



Unveiling the Power of Voice! The Impact of Educational Robots' Voice Gender and Communication Style on Users' Perception, Visual Attention and Cerebral Activity

Chen Fang^{1,2} · Fu Guo¹ · Xuegong Bao¹ · Zeyu Zhang¹ · Tony Belpaeme²

Received: 14 June 2025 / Revised: 13 May 2026 / Accepted: 3 August 2026
© The Author(s), under exclusive licence to Springer Nature B.V. 2026

Abstract

Despite increasing interest in and widespread use of robots in education, there is limited research on the effect of voice gender and communication style on affective outcomes. Based on a Q&A interaction model, a within-subject experiment was conducted with a 2 × 2 design, manipulating voice gender (male vs. female) and communication style (functional vs. relational). The study combined subjective evaluations with eye-tracking and fNIRS measurements. We confirmed the effects of voice gender and communication style on subjective perceptions, reported the influence of voice gender on users' visual attention, and demonstrated the interaction between voice gender and communication style on users' cerebral activity. The results suggest that the female voice and relational communication style are preferable for educational robots. This study enhances the understanding of users' visual attention and cerebral activity during interactions with educational robots, providing valuable insights for designing robot voices in educational contexts.

Keywords Human-robot interaction · Educational robot · Voice gender · Communication style · Eye-tracking · fNIRS

1 Introduction

Social robots are gaining popularity in the educational domain. For instance, social robots like Pepper and Nao have been applied to assist classroom teaching in primary schools [1], support teaching for autistic children outside the classroom [2], and introduce academic writing in higher education [3]. These robots are expected to benefit learners by promoting positive attitudes, interest, motivation, and satisfaction, which in turn can enhance students' engagement and knowledge acquisition [4]. A key advantage of humanoid robots in education is their ability to engage in social interactions with students through, among other channels, voice and gestures. While positive impacts of social interaction on learning have been demonstrated in human-robot

interaction (HRI) from both affective and cognitive outcomes [5], unnatural or inappropriate robot behaviors can lead to negative outcomes, such as distraction [6]. Voice, as a vital cue in social interaction, both conveys information and expresses affect. The information about the speaker's gender [7], communication style [8], and affection [9] delivered by voice communication can impact users' perceptions. Consequently, the voices of robots may influence affective and cognitive outcomes in HRI in educational contexts.

Various voice factors have been explored and found to play critical roles in HRI, including voice characteristics (i.e., pitch, speed, and gender) and speech content (i.e., communication style) [7, 8]. Nevertheless, research on the impact of voice gender of robots in HRI, particularly in educational contexts, remains limited. Existing conclusions on the influence of voice gender on users' perceptions are inconsistent. For example, McGinn and Torre [10] and Eyssel et al. [11] reported a strong effect of voice gender on users' perceptions of robots, while Song et al. [7] did not find a significant effect. Nonetheless, it is plausible that voice gender may influence users' perceptions, as gender stereotypes inferred from robot voices may impact human-robot interaction [12, 13]. Additionally, instructors' communication

✉ Fu Guo
fguo@mail.neu.edu.cn

¹ Department of Industrial Engineering, School of Business Administration, Northeastern University, NO. 195, Chuangxin Road, Hunnan District, Shenyang 110167, China

² Idlab-AIRO, Ghent University – Imec, Ghent, Belgium

style is critical for creating positive learning experiences. In real educational settings, some instructors prioritize the delivery of educational content, while others emphasize the importance of relationship-building when delivering learning content. Kim et al. [14] found that students preferred an AI instructor with a relational-oriented communication style over one with a functional-oriented style. Although these studies contribute to understanding the effect of voice gender and communication style, they examined these factors in isolation. Therefore, it is important to explore whether and how the combination of voice gender and communication styles of robots affects users' perceptions in educational contexts. Such findings could enhance the understanding of how robots' voice interactions shape user perceptions and offer valuable insights for designing educational robots that foster positive affective outcomes.

Affective outcomes, which are important outcomes of robot interventions in educational settings, encompass qualities beyond traditional learning outcomes per se [5], such as a learner's acceptance, trust, learning effort, and willingness to use. In the educational context of HRI, affective outcomes are varied and are commonly measured through self-reports. For example, willingness to use learning tools is one of such outcomes. Reich-Stiebert and Eyssel [15] developed a scale to evaluate users' future willingness to engage with educational robots. Similarly, trust, regarded as a driver of technology acceptance, is one of the important contents of affective outcomes in the interaction between students and educational robots. Qin et al. [16] adopted the interviews to determine factors affecting users' trust in AI-based educational tools. Learning effort, another affective outcome, refers to the amount of time and cognitive resources invested in the learning process [17]. Chen et al. [17] developed an item to evaluate users' learning effort during interaction with an animal companion system for learning. Additionally, understanding users' preferences towards robot voices can inform the design of educational robots that are appealing to users. Therefore, we selected willingness to use, trust, learning effort, and preference as the dependent variables to reflect affective outcomes.

