learning environments. They are seen as potential tools to create a social presence that primes learners to deeply process the learning material (Kim & Baylor, 2015). Artificial teachers are expected to play a crucial role in the years to come, as the educational landscape is changing and being reshaped by artificial intelligence (Chassignola, Khoroshavinb, Klimovac & Bilyatdinovac, 2018). Nonetheless, findings regarding their effectiveness for learning are mixed. That is, a meta-analysis revealed that agents were associated with a small but positive effect on learning (Schroeder et al., 2013), while a systematic review showed that the majority of studies found nonsignificant effects (Heidig & Clarebout, 2011; Martha & Santoso, 2019). One of the reasons for the insignificant effects of artificial agents on learning might be that their potential for subtle nonverbal behaviour that we know from human-human interaction has not been thoroughly studied (Krämer & Bente, 2010). In fact, it has been argued that nonverbal cues of artificial teachers could increase their social presence leading to learning gains (Baylor & Kim, 2009). However, there are many unanswered questions on how artificial agents' nonverbal cues should be implemented to accomplish such aims in multimedia settings.

A closer inspection of the behaviour of human teachers could be taken into consideration in order to inform the behavioural design of pedagogical artificial agents. In traditional classroom settings, teachers' nonverbal behaviour plays a crucial role in student learning. Specifically, some forms of teachers' nonverbal behaviour are found to increase 'nonverbal" immediacy" (Andersen, 1979). The nonverbal immediacy concept refers to the ability of teachers to create psychological closeness with their students through nonverbal communication (Mehrabian, 1981). This concept is grounded in approach-avoidance theory, which asserts that people "are drawn toward the person and things they like, evaluate highly, and prefer; and they avoid or move away from things they dislike, evaluate negatively, or do not prefer" (Mehrabian, 1981, p. 1). Several nonverbal cues of teachers have been found to play a crucial role in student's learning, such as proximity, eye gaze, gestures, body position, facial and vocal expressiveness (Witt & Wheeless, 2001). Cumulative evidence has revealed that human teachers' nonverbal immediacy behaviour promotes affective learning (student beliefs and motivations) and cognitive learning (recall and perceived learning) (Bambaeeroo & Shokrpour, 2017; Ellis et al., 2016; Uleanya et al., 2020; Witt et al., 2004).

Two major explanations have been proposed for explaining the effect of immediacy on learning: motivational theory (Christophel, 1990) and arousal–attention theory (Kelly & Gorham, 1988). Motivational theory suggests that some forms of teacher behaviour may increase student (state) motivation by stimulating students, directing their efforts and, in turn, influencing affective learning. Findings of a recent review on teacher immediacy and student motivation indicated that teachers' play a significant role in enhancing student motivation (Liu, 2021). Additionally, the author indicated that teachers' immediacy influences students beyond their motivation and probably interacts with learning outcomes. Nonetheless, only few empirical studies have investigated this, providing support for motivational theory (i.e., Christophel, 1990; McCroskey, Richmond & Benett, 2006). Arousal-attention theory argues that immediacy stimulates arousal, which is related to attention and memory, thus leading to greater cognitive learning. In sum, the main argument of arousal-attention theory is that immediate teachers arouse students, draw attention to themselves and to the taught material, and thereby produce more student learning (McCroskey & Richmond, 1992). Again, only few studies empirically tested the viability of arousal-attention theory (Comstock et al., 1995; Kelly & Gorham, 1988). Further examination of motivation and attention as mediators of the path from immediacy to affective and/or cognitive change has been deemed necessary (Witt et al., 2004).

Visual nonverbal cues of artificial pedagogical agents, such as the use of gestures and facial expressions, have received increasing attention over the last years (i.e., Liew et al., 2016; Wang et al., 2018). Nonetheless, it has been shown that it is mainly the artificial agent's voice that is responsible for increased learning rather than its visual presence (Atkinson, 2002; Bente et al., 2008; Krämer & Bente, 2010). Previous work found that the characteristics of a speaker's voice that can affect learning (i.e., transfer and social perception) in multimedia settings are (amongst others) mechanization (human vs machine-synthesized voice) (Atkinson et al., 2005; Mayer et al., 2003), accent (native vs. foreign accent) (Mayer et al., 2003), gender (male vs. female voice) (Linek et al., 2010), dialect (regional dialect vs. standard speech) (Rey & Steib, 2013) and slang (youth slang vs. standard speech) (Schneider et al., 2015). However, there is limited evidence on whether and, more importantly, how artificial teachers' vocal nonverbal cues can influence multimedia learning outcomes.

In fact, only a few studies have provided some evidence that indicates that vocal expressiveness can benefit learning. Liew et al. (2020) found that an enthusiastic voice of a virtual speaker positively led to increased learners' engagement and learning outcome when compared to a calm voice. However, Liew's study featured an invisible narrator that had no visual cues, such as face and body. Thus, it is not known whether the positive effect of vocal expressiveness can also manifest in a multimedia environment presented by an on-screen artificial agent that inevitably involves visual cues. Furthermore, two studies examined vocal nonverbal cues of a robot and found an effect of an expressive voice, as compared to a flat voice, on affective and cognitive learning (Kory Westlund et al., 2017; Kennedy, Baxter & Balpaeme, 2016). However, it is not known whether results of studies that used physically embodied robots also apply to artificial agents that are not physically present. To the best of our knowledge, only one study investigated effects of a vocally expressive artificial agent (i.e., use of additional pauses, louder voice and better enunciated words) on learning, as compared to a non-vocally expressive artificial agent (Valetsianos 2009). That research (Valetsianos, 2009) provided evidence of a benefit of vocal expressiveness on affective and cognitive learning. However, given that the study employed a quasi-experimental design with partial randomization (random assignment but not random selection), the risk of introducing confounding factors (i.e., differences in characteristics of people involved) affecting the outcome is high (Handley et al., 2018). What is more, further delineations of verbal expressiveness such as pitch tone, pitch variation, and speech rate require further exploration (Valetsianos, 2009).

Therefore, the aim of the current study is to examine whether vocal expressiveness (operationalized as pitch tone, pitch variation and speech rate) of an artificial agent can create immediacy and enhance learning, as it is proposed by communication literature (i.e., Ellis et al., 2016; Witt et al., 2004). What is more, the study goes one step further and tests whether the proposed underlying mechanisms of attention and motivation can explain the anticipated effect of vocal expressiveness on learning outcomes.

It has been argued that the inclusion of nonverbal cues of artificial agents is not a panacea, and caution is needed when constructing such cues, as they can be detrimental to learning (Dehn & Van Mulken, 2000; Clark & Choi, 2006; Woo, 2009; Frechette & Moreno, 2010). This is because they could impose an additional processing burden, which is known as extraneous cognitive load, on working memory, because learners have to attend to nonverbal cues by an expressive artificial agent (Sweller, 2004). In the current study, we employed an artificial agent that adopted the role of a model for enabling social learning (see, Fountoukidou et al., 2019). We argue that vocal expressiveness of an artificial model facilitates, rather than hinders, learning. Our claim is based on the cognitive theory of multimedia learning, according to which there are two separate channels (auditory and visual) for processing information that both have a limited processing capacity (Mayer, 2003). Due to the fact that modelling requires a substantial amount of processing to take place in the visual channel (i.e., demonstration), we argue that the inclusion of vocal nonverbal cues balances processing demands, since vocal nonverbal cues are being processed in the auditory channel and not in the visual channel.

Taking into consideration all the above, the aims of the current research were threefold: Firstly, to examine the effect of an artificial agent's vocal expressiveness on immediacy. We predicted that an artificial agent that shows strong vocal expressiveness will increase perceptions of immediacy, as compared to an artificial model that shows weak vocal expressiveness (H1). Secondly, to test whether an artificial agent showing strong vocal expressiveness will enhance affective and cognitive learning, as compared to an artificial agent that shows weak vocal expressiveness. We expected that strong vocal expressiveness would improve affective learning (H2) and cognitive learning (H3), when compared to weak vocal expressiveness. Thirdly, to examine whether the underlying mechanisms of motivation and attention explain the effect of immediacy (and thereby also of vocal expressiveness) on affective and cognitive learning. Specifically, we hypothesized motivation to mediate (part of) the effect of immediacy on affective learning (H4), and we hypothesized attention to mediate (part of) the effect of immediacy on cognitive learning (H5).