The aforementioned subjective measurements of affective outcomes in HRI within educational contexts rely heavily on users' subjective perceptions. However, subjective methods alone are insufficient for collecting users' momentary inner states and behaviors. Eye-tracking technology, commonly used to evaluate the quality of human-robot interaction [18], can capture and record users' real-time eye movements and extract eye-moving metrics that quantify users' visual behaviors. Users' eye-gaze behavior can serve as a behavioral indicator of attention and task engagement during educational interactions with robots [19]. The fixation duration ratio, which refers to the ratio of the sum of

fixation duration on the specific area of interest (AOI) to the total fixation time of the entire stimulus, is used to assess users' visual attention and interest. Higher fixation duration ratios indicate greater attention to the AOI [20]. Visual attention competition between the robot and the learning material is crucial for effective learning. Adopting eye-tracking technology to reveal users' visual attentional patterns is significant. Additionally, technologies such as electroencephalogram (EEG) and functional near-infrared spectroscopy (fNIRS), which collect real-time signals to reflect human cerebral activity, are increasingly used to offer insights into neural mechanisms associated with attention and affective engagement. For instance, Ren et al. [21] used self-report, eye-tracking, and EEG technology to examine users' perception of robot facial features. Guo et al. [22] combined self-report and fNIRS technology to evaluate users' emotional responses toward robot touch. Therefore, our study combines subjective methods, eye-tracking, and fNIRS technologies to investigate the effects of educational robots' voice gender and communication style on affective outcomes. The combination of subjective and objective methods helps unravel the subjective perception, visual attention, and cerebral activity of users during interaction with educational robots. These findings will deepen our understanding of users' perceptions and provide guidelines for designing effective robot voices in educational settings.

2 Theoretical Background and Research Questions

2.1 The Impact of Voice Gender

Voice gender is an important characteristic of robots' voice that has received attention in HCI studies, but has received relatively little attention in HRI [23, 24]. There is considerable disagreement over the conclusions about voice gender in previous studies in HRI. Specifically, some research emphasized the critical role of voice gender on users' perception of robots [10, 11], while other studies did not find a significant main effect of voice gender [7]. These mixed results appear across various fields of robotics applications. To our knowledge, experimental studies examining the role of the voice gender of robots in educational contexts are limited.

Voice gender, as a social trait, is important for interpersonal relationships and can evoke social stereotypes that trigger specific responses from users in HRI. Firstly, the social robots' gender can help establish common ground between humans and robots, facilitating more intuitive interaction [13]. Secondly, gender stereotypes are often activated automatically [25]. Individuals hold biases about

certain occupations based on gender, which may influence their perceptions of those who do not conform to these stereotypes in human-human interactions. In educational contexts, stereotypes about male and female teachers are prevalent, and a similar trend exists in HRI [13]. In addition, the voice gender of robots can influence their persuasive power [26] and task suitability [13]. Male voices typically have a lower pitch than female voices [27], which leads to perceptions of higher maturity, dominance, assertiveness, and lower warmth and agreeableness compared to female voices [9]. Female voices are generally perceived as more attractive, particularly by male users [9, 28]. Chang et al. [29] found that elderly individuals preferred female-extroverted robot voices for home companionship. Dou et al. [30] also suggested that users have varying affective evaluations of robots' voices across different application fields. They found that adult female and male voices are the most acceptable voice for educational robots, while adult male and child voices are preferred for retail and customer-facing robots and home companion robots. Hence, it is reasonable to speculate that the voice gender of robots may affect users' perceptions and subsequent interactions in educational contexts.

2.2 The Impact of Communication Style

In human-human interaction, an instructor's communication style plays a vitally important role in shaping positive learning experiences [31] and impacting students' affective, cognitive, and behavioral outcomes [32]. Based on different goal orientations, teachers' communication strategies can be broadly categorized into two types: task-oriented and relational-oriented [33]. The task-oriented strategy focuses on the content and tasks the instructor aims to convey, emphasizing clarity and content relevance. In other words, teachers using a task-oriented approach tend to deliver content in a direct way. Relational-oriented strategy, by contrast, prioritizes building relationships and fostering ongoing connections between the instructor and students. In this context, communication style in education can be classified as either functional or relational [14]. Frisby and Buckner [34] reported that students' learning behaviors and efficacy improved when teachers developed and maintained positive relationships with their students.

Similarly, the communication style of educational robots in HRI can be broadly understood as lying along two dimensions: a functional one and a relational one [35]. Robots with a functional communication style are task-oriented and focus on delivering content efficiently, while those with a relational communication style prioritize meeting users' emotional needs and fostering great relationships with users. Nevertheless, research has investigated

the different communication styles in HRI and has yielded mixed results. Some studies suggested that the relational style is more effective than the functional style in enhancing users' satisfaction [35, 36]. For example, in the educational context, Kim et al. [14] indicated that AI instructors with a relational style received more favorable ratings from students compared to those using a functional style during online teaching. Yet, this needs to be balanced against the effectiveness as teaching aids. Studies have shown that functional communication styles can be more effective [37, 38]. For instance, older adults exhibited better cognitive and functional outcomes when interacting with conversational digital assistants that used a functional style, rather than a relational one [38]. Generally, functional style in chatbots conveys "low warmth but high competence", while relational style is perceived as "high warmth but low competence" [37]. Xu et al. [37] suggested that users with low competency prefer functional style, whereas those with high competency favor relational style when interacting with AI-based shopping assistants.

2.3 Multidimensional Measurement in HRI

In light of the above, most existing studies have reported insightful yet mixed findings regarding the effects of voice gender or communication style, largely based on subjective measurement methods. The self-reported techniques provide only a partial understanding of users' responses during HRI. To obtain a more comprehensive picture, multidimensional measurement approaches that combine subjective and objective indicators have become increasingly common in HRI research [22, 39, 40]. For example, eye-tracking is widely used to capture users' visual attention and reveal attentional allocation during interaction, offering behavioral evidence that complements self-reported results. Likewise, neurophysiological techniques such as fNIRS and EEG enable researchers to infer the cognitive or emotional processing through the detection of neural activity in relevant brain regions.

In educational HRI, users' visual attention reflects task engagement, attentional competition between the robot and the learning material, and provides meaningful insight into how they allocate cognitive resources during learning [19]. Prior research has shown that female voices are more attractive than male voices [9, 28], implying that the robot's voice gender may influence how users allocate their visual attention during learning. Similarly, the impact of communication style has been partially revealed. Compared with a task-oriented communication style, a relational-oriented communication style places greater emphasis on maintaining social connections and is generally perceived as warmer [37]. This increased sense of warmth may heighten users'

perception of the robot's social qualities, thereby drawing more visual attention to the robot rather than the learning task itself. In parallel, cerebral activity offers a complementary understanding of users' cognitive and affective processes. Cerebral activity in the left DLPFC is associated with positive emotions [41–43] and off-task thoughts [44]. The FPC plays a key role in higher-order cognitive operations, particularly those involved in evaluating one's own judgments in relation to those of others [45, 46]. Neurophysiological studies suggested that communicative cues can elicit different neural responses [39, 40, 47]. Accordingly, both voice gender and communication style may evoke different levels of engagement at the neural level.

2.4 Research Questions and Overview of Studies

In light of the above, we proposed the following three research questions:

RQ1. Do the voice gender and communication style of educational robots impact users' subjective perceptions?

RQ2. Do the voice gender and communication style of educational robots influence users' visual attention?

RQ3. Do the voice gender and communication style of educational robots impact users' cerebral activity?

Addressing these research questions allows the study to offer multidimensional empirical evidence regarding the influence of the robot's voice gender and communication styles on users' perceptions and attention. In general, this study reveals that educational robots with female voices and relational communication styles help shape positive affective outcomes during the learning interaction. This is demonstrated by higher ratings of trust, learning effort, willingness to use, and preference, along with increased visual attention to the robot's screen (learning material) and heightened cerebral activity in the FPC. Moreover, this study reveals users' visual attention patterns toward the robot during the learning process, showing that the robot's head and trunk compete with the learning task (present on the robot's screen) for the learner's visual resources. Furthermore, while our overall results favor female voices and relational communication styles, the cerebral activity detected in the left dorsolateral prefrontal cortex (DLPFC) suggests that male voices may be comparatively more suited to functional communication styles. Note that learning outcomes are not the subject of this study, as learning gains have been convincingly demonstrated in earlier studies [5].

3 Methods

A 2 (voice gender: male vs female) $\times$ 2 (communication style: functional vs relational) within-subjects experiment with the Pepper robot was conducted, combining self-reports, eye-tracking, and fNIRS technology to collect both subjective and physiological data.

3.1 Participants

Twenty-nine participants (15 males, $M_{\text{age}}=22.552$, $SD_{\text{age}}=3.203$) were recruited from the student population of a local university. A power analysis indicated that the minimum required sample size was 24, using G*power v3.1.9.7 for F-test repeated measures with a power of 0.8, a medium effect size of 0.25, and an α level of 0.05 [48, 49]. Thus, with 29 participants the study achieved sufficient power. All participants were healthy with normal or corrected-to-normal visions and normal hearing. They were unfamiliar with Pepper and participated voluntarily, and received a gift as compensation after the experiment. The experiment was approved by Northeastern University's ethical committee.

3.2 Robot Platform and Experimental Tasks

A social robot (Pepper, Aldebaran) was used in this study, which has proven its effectiveness in educational contexts [50]. Pepper is 120 cm tall and can be reprogrammed to offer different interaction styles. It can integrate voice packages from other speech synthesis platforms, enabling it to use a variety of voices.