Overall, the current study aimed at contributing to the topic of artificial agents' nonverbal cues, such as vocal expressiveness, as a means to promote immediacy and in turn enhance learning. This process has received limited attention, despite the fact that the voice of artificial agents has been argued to have a profound impact on learning (Atkinson, 2002; Bente et al., 2008; Krämer & Bente, 2010). More importantly, the study aims at extending earlier findings by examining the underlying mechanisms of attention and motivation through which nonverbal cues of an artificial agent improve affective and cognitive learning.

2 | METHOD

2.1 | Participants

The participants of this study were 144 individuals (38% females and 62% males). The majority of the population were students from Eindhoven University of Technology. Specifically, 92 participants (63%) were educated to undergraduate level or higher, 45 participants (31.2%) had completed high school, and seven participants (4.86%) chose not to disclose their education.

The study used a between-participants design, with the participants being randomly assigned to one of the two experimental conditions: artificial modelling with strong expressiveness and artificial modelling with weak vocal expressiveness. The dependent variables of the study were immediacy, affective learning, and cognitive learning (perceived and actual). Only participants who indicated to speak, read and write English fluently were invited to participate. The experiment lasted for approximately 30 min, and participants were compensated for their participation (5 euros).

2.2 | Materials

The 3D animated artificial agent, employed in this research, was designed using the *CrazyTalk 8* software.

The study's instructional script explains how to use eye-tracking software, called *GazeTheWeb* (GTW). GTW is a gaze-controlled web-browser, which works with an eye-tracking hardware. However, this software was only demonstrated, but not actually tried out by the participants.

The vocal parameters that used to distinguish the strong vocal expressiveness from the weak vocal expressiveness were pitch (pitch tone and pitch variation) and speech rate. The actor's voice was recorded using *Audacity* software. Pitch analysis of these audio recordings was performed with the use of *Praat* software.

2.3 | Artificial agent

The artificial agent used in the current study was developed such that it shared some common characteristics with participants in terms of their appearance. These characteristics were derived from earlier literature (Rosenberg-Kima et al., 2008). That is, since the majority of the participants were students at a Dutch University, the agent was designed to be young (<30 years old), attractive (in terms of the artificial agent's facial characteristics) and "cool" (in terms of the artificial agent's clothing and hairstyle). What is more, the artificial agent intentionally lacked strong facial expressions.

Concerning its educational role, the artificial agent took the role of a model, demonstrating the GTW system's functionalities by moving its head and eyes, while providing verbal explanations at the same time (see Figure 1).

The design of the artificial agent (both in terms of appearance and educational role,) was identical in both experimental conditions, and the only difference was the level of the artificial agent's vocal expressiveness (explained in the subsection 2.2.3).

2.4 | Instructional script

Initially, one basic version of an instructional script was created, which familiarized participants to the use of a novel web browser called GazeTheWeb (GTW), accessible solely by eye gaze input. Specifically, the script pertained to a description on how to conduct a web search and the basic functionalities one could use (e.g., scrolling and tab manager) when navigating the internet with this gaze-controlled web browser. The GTW browser was unknown to the study's population. In both videos, the agent demonstrated how to proceed with specific activities that he explained in the script by corresponding movements of the head and the eyes, thereby acting as a behavioural model that enables social learning. These movements did not differ between the two versions of the video. Then, two versions of this script were developed. The only difference between these two versions was the level of the artificial agent's vocal expressiveness (strong vs. weak). That is, the two versions differed in terms of average pitch tone, pitch variation, and speech rate (see the 2.2.3 subsection for more details). A male actor, whose voice was recorded and used by the artificial agent, performed both instructional versions. The selection of the voice actor was based on two requirements: clear English pronunciation and good voice acting skills. English was chosen as the artificial teacher's language as this is the language that is used for the lectures of most Bachelor and Master students at our university.

2.5 | Vocal expressiveness

The study concurrently manipulated the vocal parameters of (a) pitch tone, (b) pitch variation, and (c) speech rate, so as to create two different levels of vocal expressiveness (strong vs. weak). The decision to manipulate the aforementioned vocal parameters was based on earlier research suggesting that the combination of a speaker's temporal (i.e., speech rate) and expressive (i.e., pitch) vocal features has the greatest impact on both emotions and cognition (Breitenstein et al., 2001).

Pitch is defined as the degree of highness or lowness of a tone, determined by the vibration of the vocal folds (i.e., the faster the vibration per second [Hz], the higher the pitch). It is generally measured as the fundamental frequency of the sound wave. There is no universally optimal. Rather, pitch tone is determined by factors such as culture and context (Gudykunst et al., 1988). Nevertheless, according to general guidelines, the average fundamental frequency

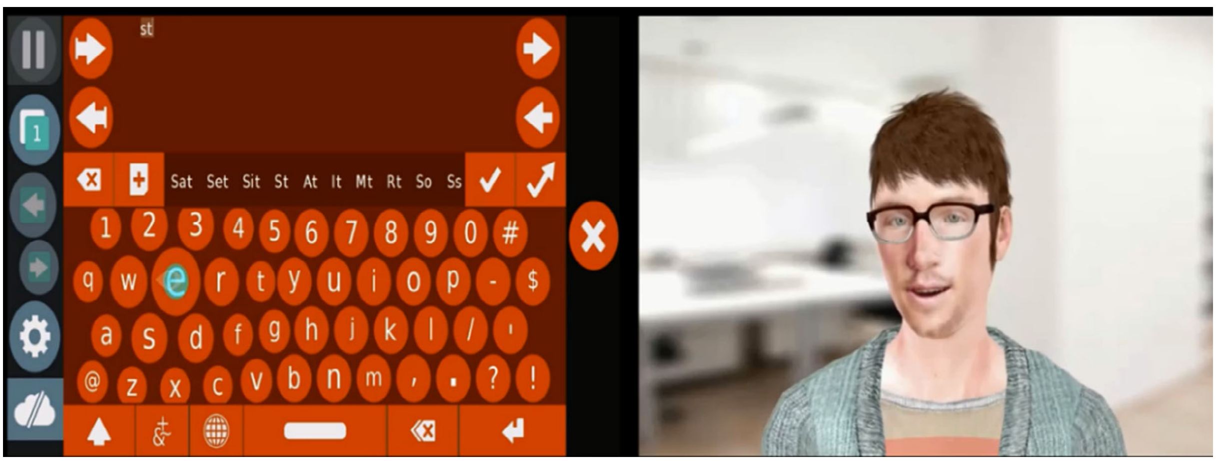

FIGURE 1 Artificial modelling: On the right side, the agent appeared to demonstrate an action (i.e., typing), while providing verbal explanations; on the left side, the light blue highlights the effect of the artificial agent 's action

of a male adult's speech is 120 Hz (Hollien & Shipp, 1972; Hsiao et al., 1994; Mizuno & Nakajima, 1998). Thus, in this study, the pitch tone boundary to differentiate the two levels of vocal expressiveness (strong and weak) was set at ~120 Hz. To control for pitch tone variations within and between the study's two conditions (strong and weak vocal expressiveness), we employed a two-step process. Firstly, for both conditions, during the audio recordings, we constantly monitored the minimum and maximum pitch (using Audacity) for each ~1 min audio part, thus providing an average pitch. This goal of such monitoring was to ensure that the predefined boundary of 120 hz has been successfully met (i.e, strong vocal expressiveness >120 Hz; weak vocal expressiveness <120 Hz). Secondly, once the audio recordings for both conditions were completed, we analysed samples for each condition (i.e., 10 samples of 10 seconds for each condition using Praat), which were randomly selected but were the same for both conditions, so as to maintain consistency. All in all, according to our analysis, the average pitch tone of the strong vocal expressiveness condition was 260 Hz, while the average pitch tone of the weak vocal expressiveness condition was 115 Hz.

In addition to the pitch tone, pitch variation (i.e., intonation) was also manipulated. That is, there was more pitch variation (i.e., voice rises and then falls before it rises again) in the strong vocal expressiveness condition than in the weak vocal expressiveness condition. Hence, the weak vocal expressiveness condition, apart from its lower pitch tone, was also constructed to be "flat" in terms of pitch variation. Concerning the strong vocal expressiveness condition, the artificial agent's voice raising was congruent with important information that participants needed to recall. The development of pitch tone and variation in the strong vocal expressiveness condition was intentionally prepared to emphasize both, affective nonverbal communication (i.e., speaker's feeling and attitude conveyance) and cognitive nonverbal communication (i.e., help in the encoding of new information) (Frechette & Moreno, 2010).

Speech rate is defined as the speed at which one speaks. It's calculated by the number of words spoken in a minute. An average number of words per minute (wpm) can vary hugely both between people and within a person. This is because speech rate is inextricably bound to the speaker's culture, geographical location, subject matter, gender, emotional state, fluency, profession or audience. Nonetheless, according to some general guidelines, that conversational speech generally falls between 125 wpm at the slow end, to 150 wpm in the fast range (Simonds et al., 2006). Overall, the speech rate in the strong vocal expressiveness condition was 133 wpm, as opposed to 119 wpm in the weak vocal expressiveness condition. Inevitably, this manipulation led to a relatively small difference in the video duration between the two conditions (9.5 min in the strong vocal expressiveness condition as opposed to 10 min in the weak vocal expressiveness condition).

3 | MEASURES

As a manipulation check, participants were asked to evaluate the artificial agent's vocal expressiveness via a self-constructed scale, assessing the vocal parameters of pitch tone, pitch variation and speech rate. This scale consisted of three items (use of high vs. low tone of voice; use of vocal variety vs. flat voice; use of fast vs. slow speech rate) and it was administered through a 7-point semantic differential scale. By averaging participants' answers to these questions, we constructed a measure of perceived vocal expressiveness with acceptable reliability (Cronbach's $a = 0.68$).

Nonverbal immediacy (H1) was assessed using a scale consisting of six items and was administered through a 7-point semantic differential scale (i.e., pleasant vs. unpleasant voice, enthusiastic vs. boring voice etc.). This scale was adapted from earlier versions measuring not only vocal cues but a variety of other nonverbal cues (i.e., facial expressiveness) (Mehrabian, 1981; Richmond, Gorham & McCroskey, 1987; Richmond, McCroskey & Johnson, 2003; Servilha & Costa, 2015). The scale has a high reliability (Cronbach's $a = 0.87$).

Affective learning (H2) was assessed by asking participants to estimate three components of their affective perceptions towards the instructional material, towards the artificial instructor, and the likelihood of following the same artificial instructor for other instructional videos. These components of affective learning were administered through an established 7-point semantic differential scale (Andersen, 1979; Scott & Wheeless, 1975). We constructed reliable measures of affective perceptions towards the instructional material (example item: good/bad, Cronbach's $a = 0.86$), towards the artificial instructor (example item: good/bad, Cronbach's $a = 0.88$) and likelihood of following the same artificial instructor for other instructional videos (example item: unlikely/likely, Cronbach's $a = 0.90$), by averaging participants answers to each set of questions.

Cognitive learning (H3) was assessed both objectively (through a recall test in two forms) as well as subjectively (by assessing perceived cognitive learning), leading to three separate indicators Specifically, recall of the content of the instructional video was measured as an index of cognitive learning, and it was measured through a self-constructed recall test. This recall test contained two methods of knowledge assessment: (1) A "fill-in-the-blanks" test consisted of nine recall items, and, (2) a multiple-choice test consisted of 18 questions. We use both scores as separate indicators of recall. For the gap filling test, participants were asked to recall keywords (exact words or synonyms) spoken by the artificial agent during the video, and to fill in the blanks of a written transcript. For the multiple-choice test, participants were asked to answer a series of questions by selecting the correct answer amongst four optional answers. We constructed two measures of cognitive learning by counting participants' number of correct answers to each test separately. The participants' performance scores for the gap filling test were calculated by two researchers independently. There was a 100% agreement on the gap filling performance scores between the two researchers.

Furthermore, perceived cognitive learning was assessed by asking participants' to answer two questions that had to be answered on 7-point Likert scales (Richmond, Gorham & McCroskey, 1987). Next, a "learning loss" score was computed by subtracting the score on the first question (i.e., How much did you learn during the video lesson?)

from the score of the second question (i.e., How much do you think you could have learned from this video had you had this ideal instructor?), indicating a learner's overall perceived cognitive learning score. Reliability of this measure in previous research was reported at 0.94 (Gorham, 1988). Overall, this way of constructing a learning loss score has been widely used in communication research as an index of cognitive learning (e.g., Chesebro & McCroskey, 2000).

The main dependent variable of the fourth hypothesis was (state) motivation. Motivation was assessed by asking participants to answer nine questions on how they felt about the instructional video they watched. The questionnaire was administered using a 7-point semantic differential scale taken from Christophel (1990) (we selected the nine out of the 12 questions that were relevant to our study). We constructed a reliable measure of motivation (Cronbach's $a = 0.91$), by averaging participants' answers to each set of questions.

The main dependent variable of the fifth hypothesis was attention. Attention was assessed by asking participants to answer four established questions, administered using a 7-point scale, about the level of their attention to the instructional video. These were taken from Yi and Davis (2003). We constructed a reliable measure of motivation (Cronbach's $a = 0.86$), by averaging participants' answers to each question.

Lastly, we explored whether vocal expressiveness would influence how learners perceive the likeability of an artificial agent. Furthermore, and more importantly, likeability of the artificial agent was taken into consideration to test whether the influence of vocal expressiveness on affective learning is still mediated by immediacy even when another possible mediating path is considered (i.e., likeability of the artificial agent). The agents' likeability was measured with a subscale of the "Godspeed" scale, developed to assess key concepts of Human-Computer interaction (Bartneck, Croft & Kulic, 2008). The scale was formatted in a 7-point semantic differential, scale. We constructed a reliable scale of likeability (Cronbach's $a = 0.93$) by averaging participants' answers to each set of questions.

3.1 | Procedure

Participants were welcomed in the main hall of the lab building. Each participant was required to read and sign an informed consent form, explaining the general aim of the study and their willingness to participate. Next, they were randomly assigned to one of the two experimental conditions and they were requested to watch an instructional video on a computer monitor regarding the use of GTW browser. The video screen was split into two sides: on the right-hand side, an artificial agent appeared to use the GTW system by moving the head and eyes, while explaining the system functionalities being demonstrated; on the left-hand side, the actual system was displayed, showing the effects of the artificial agent's actions on the system in real time (see Figure 1). Participants in both conditions were provided with an identical instructional video, with the only difference being the level of vocal expressiveness in terms of pitch tone, pitch variation and rate.

As a next step, they filled in an online survey and a recall test. Lastly, they were debriefed, paid and thanked for their participation.

4 | RESULTS

Manipulation check: An independent sample *t* test analysis was conducted to check the study's manipulation of vocal expressiveness (i.e., perceptions of vocal parameters of pitch tone, pitch variety and speech rate) between the strong vocal expressiveness and weak vocal expressiveness condition. As expected, the results revealed a statistically significant effect on vocal expressiveness, $t(142) = 9.2$, $p < 0.001$, with participants in the strong vocal expressiveness condition ($N = 78$, $M = 3.7$, $SD = 0.96$) to report stronger perceptions of vocal expressiveness as compared to participants in the weak vocal expressiveness condition ($N = 66$, $M = 2.3$, $SD = 0.84$).

The mean values and SDs of all dependent variables in the following hypothesis testing are shown in Table 1 for both groups of participants.

Immediacy: To test H1, an independent sample *t* test analysis was conducted to examine the effect of the level of the artificial model's vocal expressiveness on individuals' perceptions of immediacy. Results supported our hypothesis, revealing a statistically significant effect, $t(142) = 6.873$, $p < 0.001$, with participants in the strong vocal expressiveness condition ($N = 78$, $M = 4.19$, $SD = 1.3$) to report higher perceptions of immediacy as compared to participants in the weak vocal expressiveness condition ($N = 66$, $M = 2.84$, $SD = 1.0$).