We focused on the role of robots as teaching assistants because this role is widely used in the educational domain [5]. To engage students in learning, we adopted an interaction model adapted from the Q&A interaction framework developed by You et al. [51]. The Q&A model is one of the common models of collaboration between students and education robots. In this study, participants were required to collaborate with the Pepper robot to complete the Question and Answer tasks. The test questions were displayed on Pepper's chest-mounted screen. Participants were asked to judge whether each test item was correct or incorrect, after which Pepper provided feedback using the designed voices according to whether their answers were correct or incorrect. To isolate the affective outcomes of the robot's voice gender and communication style, test performance was experimentally controlled using a simulated Q&A task in which participants received preset answers [52, 53]. This approach ensured that every participant received an equal number of correct and incorrect feedback instances across all conditions, thereby preventing performance-related factors, such as frustration following errors or satisfaction after

Table 1 The Q&A script

The content of the knowledge tests	Participants' answers
The biosphere is the largest ecosystem on Earth.	I think this sentence is wrong.
Plant cells have additional structures than animal cells, including cell walls, vacuoles, and chloroplasts.	I think this sentence is right.
The ootid stores a large amount of nutrients and has a large volume, which is conducive to the exchange of substances with the outside world.	I think this sentence is right.
Bacteria have no proteins, just naked DNA.	I think this sentence is wrong.

Table 2 Robot's voice script (English translated from original mandarin Chinese)

Communication Style	Voice content
Functional	Correct answer. Incorrect answer.
Relational	Congratulation! Your answer is correct. Keep it up! Regrettably! Your answer is incorrect. Better luck next time!

**Fig. 1** Experimental scene

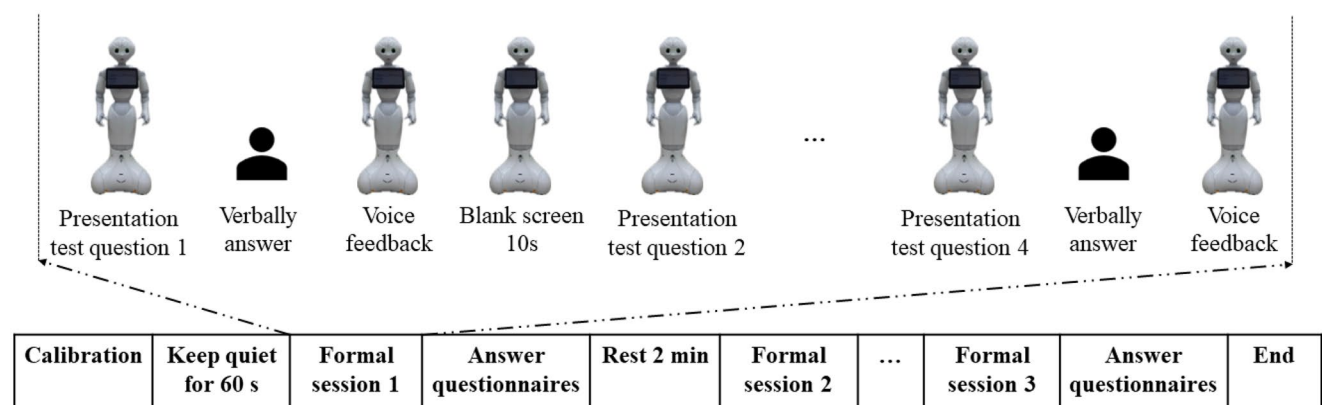
success, from confounding the affective outcomes. Note that this study only focuses on the affective outcomes rather than the cognitive outcomes.

The Q&A script is shown in Table 1. The robot voice was generated through the Mandarin Text-To-Speech system developed by iFlytek (Hefei, China) and then integrated into

the Pepper robot by Choregraphe v2.5.11.13. The robot's voice scripts are shown in Table 2.

3.3 Procedure

The experiment was conducted in the Human Factors Engineering Lab with light control, as shown in Fig. 1. The specific procedure of the experiment is shown in Fig. 2. Firstly, participants were welcomed, informed of the experiment contents, and voluntarily filled in the informed consent form. Then, they were seated in a chair facing Pepper. Subsequently, participants were equipped with the fNIRS probes and eye-tracking glasses. Eye-tracking calibration was performed and typically required 10 to 20 seconds per attempt, with repeated attempts when necessary. Subsequently, fNIRS channel-quality calibration was conducted, lasting at least 2 minutes per participant and frequently requiring multiple attempts until all channels achieved acceptable signal levels. Throughout this time, participants were asked to sit still and keep their eyes closed, and their minds rested [54]. Then, participants completed four formal sessions. The four formal sessions consisted of the four experimental conditions in a sequence of Latin Square design. Before each session, participants were asked to keep quiet for 60 seconds to collect a stable baseline of physiological signals. In each session, participants completed four knowledge tests. For each test, participants would see the test question on the screen on Pepper's chest and then verbally answer the questions according to the Q&A script that had been shown to

**Fig. 2** Experimental procedure

participants before. Subsequently, Pepper gave voice feedback to participants on whether the answer was correct. After each session, participants completed questionnaires immediately and had 2 min to rest. The participants were unaware that the robot was being controlled using a Wizard-of-Oz method, with the robot being teleoperated from another room. The experiment lasted approximately 30 to 40 minutes.