Affective learning: To test H2, a one-way multivariate analysis of variance (MANOVA) was conducted to examine the effect of the level

TABLE 1 Mean immediacy, affect towards instructional material, affect towards instructor, likelihood of repetition, recall (performance), recall (multiple choice), perceived cognitive learning, and perceived likability of the agent (and standard deviations) by vocal expressiveness

	Vocal expressiveness	
	Strong	Weak
Perceptions of immediacy	4.19a (1.3)	2.84b (1.0)
Affective perceptions towards instructional material	5.6a (1.0)	5.1b (1.0)
Affective perceptions towards the instructor	5.4_a (1.1)	4.5_b (1.1)
Perceived likelihood of same instructor for other instructional material	4.4_a (1.5)	3.3_b (1.4)
Recall–performance	7.9_a (2.7)	6.9_b (2.2)
Recall–multiple choice	10.4_a (2.8)	10.6_a (2.9)
Perceived cognitive learning (learning loss)	$.41_a$ (SE = 0.11)	$.83_b$ (SE = 0.13)
Perceived likability of artificial agent	5.4_a (1.0)	4.6_b (1.0)

Note: Means with different subscripts are different, $p < 0.05$.

of the artificial model's vocal expressiveness on the three affective dependant variables (affective perceptions towards the instructional material, towards the artificial instructor, and perceived likelihood to follow the same instructor on other instructional material). The results revealed a statistically significant effect of the level of the artificial model's vocal expressiveness on the three dependent variables combined, Wilk's $\Lambda = 0.85$, $F(3,140) = 7.88$, $p < 0.001$, $\eta p^2 = 0.14$.

That is, supporting this hypothesis, separate univariate ANOVAs on the outcome variables revealed a significant treatment effect on: (1) affective perceptions towards instructional material, $F(1, 142) = 7.23$, $p < 0.01$, $\eta p^2 = 0.48$, with participants' affect towards the instructional material to be more positive in the strong vocal expressiveness condition ($N = 78$, $M = 5.6$, $SD = 0.98$) as compared to participants in the weak vocal expressiveness condition ($N = 66$, $M = 5.1$, $SD = 1.0$); (2) affective perceptions towards the artificial instructor, $F(1, 142) = 21.39$, $p < 0.001$, $\eta p^2 = 0.13$, with participants' affect towards the artificial instructor to be more positive in the strong vocal expressiveness condition ($N = 78$, $M = 5.4$, $SD = 1.1$) as compared to participants in the weak vocal expressiveness condition ($N = 66$, $M = 4.5$, $SD = 1.1$); 3) the likelihood of following the same artificial instructor for other instructional material, $F(1,142) = 17.82$, $p < 0.001$, $\eta p^2 = 0.11$, with participants' perceived likelihood to follow the same instructor on other instructional material to be more positive in the strong vocal expressiveness condition ($N = 78$, $M = 4.4$, $SD = 1.5$) as compared to participants in the weak vocal expressiveness condition ($N = 66$, $M = 3.3$, $SD = 1.4$).

Cognitive learning: To test H3, a one-way multivariate analysis of variance (MANOVA) was conducted to examine the effect of the level of the artificial model's vocal expressiveness on individuals' recall. Recall was measured with a fill-in-the-blanks test and a multiple-choice test. The results revealed a statistically significant effect of the level of the artificial model's vocal expressiveness on the two dependent variables combined, Wilk's $\Lambda = 0.94$, $F(2, 141) = 4.33$, $p = 0.01$, $\eta p^2 = 0.6$.

In line with our hypothesis, separate univariate ANOVAs on the outcome variables revealed a significant treatment effect on fill-in the-blanks test, $F(1, 142) = 5.25$, $p = 0.02$, $\eta p^2 = 0.36$, with participants' recall performance to be better in the strong vocal expressiveness condition ($N = 78$, $M = 7.9$, $SD = 2.7$) as compared to participants in the weak vocal expressiveness condition ($N = 66$, $M = 6.9$, $SD = 2.2$). Results showed a non-significant treatment effect on the multiple-choice test, $F(1, 142) = 0.31$, $p > 0.05$, between participants in the strong vocal expressiveness condition ($N = 78$, M = 10.4, SD = 2.8) and participants in the weak vocal expressiveness condition ($N = 66$, M = 10.6, SD = 2.9).

Perceived cognitive learning (learning loss): An independent sample t test analysis was conducted to examine the effect of the level of the artificial model's vocal expressiveness on individuals' perceptions of learning. As expected, the results revealed a statistically significant effect, $t(142) = -2.36$, $p = 0.02$, $r = 0.20$, with participants in the strong vocal expressiveness condition ($N = 78$, $M = 0.41$, $SE = 0.11$) to report less learning loss (therefore more perceived cognitive learning) as compared to participants in the weak vocal expressiveness condition ($N = 66$, $M = 0.83$, $SE = 0.13$).

Artificial agent's likeability: An independent sample t test analysis was conducted to examine the effect of the level of the artificial model's vocal expressiveness on individuals' judgements about the artificial agent's likeability. The results revealed a statistically significant effect, $t(142) = 4.44$, $p < 0.001$, with participants' judgements on the artificial agent's likeability to be more positive in the strong vocal expressiveness condition ($N = 78$, $M = 5.4$, $SD = 1.0$) as compared to participants' judgements in the weak vocal expressiveness condition ($N = 66$, $M = 4.6$, $SD = 1.0$).

4.1 | Path analyses

Our final aim was to test whether motivation and attention could explain (parts of) the anticipated effect of immediacy on affective learning (affective perceptions towards the instructional material, towards the artificial instructor and the likelihood of following the same artificial instructor for other instructional material) (H4) and cognitive learning (fill-the-blanks recall test and perceived cognitive learning) (H5). In addition, we also included in the analysis the examination of vocal expressiveness as a predictor of immediacy. Hence, path analyses were conducted in STATA 14 to test the model (see Figure 2). This type of analysis provides a comprehensive picture of the nature of the associations between the predictor and dependent variables of interest. The overall fit of the models was assessed by the chi-square goodness of fit (Χ2), comparative fit index (CFI), and root mean square error of approximation (RMSEA). Detailed results of the path analysis for each affective and cognitive learning variable are reported in the following subsections.

4.2 | Path analyses of affective learning outcomes

Concerning affective learning, three path analyses models were tested for the three affective learning outcomes (affect towards the artificial teacher, likelihood of following the same artificial instructor for other instructional material and affect towards the instructional material). Results are presented in Figures 3, 4 and 5.

In more detail, as seen in Figure 3, vocal expressiveness was a positive significant predictor of immediacy. Further, immediacy was found to be a positive significant predictor of affect towards the artificial teacher. Motivation was a significant positive predictor of affect towards the artificial teacher. To the opposite, the path coefficient from attention to affect towards the artificial teacher was positive but non-significant. These findings suggest that motivation partially mediates the effect of immediacy on affect towards the artificial teacher. Attention, however, does not show any mediating effect. Overall, the model explained 43.8% of variation in affect towards the artificial teacher. For further clarification, a mediation analysis (based on 1000 bootstrap samples) was conducted. The total effect of immediacy on affect towards the artificial teacher was found to be significant ($\beta = 0.57$, $p < 0.001$). The indirect

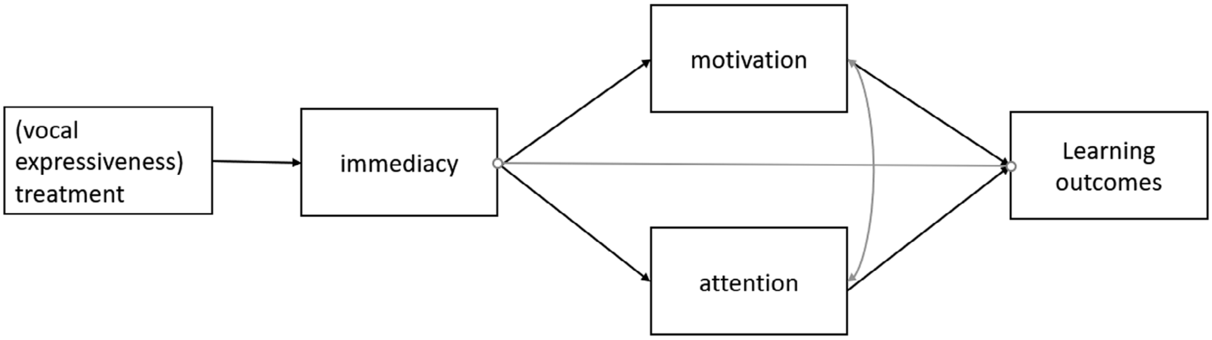

FIGURE 2 Hypothesized path analysis model

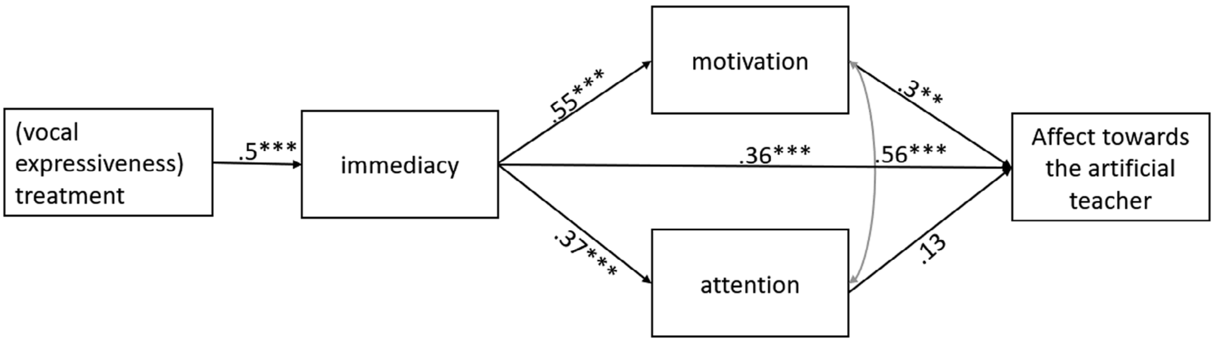

FIGURE 3 Results of the path analysis for affect towards the artificial teacher. Standardized coefficients are presented. Grey lines indicate non-hypothesized relationships, **p* 0.05; RMSEA <0.001; CFI > 0.95