3.4 Measurements

3.4.1 Subjective Perception

To investigate users' subjective perception, perceived trust, perceived learning effort, willingness to use, and affective evaluation were measured. The items of the subjective evaluation questionnaire are shown in Table 3. The perceived trust was assessed using the items adapted from Song et al. [55]. Reliability analyses showed a great Cronbach's alpha value of 0.959. To measure perceived learning effort, we used an item from Chen et al. [17]. Willingness to use was measured by the items of Choi and Ji [56]. The Cronbach's alpha value is 0.976, which shows great reliability. The subjective data were tested using 2×2 repeated measures analysis of variance (ANOVA) in SPSS v25. The Bonferroni method was applied to correct the results for multiple comparisons.

3.4.2 Visual Attention

To measure users' visual attention during interaction with the robot, an SMI wireless Eye Tracking Glasses 2 (Senso Motoric Instruments, Teltow) was used. The eye-tracker was calibrated in a three-point method, with a tracking

accuracy of 0.5° and a sampling rate of 60 Hz. Eye tracking data were analyzed using Begaze 3.6 software. To analyze eye tracking metrics, we defined four areas of interest (AOIs) frame-by-frame, namely, the area of the robot's head (AOI-1), robot's screen (AOI-2), robot's left and right arms (AOI-3), and robot's trunk (AOI-4). The fixation duration ratio, which means the ratio of total fixation duration on the AOI to the total duration the AOI is presented, was extracted from each AOI on each condition to examine users' selective attention. The fixation duration ratio was tested through 2×2 repeated measures ANOVA in SPSS v25. The Bonferroni method was applied to correct the results of multiple comparisons and simple effect analysis.

3.4.3 Cerebral Activity

The LIGHT-NIRS system (Shimadzu Corporation Co.), which involves 22 channels consisting of eight light sources (emitting near-infrared light at 780, 805, and 830 nm) and eight detectors, was utilized to measure the cerebral activity of prefrontal cortex regions (Fig. 3). The center of the below probe set row was placed at the Frontal Polar Midline (FPZ) point in the 10–10 international system. The fNIRS signals were recorded at a 13.333 Hz sampling frequency. Besides, a FASTRAK system (Polhemus Co.) was used to localize the spatial coordinates of optical probes with five reference points (NZ, CZ, IZ, AL, and RL). Based on the estimation result of MNI space [57], we focused on five regions of interest (ROIs) (the definition shown in Supplementary Appendix Table A1), including the left DLPFC¹, the right DLPFC, the left Broca's area, the right Broca's area and the Frontopolar area (FPC). The fNIRS data were preprocessed in the NIRS_KIT v2.0 toolbox [58] in MATLAB (2020b). We focused on HbO₂ signals for analysis as its high signal-to-noise ratio and its sensitivity to task response [59]. First, the optical density changes were transformed into concentration changes of oxyhemoglobin (HbO₂), deoxyhemoglobin (HbR), and total hemoglobin (HbT) according to the modified Beer-Lambert law [60]. Then, the HbO₂ signals were removed the artefacts using 1st-order polynomial regression, were corrected motion using correlation-based signal improvement (CBSI) method [61], and band-pass filtered with a third-order IIR filter of 0.01–0.08 Hz. Four predictors were generated based on the onset and duration of four experimental conditions using the general linear model to model the hemodynamic response. Then, the β values of four experimental conditions were exported and tested by 2×2 repeated measures ANOVA in SPSS v25. The Bonferroni method was applied to correct the results for multiple comparisons.

Table 3 The items of the subjective evaluation questionnaire

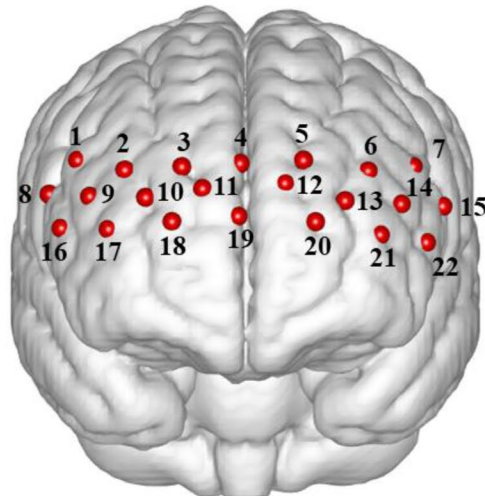
Subjective perception	Description	Items
Trust	Participants' perceived trust.	I think this robot is credible. I think this robot is sincere. I think this robot is honest. I think this robot is convincing.
Willingness to use	Participants' willingness to use the robot to study.	I intend to use the robot in the future. I expect that I will use the robot in the future. I plan to use the robot in the future.
Learning effort	The degree of effort participants put into the learning process.	Do I make more effort to learn with the robot?
Preference	Participants' preference towards the robot for studying.	Non-preferred - Preferred

Note: All items are 7-point Likert scale

¹ Dorsolateral prefrontal cortex.



(a) Sensor layout for LIGHTNIRS device with a source-detector distance of 30mm



(b) Visual representation of the channels

Fig. 3 Sensor placement and visual representation. Five ROIs were categorized: the left DLPFC comprising CH06, CH13, CH21 [Brodmann area 46]; the right DLPFC comprising CH02, CH10, CH17 [Brodmann area 46]; the left Broca's area comprising CH07, CH14,

CH22 [Brodmann area of 45]; the left Broca's area comprising CH01, CH09, CH16 [Brodmann area 45]; and the FPC comprising CH11, CH12, CH18, CH19, CH20 [Brodmann area 10]

4 Results

4.1 Subjective Perception

Figure 4 and Table 4 present the means and standard deviations of subjective perceptions. The results of the repeated measures ANOVA indicated that voice gender significantly impacted users' perceived learning effort [$F(1, 28)=4.963$, $p=0.034$, $\eta^2=0.151$] and preference [$F(1, 28)=10.808$, $p=0.003$, $\eta^2=0.278$]. Multiple comparison analysis revealed that female voices elicited higher levels of perceived learning effort ($p<0.05$) and preference ($p<0.01$) compared to male voices. Additionally, communication styles significantly influenced users' perceived trust [$F(1, 28)=11.290$, $p=0.002$, $\eta^2=0.287$], learning effort [$F(1, 28)=26.004$, $p<0.001$, $\eta^2=0.482$], willingness to use [$F(1, 28)=15.871$, $p<0.001$, $\eta^2=0.362$], and preference [$F(1, 28)=22.594$, $p<0.001$, $\eta^2=0.447$]. The multiple comparison tests demonstrated that the relational styles resulted in higher perceived trust ($p<0.01$), learning effort ($p<0.001$), willingness to use ($p<0.001$), and preference ($p<0.001$) than the functional styles (Fig. 4 and Table 4). However, no

significant interaction effect between voice gender and communication styles was observed.

4.2 Eye-Tracking Metric

Figure 5 illustrates participants' visual attention towards educational robots utilizing four different voices. Table 5 presents the means and standard deviations of the fixation duration ratio across various AOIs. The results of the repeated measures ANOVA showed a significant main effect of voice gender [$F(1, 28)=10.437$, $p=0.003$, $\eta^2=0.272$] on fixation duration ratio. However, Mauchly's test indicated that the assumption of sphericity was violated for both the main effect of AOI ($p<0.05$) and the interaction between voice gender and AOI ($p<0.05$). Consequently, a Greenhouse-Geisser correction [62] was applied [63], and the corrected results were reported. A significant main effect of AOI [$F(3, 40.756)=78.702$, $p<0.001$, $\eta^2=0.738$] on fixation duration ratio was observed. Additionally, the results indicated a significant interaction effect between voice gender and AOI on fixation duration ratio [$F(3, 35.398)=6.029$, $p=0.001$, $\eta^2=0.177$]. Simple effect analysis showed that participants' fixation duration ratio on AOI-2 was higher

Fig. 4 The mean and standard deviation of subjective perception. * depicts $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