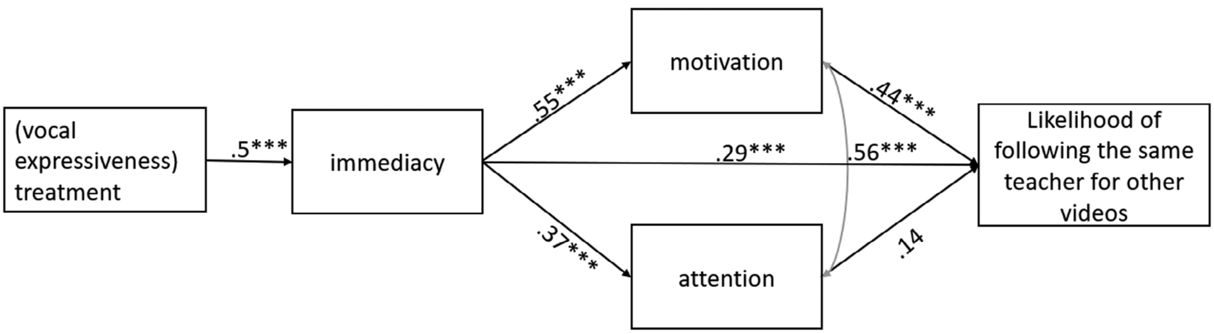

FIGURE 4 Results of the path analysis for likelihood of following the same artificial instructor for other instructional material. Standardized coefficients are presented. Grey lines indicate non-hypothesized relationships, **p* 0.05; RMSEA <0.001; CFI > 0.95

effect through motivation was also significant ($\beta = 0.166$, 95% *CI* [0.048, 0.338]). The indirect effect through attention was non-significant ($\beta = 0.047$, 95% *CI* [−0.020, 0.118]).

Furthermore, to examine whether immediacy, and not likeability mediated the effect of expressiveness on affective learning, a mediation analysis was conducted. Immediacy remained a mediator of the effect of vocal expressiveness on all three affective outcomes even when likeability of artificial agent was included as a second mediator (likeability of artificial agent also mediated the effect of vocal expressiveness on all three forms of affective learning).

Similar results were found for participants' likelihood of following the same artificial instructor for other instructional material, as seen in Figure 4. Again, these results suggest that only motivation partially mediates the effect of immediacy on likelihood of following the same artificial instructor for other instructional material. Overall, the model explained 54.6% of likelihood of following the same artificial instructor for other instructional material. For further clarification, a mediation analysis was conducted (based on 1000 bootstrap samples). The total effect of immediacy on likelihood of following the same artificial instructor for other instructional material was found to be significant ($\beta = 0.58$, $p < 0.001$). The indirect effect through motivation was also significant ($\beta = 0.241$, 95% *CI* [0.138, 0.348]). The indirect effect through attention was non-significant ($\beta = 0.05$, 95% *CI* [−0.001, 0.104]).

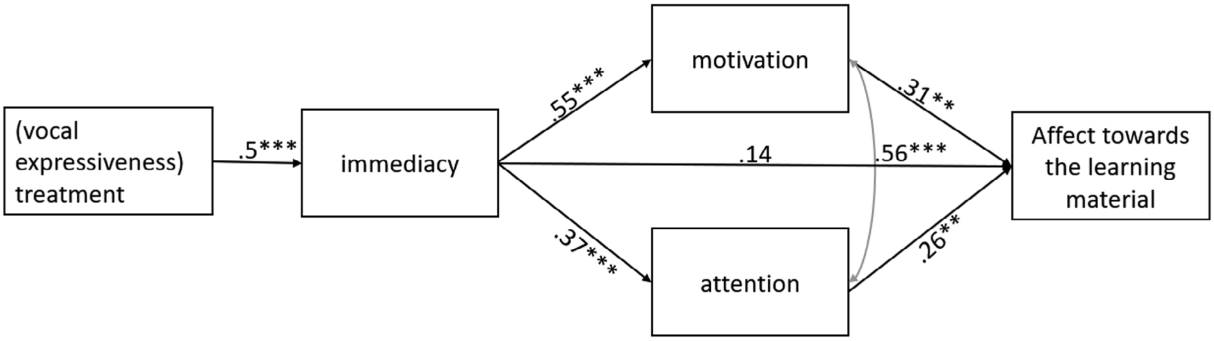

FIGURE 5 Results of the path analysis for affect towards the artificial teacher. Standardized coefficients are presented. Grey lines indicate non-hypothesized relationships, *p 0.05; RMSEA <0.001; CFI > 0.95

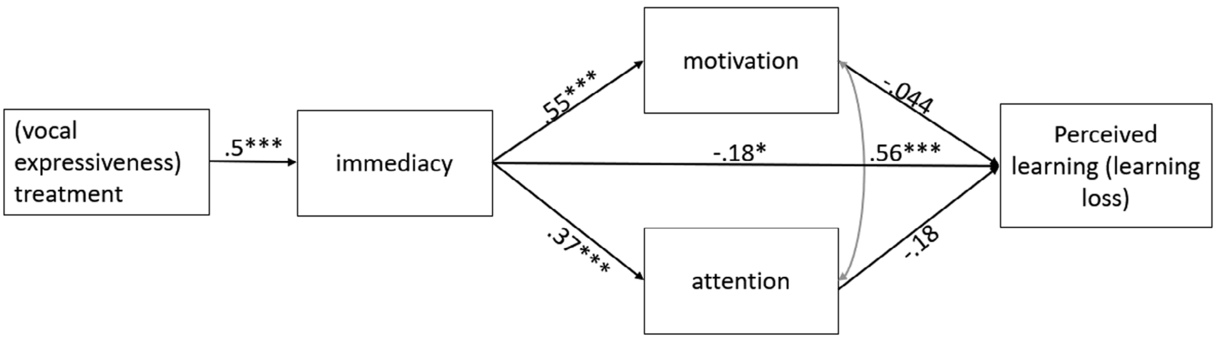

FIGURE 6 Results of the path analysis for perceived cognitive learning (learning loss). Standardized coefficients are presented. Grey lines indicate non-hypothesized relationships, *p 0.05; RMSEA <0.001; CFI > 0.95

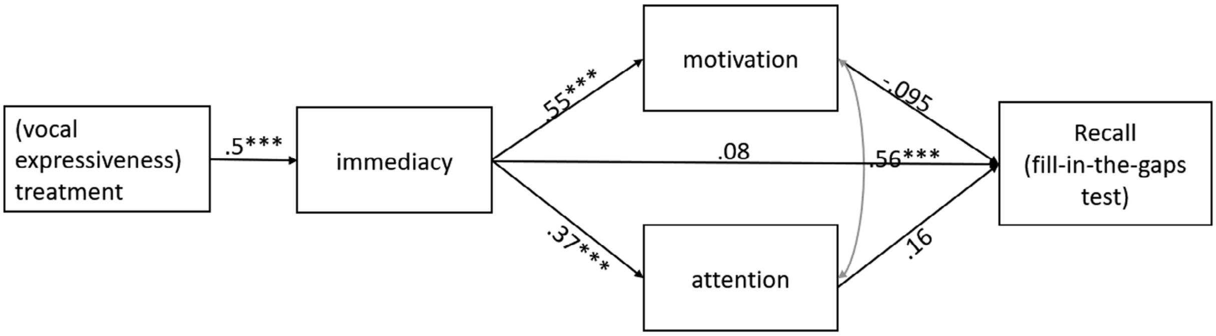

FIGURE 7 Results of the path analysis for recall scores. Standardized coefficients are presented. Grey lines indicate non-hypothesized relationships, *p 0.05; RMSEA <0.08; CFI > 0.95