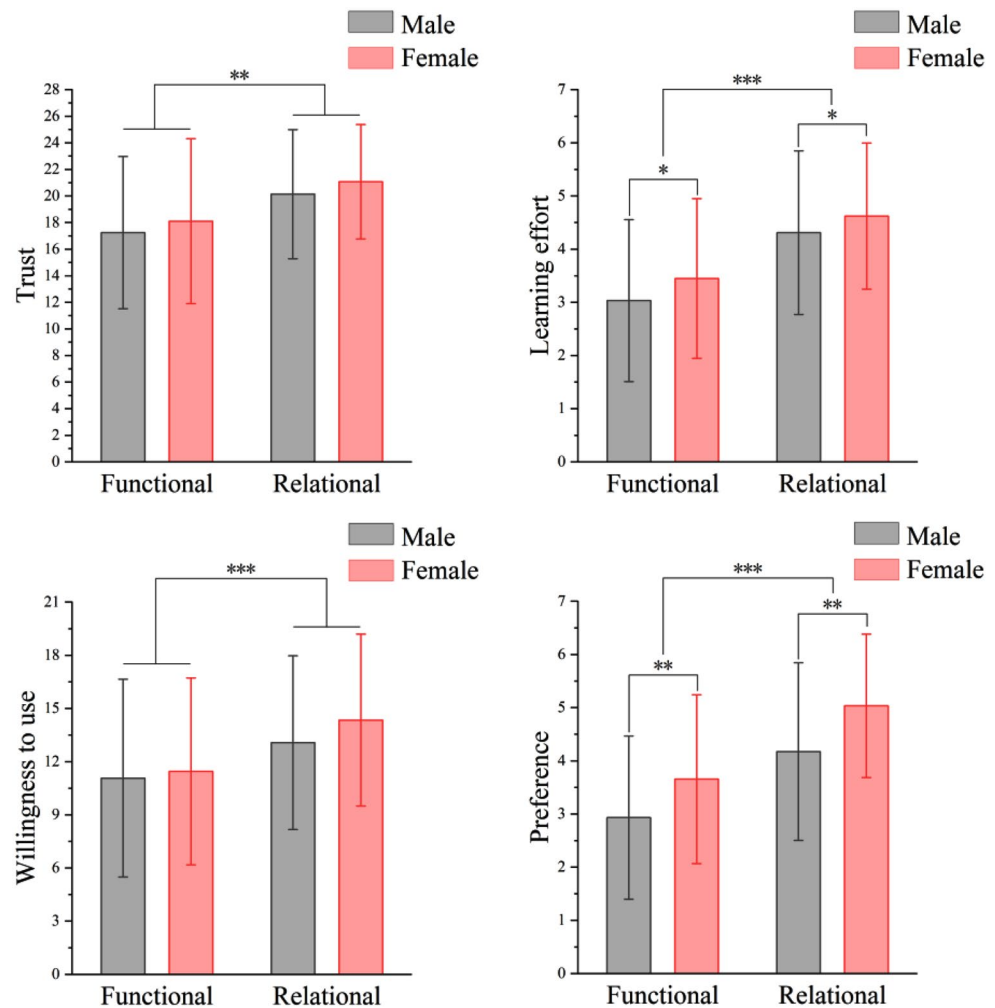


Table 4 Means and standard deviations of subjective perception

Voice gender	Communication styles	Trust		Learning effort		Willingness to use		Preference	
		M	SD	M	SD	M	SD	M	SD
Male	Functional	17.241	5.724	3.034	1.523	11.069	5.580	2.931	1.534
	Relational	20.138	4.853	4.310	1.538	13.069	4.898	4.172	1.671
Female	Functional	18.103	6.201	3.448	1.502	11.448	5.275	3.655	1.587
	Relational	21.069	4.301	4.621	1.374	14.345	4.850	5.034	21.069

when the robot used female voices compared to male voices ($p < 0.01$). For both male and female voices, the fixation duration ratio on AOI-2 exceeded that on AOI-1 ($p < 0.01$), AOI-3 ($p < 0.001$), and AOI-4 ($p < 0.001$). Furthermore, the fixation duration ratio on AOI-1 was greater than that on AOI-3 ($p < 0.001$) and AOI-4 ($p < 0.001$). No other significant main effects and interaction effects were identified.

4.3 fNIRS Data

Table 6 shows the means and standard deviations of β values across different ROIs during the Q&A interaction. Figure 6 presents users' cerebral activity in this context. The results

of the repeated measures ANOVA showed that voice gender significantly affected the β value in the left DLPFC [$F(1, 28) = 6.206$, $p = 0.019$, $\eta^2 = 0.181$]. Additionally, the interaction effect of communication styles and voice gender on the β values in the left DLPFC was significant [$F(1, 28) = 5.903$, $p = 0.022$, $\eta^2 = 0.174$]. Specifically, the simple effect analysis revealed that the functional communication styles induced higher hemodynamic responses in the left DLPFC than the relational communication styles when educational robots used male voices ($p = 0.083$, marginal significance). Furthermore, when educational robots employed functional communication styles, male voices elicited higher hemodynamic responses in the left DLPFC compared to female

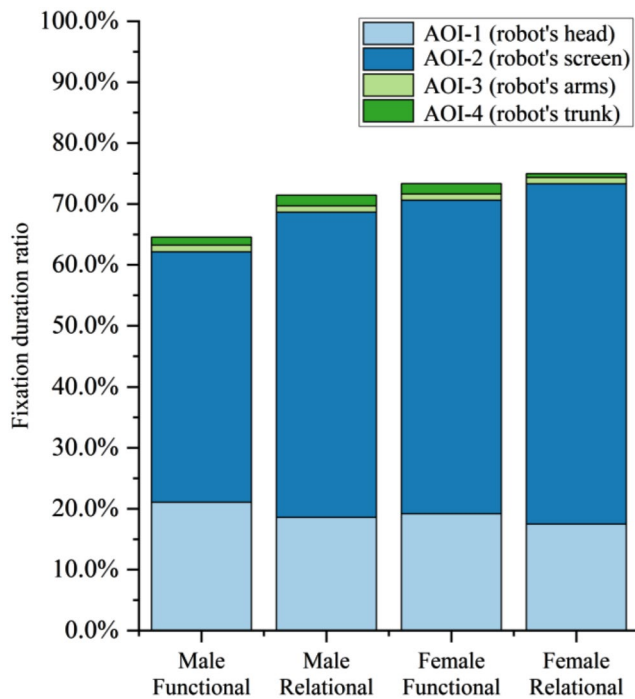


Fig. 5 The participants' visual attention for different voice genders and communication styles

voices ($p < 0.01$). The main effect of communication styles on hemodynamic responses in the FPC was also significant [$F(1, 28) = 4.871$, $p = 0.036$, $\eta^2 = 0.148$], with multiple comparison tests indicating that the relational communication styles induced higher hemodynamic responses in the FPC than the functional communication styles ($p < 0.05$). We did not find other main effects of voice gender and communication styles. No other significant interaction effects of communication style and voice gender were found.