However, as Figure 5 illustrates, both motivation and attention appeared to be significant positive predictors of affect towards the instructional material, while the path coefficient from immediacy to affect towards the instructional material was positive but no longer significant. We can conclude that affect towards the instructional material is fully mediated by both attention and motivation. Overall, the model explained 36% of variation in affect towards the instructional material. For further clarification, a mediation analysis was conducted (based on 1000 bootstrap samples). The total effect of immediacy on affect towards the instructional material was found to be significant ($\beta = 0.40$, $p < 0.001$). The indirect effect through motivation was also significant ($\beta = 0.172$, 95% *CI* [0.053, 0.291]). Similarly, the indirect effect through attention was significant ($\beta = 0.095$, 95% *CI* [0.019, 0.176]).

To summarize, the results of the path analyses as also further mediation analyses supported our hypothesis 4, showing that motivation explained (part of) the effect of immediacy on affective learning outcomes.

4.3 | Path analyses of cognitive learning outcomes

Concerning cognitive learning, two path analyses were conducted for the two cognitive learning outcomes (perceived cognitive learning and recall scores). Results are presented in Figures 6 and 7. In more detail, as seen in Figure 6, in line with our expectations, immediacy was found to be a negative significant predictor of perceived cognitive learning (it is negative as it has been measured as learning loss). At the same time, the path coefficients from both motivation and attention to perceived cognitive learning were non-significant. These findings suggest that neither motivation nor attention mediated the effect of immediacy on perceived cognitive learning.

Lastly, as illustrated in Figure 7, neither immediacy nor motivation and attention were found to be significant predictors of participants' recall. The results of the path analyses did not provide support for our hypothesis 5, as attention was not found to be a mediator.

5 | DISCUSSION

The current research investigated the influence of an artificial model with strong vocal expressiveness, as compared to the (same) artificial model with weak vocal expressiveness, as a means to increase non-verbal immediacy and subsequently to enhance individuals' affective and cognitive learning. What is more, we tested the proposed underlying mechanisms of motivation and attention to explain the anticipated effect of artificial model's immediacy on affective and cognitive learning respectively.

Our results supported our first hypothesis, showing that an artificial model that shows strong vocal expressiveness can increase perceptions of immediacy, as compared to an artificial model that shows weak vocal expressiveness. This is in accordance with the vast body of literature on human teachers' nonverbal immediacy (i.e., Bambaeeroo & Shokrpour, 2017; Ellis et al., 2016; Uleanya et al., 2020; Witt et al., 2004) but new in the field of research on virtual pedagogical agents. Therefore, the study provides evidence that, similar to a human teacher, strong nonverbal cues, such as vocal expressiveness, can influence learners' perceptions of psychological closeness (i.e., immediacy) with an artificial teacher.

Furthermore, the study's results supported our second hypothesis, showing that an artificial model with strong vocal expressiveness can enhance individuals' affective learning, as compared to an artificial model with weak vocal expressiveness. Specifically, according to the current study's findings, participants in the strong vocal expressiveness condition indicated increased affective evaluation of the instructional material, the artificial teacher, as also an increased likelihood of following the same artificial instructor for other instructional videos. Undeniably, students' affective experiences are important as they have been found to be the central mediator linking teaching behaviours to student reports of learning and other important classroom outcomes (Bolkan, 2015; Grawemeyer et al., 2017; Tyng et al., 2017).

Next, results provided partial support for our third hypothesis, showing that an artificial model that shows strong vocal expressiveness can impact learners' recall when assessed with the fill-in-the-blanks test as compared to an artificial model that shows weak vocal expressiveness. At the same time, no evidence for a difference between the two levels of vocal expressiveness was found when recall was assessed with a multiple-choice test. These mixed results might be an aftermath of the difference between the gap filling and multiple-choice tests as methods of knowledge assessment. Our results are in accordance with earlier studies, which found a significant difference in learners' scores on the two types of tests, with learners' multiple-choice scores to be significantly better than their gap filling scores (Medawela et al., 2018; Utari, 2013). These studies did not test why this is the case, and more exploration of this issue is required.

Additionally, we found evidence that an artificial model with strong vocal expressiveness also affects *perceived* cognitive learning. Specifically, the study's findings provide evidence that strong vocal expressiveness has a positive influence on perceptions of learning as compared to weak vocal expressiveness. Despite the fact that perceived cognitive learning is not as strong as measuring actual cognitive learning (as they have been found to be moderately correlated), it is often used as an indication of the nature of relationships (Chesebro & McCroskey, 2000).

Furthermore, according to our fourth hypothesis (concerning mediating psychological processes of motivation on affective learning), findings of this study suggest that learners' motivation explains part of the effect of immediacy (and, thus, of vocal expressiveness) on affective learning outcomes. Thus, the current study provides support for motivation theory (Christophel, 1990), which argues what this study's findings revealed; strong nonverbal cues, such as vocal expressiveness, increase perceptions of immediacy, which has a positive effect on motivation leading to enhanced affective learning.

However, contrary to our fifth hypothesis (concerning mediating psychological processes of attention on cognitive learning), learners' attention to the instructional material does not appear to explain the effect of immediacy (and indirectly of vocal expressiveness) on cognitive learning (immediate recall) and perceived cognitive learning. Thus, results of the current study do not provide evidence in favour of arousal–attention theory, which posits that immediacy stimulates arousal, which affects attention and memory leading to greater cognitive learning (Kelly & Gorham, 1988). However, other psychological studies demonstrated the important role of arousal in altering both attention and consolidation of memories (Christianson & Loftus, 1991a; Eysenck, 1976a, 1976b; Heuer & Reisberg, 1992; Revelle & Loftus, 1992). Such studies suggest that if arousal acts specifically on memory consolidation, its influence magnifies following a delay, as consolidation is a process that occurs over time. Thus, future research might examine whether attention mediates the effect of immediacy on delayed recall (e.g., one-week past treatment). How the effect of vocal expressiveness on cognitive learning can be explained is still unclear.

Collectively, the current findings are in line with past work that emphasized the vital role of *human* teachers' nonverbal cues in increasing students' affective, cognitive and perceived cognitive learning in traditional classroom settings (Bambaeeroo & Shokrpour, 2017;

Ellis et al., 2016; Uleanya et al., 2020; Witt et al., 2004). Similarly, the study showed that an *artificial* teacher's nonverbal cues are related to learning outcomes because they promote immediacy. Nonetheless, earlier studies examined a plethora of nonverbal cues used by an artificial agent together (e.g., facial cues, posture), as it is difficult to disentangle such cues from each other when human teachers are employed. In addition, the majority of past studies utilized a survey research design, which has been argued to be of limited usefulness when it comes to drawing conclusions about students' learning (Comstock et al., 1995; Hess & Smythe, 2001; Witt & Wheeless, 2001). Therefore, an advantage of the current study is that by employing artificial agents, it was able to experimentally show the single effects of nonverbal cues, such as vocal expressiveness, on immediacy and learning outcomes. All in all, the fact that the same learning mechanisms were found to be effective for both human and artificial teachers (i.e., alignment of these findings with human teachers' past research) suggests that the study's results are pertinent for human teachers, too.

In addition, the study's findings are in accordance with the few studies that have provided evidence of the positive effect of vocal expressiveness of artificial agents (Valetsianos, 2009) and robots (Kory Westlund et al., 2017; Kennedy, Baxter & Balpaeme, 2016). However, the current research goes beyond these earlier studies by examining the underlying mechanisms of the effect of vocal expressiveness on both affective and cognitive outcomes. This is, our study tested and revealed that vocal expressiveness is related to learning outcomes, because it reduces psychological distance, thus, promoting immediacy. What is more, our research further examined motivation and attention as mediators of the path from immediacy to affective and cognitive learning. These mechanisms were reported as potential explanations in earlier studies, though they were not empirically tested (Valetsianos, 2009). Lastly, we show that the effect of the combination of pitch and speech rate on learning has its own importance, as it can help designers' choice of vocal parameters when constructing vocal expressiveness of artificial teachers.

Despite the study's aforementioned advantages, caution is needed in generalizing the results beyond the study's population characteristics. This is because nonverbal cues are highly inferential, and they vary culturally and contextually (Gudykunst et al., 1988). Future research could explore whether different contexts (i.e., geographical location) would produce different results on learning. Furthermore, the artificial model's vocal expressiveness consisted of both pitch (tone and variation) and speech rate. Future research could examine the single effect of each vocal parameter on learning outcomes. Another limitation of the study pertains to the short duration of multimedia learning (~10 min). Though the artificial agent's strong vocal expressiveness was shown to increase affective and cognitive learning, the effects of repeated and prolonged exposure to nonverbal cues are not known.

The results, together with the findings of related research, also have practical implications. Earlier findings (Fountoukidou et al., 2019) indicated that a pedagogical artificial agent that uses behavioural modelling as an instructional method facilitates students' learning outcomes when compared to other instructional methods (with or without an agent) that are traditionally used in online or blended learning. The new findings indicate *how* the pedagogical agent could "behave" as a teacher in blended or online learning in order to increase student motivation, liking of the teacher and the material, as well as some cognitive learning outcomes. Pedagogical artificial agents could be designed in such a way that they demonstrate a strong vocal expressiveness during their teaching, meaning that they exhibit variation in pitch while they talk, speak moderately quickly, and use a somewhat higher pitch. At least the combination of all three elements should lead to a sufficiently strong vocal expressiveness of the pedagogical artificial agent.

In conclusion, the current findings revealed that an artificial agent that shows strong vocal expressiveness increases perceptions of immediacy and thereby enhances learners' affective learning and perceived cognitive learning. The effect of vocal expressiveness on participants recall is unclear. Also, the current study provides evidence that motivation explains part of the effect of immediacy on affective learning. Such findings verify the important role of nonverbal immediacy found in traditional educational settings. Showing that these results also apply to artificial teachers is essential, given that the educational landscape is changing and being reshaped by artificial intelligence (Chassignol et al., 2018). Thus, taking into consideration the role of immediacy in the development of artificial teachers or voice assistants like Alexa, Siri, and Google assistant as a way to enhance learners' affective experience is of imperative value, since such artificial agents will be found more and more in our societies.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12632.

DATA AVAILABILITY STATEMENT

Data available on request from the authors

ORCID

Sofia Fountoukidou https://orcid.org/0000-0001-5651-6223

REFERENCES

Andersen, J. F. (1979). Teacher immediacy as a predictor of teaching effectiveness. *Annals of the International Communication Association*, *3*(1), 543–559. https://doi.org/10.1080/23808985.1979.11923782

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, *94*(2), 416–427. https://doi.org/10.1037/0022-0663.94.2.416

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

Bambaeeroo, F., & Shokrpour, N. (2017). The impact of the teachers' nonverbal communication on success in teaching. *Journal of Advances in Medical Education & Professionalism*, *5*(2), 51. https://doi.org/10.1016/j59

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2008). Measurement instruments for the anthropomorphism, Animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Baylor, A. L., & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior*, *25*(2), 450–457. https://doi.org/10.1016/j.chb.2008.10.008

Bente, G., Krämer, N., & Eschenburg, F. (2008). Is there anybody out there? Analyzing the effects of embodiment and nonverbal behavior in avatar mediated communication. In E. Konijn (Ed.), *Mediated interpersonal communication* (pp. 131–157). Routledge.

Bolkan, S. (2015). Students' affective learning as affective experience: Significance, reconceptualization, and future directions. *Communication Education*, *64*(4), 502–505. https://doi.org/10.1080/03634523.2015.1058963

Breitenstein, C., Lancker, D. V., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion*, *15*(1), 57–79. https://doi.org/10.1080/0269993004200114

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial intelligence trends in education: A narrative overview. *Procedia Computer Science*, *136*, 16–24. https://doi.org/10.1016/j.procs.2018.08.233

Chesebro, J. L., & Mccroskey, J. C. (2000). The relationship between students' reports of learning and their actual recall of lecture material: A validity test. *Communication Education*, *49*(3), 297–301. https://doi.org/10.1080/03634520009379217

Choi, S., & Clark, R. E. (2006). Cognitive and affective benefits of an animated pedagogical agent for learning English as a second language. *Journal of Educational Computing Research*, *34*(4), 441–466. https://doi.org/10.2190/a064-u776-4208-n145

Christianson, S., & Loftus, E. F. (1991a). Remembering emotional events: The fate of detailed information. *Cognition & Emotion*, *5*(2), 81–108. https://doi.org/10.1080/02699939108411027

Christianson, S. Å., & Loftus, E. F. (1991a). Remembering emotional events: The fate of detailed information. *Cognition & Emotion*, *5*(2), 81–108.

Christophel, D. M. (1990). The relationships among teacher immediacy behaviors, student motivation, and learning. *Communication Education*, *39*(4), 323–340. https://doi.org/10.1080/03634529009378813

Comstock, J., Rowell, E., & Bowers, J. W. (1995). Food for thought: Teacher nonverbal immediacy, student learning, and curvilinearity. *Communication Education*, *44*(3), 251–266. https://doi.org/10.1080/03634529509379015

Dehn, D. M., & Mulken, S. V. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, *52*(1), 1–22. https://doi.org/10.1006/ijhc.1999.0325

Ellis, R. B., Carmon, A., & Pike, C. (2016). A review of immediacy and implications for provider–patient relationships to support medication management. *Patient Preference and Adherence*, *9*, 1. https://doi.org/10.2147/ppa.s95163

Eysenck, M. W. (1976a). Arousal, learning, and memory. In *Arousal, learning and memory*. University of London.

Eysenck, M. W. (1976b). Arousal, learning, and memory. *Psychological Bulletin*, *83*(3), 389–404.

Fountoukidou, S., Ham, J., Matzat, U., & Midden, C. (2019). Effects of an artificial agent as a behavioral model on motivational and learning outcomes. *Computers in Human Behavior*, *97*, 84–93. https://doi.org/10.1016/j.chb.2019.03.013

Frechette, C., & Moreno, R. (2010). The roles of animated pedagogical Agents' presence and nonverbal communication in multimedia learning environments. *Journal of Media Psychology*, *22*(2), 61–72. https://doi.org/10.1027/1864-1105/a000009

Gorham, J. (1988). The relationship between verbal teacher immediacy behaviors and student learning. *Communication Education*, *37*(1), 40–53. https://doi.org/10.1080/03634528809378702

Grawemeyer, B., Mavrikis, M., Holmes, W., Gutiérrez-Santos, S., Wiedmann, M., & Rummel, N. (2017). Affective learning: Improving engagement and enhancing learning with affect-aware feedback. *User Modeling and User-Adapted Interaction*, *27*(1), 119–158. https://doi.org/10.1007/s11257-017-9188-z

Gudykunst, W. B., Ting-Toomey, S., & Chua, E. (1988). *Culture and interpersonal communication*. Sage Publications.

Handley, M. A., Lyles, C. R., McCulloch, C., & Cattamanchi, A. (2018). Selecting and improving quasi-experimental designs in effectiveness and implementation research. *Annual Review of Public Health*, *39*, 5–25. https://doi.org/10.1146/annurev-publhealth-040617-014128

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

Hess, J. A., & Smythe, M. J. (2001). Is teacher immediacy actually related to student cognitive learning? *Communication Studies*, *52*(3), 197–219. https://doi.org/10.1080/10510970109388554

Heuer, F., & Reisberg, D. (1992). Emotion, arousal, and memory for detail. In S. Christianson (Ed.), *The handbook of emotion and memory* (pp. 151–164). Erlbaum.

Heuer, F., & Reisberg, D. (1992). Emotion, arousal, and memory for detail. In I. S. Christianson (Ed.), *The handbook of emotion and memory: Research and theory* (pp. 1151–1180). Erlbaum Associates.

Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research*, *15*(1), 155–159. https://doi.org/10.1044/jshr.1501.155

Hsiao, T., Solomon, N. P., Luschei, E. S., & Titze, I. R. (1994). Modulation of fundamental frequency by laryngeal muscles during vibrato. *Journal of Voice*, *8*(3), 224–229. https://doi.org/10.1016/s0892-1997(05)80293-0

Kelly, D. H., & Gorham, J. (1988). Effects of immediacy on recall of information. *Communication Education*, *37*(3), 198–207. https://doi.org/10.1080/03634528809378719

Kennedy, J., Baxter, P., & Belpaeme, T. (2016). Nonverbal Immediacy as a Characterisation of Social Behaviour for Human–Robot Interaction. *International Journal of Social Robotics*, *9*(1), 109–128. https://doi.org/10.1007/s12369-016-0378-3

Kim, Y., & Baylor, A. (2015). Research-Based Design of Pedagogical Agent Roles: A Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education*, *26*, 160–169. https://doi.org/10.1007/s40593-015-0055-y

Krämer, N. C., & Bente, G. (2010). Personalizing e-learning. The social effects of pedagogical agents. *Educational Psychology Review*, *22*(1), 71–87. https://doi.org/10.1007/s10648-010-9123-x

Liew, T. W., Tan, S., Tan, T. M., & Kew, S. N. (2020). Does speaker's voice enthusiasm affect social cue, cognitive load and transfer in multimedia learning? *Information and Learning Sciences*, *121*(3/4), 117–135. https://doi.org/10.1108/ils-11-2019-0124

Liew, T. W., Zin, N. A. M., Sahari, N., & Tan, S. M. (2016). The effects of a pedagogical agent's smiling expression on the learner's emotions and motivation in a virtual learning environment. *The International Review of Research in Open and Distributed Learning*, *17*(5).

Linek, S. B., Gerjets, P., & Scheiter, K. (2010). The speaker/gender effect: Does the speaker's gender matter when presenting auditory text in multimedia messages? *Instructional Science*, *38*(5), 503–521. https://doi.org/10.1007/s11251-009-9115-8

Liu, W. (2021). Does teacher immediacy affect students? A systematic review of the association between teacher verbal and non-verbal immediacy and student motivation. *Frontiers in Psychology*, *12*, 2475. https://doi.org/10.3389/fpsyg.2021.713978

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

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

Mccroskey, J. C., Richmond, V. P., & Bennett, V. E. (2006). The relationships of student end-of-class motivation with teacher communication behaviors and instructional outcomes. *Communication Education*, *55*(4), 403–414. https://doi.org/10.1080/03634520600702562

Medawela, R. S., Ratnayake, D. R., Abesinghe, W., Udari, M., Abeyasinghe, L., Jayasinghe, R. D., & Marambe, K. N. (2018). Effectiveness of "fill in the blanks" over multiple choice questions in assessing final year dental undergraduates. *Educación Médica*, *19*(2), 72–76. https://doi.org/10.1016/j.edumed.2017.03.010

Mehrabian, A. (1981). *Silent messages: Implicit communication of emotions and attitudes*. Wadsworth Pub.

Mizuno, O., & Nakajima, S. Y. (1998). A new synthetic speech/sound control language. In *Fifth international conference on spoken language processing*, 5(2007–2010). Proceedings of ICSLP98.

Revelle, W., & Loftus, D. A. (1992). The implications of arousal effects for the study of affect and memory. In S.-Å. Christianson (Ed.), *The handbook of emotion and memory: Research and theory* (pp. 113–149). Lawrence Erlbaum Associates, Inc.

Rey, G. D., & Steib, N. (2013). The personalization effect in multimedia learning: The influence of dialect. *Computers in Human Behavior*, *29*(5), 2022–2028. https://doi.org/10.1016/j.chb.2013.04.003

Richmond, V. P., Gorham, J. S., & Mccroskey, J. C. (1987). The relationship between selected immediacy behaviors and cognitive learning. *Annals of the International Communication Association*, *10*(1), 574–590. https://doi.org/10.1080/23808985.1987.11678663

Richmond, V. P., Mccroskey, J. C., & Johnson, A. D. (2003). Development of the nonverbal immediacy scale (NIS): Measures of self-and other-perceived nonverbal immediacy. *Communication Quarterly*, *51*(4), 504–517. https://doi.org/10.1080/01463370309370170

Rosenberg-Kima, R. B., Baylor, A. L., Plant, E. A., & Doerr, C. E. (2008). Interface agents as social models for female students: The effects of agent visual presence and appearance on female students' attitudes and beliefs. *Computers in Human Behavior*, *24*(6), 2741–2756. https://doi.org/10.1016/j.chb.2008.03.017

Schneider, S., Nebel, S., Pradel, S., & Rey, G. D. (2015). Introducing the familiarity mechanism: A unified explanatory approach for the personalization effect and the examination of youth slang in multimedia learning. *Computers in Human Behavior*, *43*, 129–138. https://doi.org/10.1016/j.chb.2014.10.052

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

Scott, M. D., Yates, M. P., & Wheeless, L. R. (1975). An Exploratory Investigation of the Effect of Communication Apprehension in Alternate Systems of Instruction. Paper presented at the International Communication Association, Chicago, IL.

Servilha, E. A., & Costa, A. T. (2015). Conhecimento vocal e a importância da voz como recurso pedagógico na perspectiva de professores universitários. *Revista CEFAC*, *17*(1), 13–26. https://doi.org/10.1590/1982-0216201514813

Simonds, B. K., Meyer, K. R., Quinlan, M. M., & Hunt, S. K. (2006). Effects of instructor speech rate on student affective learning, recall, and perceptions of nonverbal immediacy, credibility, and clarity. *Communication Research Reports*, *23*(3), 187–197. https://doi.org/10.1080/08824090600796401

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, *32*(1/2), 9–31. https://doi.org/10.1023/b:truc.0000021808.72598.4d

Tyng, C. M., Amin, H. U., Saad, M. N., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, *8*, 1454. https://doi.org/10.3389/fpsyg.2017.01454

Uleanya, M. O., Uleanya, C., Taiwo, M. B., & Shobiye, T. (2020). Impact of non-verbal communication skills on learning outcomes in English language among senior secondary school students. *African Journal of Gender, Society and Development (formerly Journal of Gender, Information and Development in Africa)*, *9*(2), 33–58. https://doi.org/10.31920/2634-3622/2020/9n2a2

Utari, A. Y. S. (2013). The usage of multiple choice and gap filling in measuring grade-schoolers' understanding of grammar. *Journal of English Language Teaching in Indonesia*, *1*(1), 69–79.

Veletsianos, G. (2009). The impact and implications of virtual character expressiveness on learning and agent-learner interactions. *Journal of Computer Assisted Learning*, *25*(4), 345–357. https://doi.org/10.1111/j.1365-2729.2009.00317.x

Wang, F., Li, W., Mayer, R. E., & Liu, H. (2018). Animated pedagogical agents as aids in multimedia learning: Effects on eye-fixations during learning and learning outcomes. *Journal of Educational Psychology*, *110*(2), 250–268. https://doi.org/10.1037/edu0000221

Westlund, J. M., Jeong, S., Park, H. W., Ronfard, S., Adhikari, A., Harris, P. L., … Breazeal, C. L. (2017). Flat vs. expressive storytelling: Young Children's learning and retention of a social Robot's narrative. *Frontiers in Human Neuroscience*, *11*, 11. https://doi.org/10.3389/fnhum.2017.00295

Witt, P. L., & Wheeless, L. R. (2001). An experimental study of teachers' verbal and nonverbal immediacy and students' affective and cognitive learning. *Communication Education*, *50*(4), 327–342. https://doi.org/10.1080/03634520109379259

Witt, P. L., Wheeless, L. R., & Allen, M. (2004). A meta-analytical review of the relationship between teacher immediacy and student learning. *Communication Monographs*, *71*(2), 184–207. https://doi.org/10.1080/036452042000228054

Woo, H. (2009). Designing multimedia learning environments using animated pedagogical agents: Factors and issues. *Journal of Computer Assisted Learning*, *25*(3), 203–218. https://doi.org/10.1111/j.1365-2729.2008.00299.x

Yi, M. Y., & Davis, F. D. (2003). Developing and validating an observational learning model of computer software training and skill acquisition. *Information Systems Research*, *14*(2), 146–169. https://doi.org/10.1287/isre.14.2.146.16016

How to cite this article: Fountoukidou, S., Matzat, U., Ham, J., & Midden, C. (2022). The effect of an artificial agent's vocal expressiveness on immediacy and learning. *Journal of Computer Assisted Learning*, *38*(2), 500–512. https://doi.org/10.1111/jcal.12632