4.4 Correlations Between Subjective and Objective Metrics

The correlation analysis showed that users' perceived trust was significantly positively correlated with their learning effort, willingness to use, and preference (see Table 7). Furthermore, users' learning effort exhibited significant positive correlations with both willingness to use and preference, as well as fixation duration ratio on AOI-3. Users' willingness to use was positively correlated with their preference and the fixation duration ratio on AOI-3. Users' preference was positively correlated with the fixation duration ratio on AOI-3. The fixation duration ratio on AOI-2 showed a significant negative association with the fixation duration ratio on AOI-1 and AOI-4. In addition, the hemodynamic responses in the left DLPFC were significantly positively associated with the responses in the right DLPFC and the left Broca's area. The hemodynamic responses in the right DLPFC positively correlated with those in the left Broca's area, the right Broca's area, and the FPC. Lastly, the hemodynamic responses in the right Broca's area were positively associated with those in the left Broca's area and the FPC. No other significant correlations were observed.

5 Discussion

This study investigated whether and how voice gender and communication style of educational robots impact users' perceptions, visual attention, and cerebral activities, responding to the call for research on educational robots [5]. Overall, the results revealed a significant main effect of voice gender on users' subjective perceptions, visual behaviors, and cerebral activities, suggesting that voice gender may function as salient social signals during HRI. Consistent with prior HRI research [9, 13, 25], users may

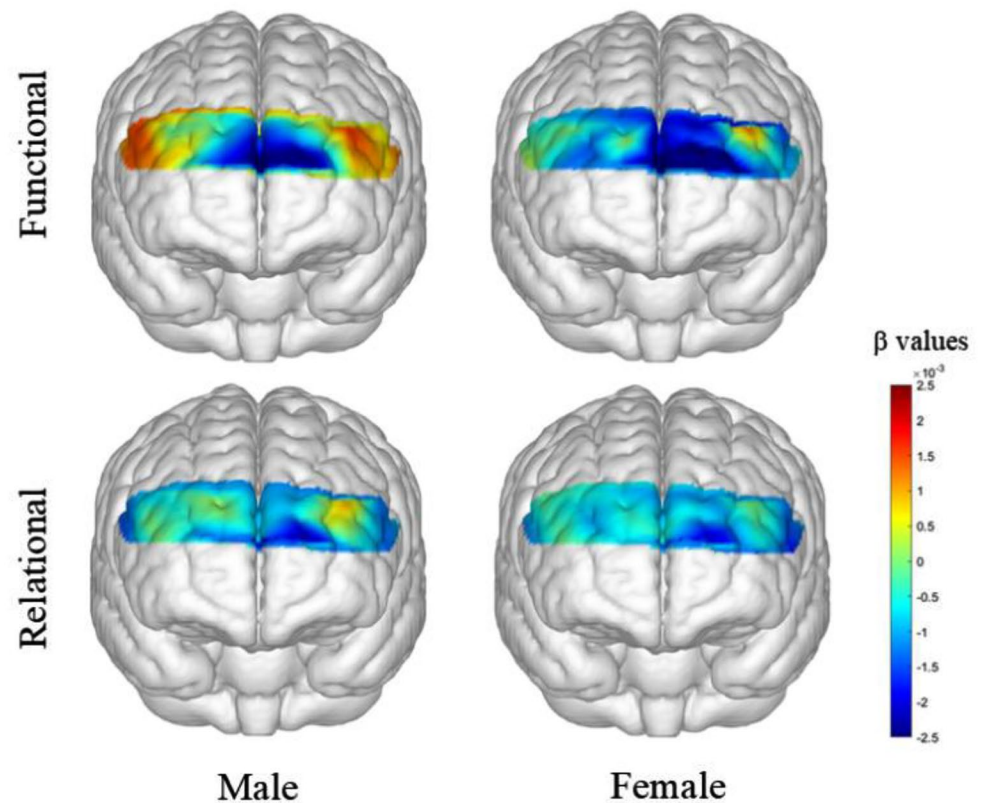
Table 5 Means and standard deviations of fixation duration ratio on different AOIs

Voice gender	Communication styles	AOI-1		AOI-2		AOI-3		AOI-4	
		M	SD	M	SD	M	SD	M	SD
Male	Functional	0.211	0.190	0.411	0.213	0.011	0.029	0.013	0.048
	Relational	0.186	0.172	0.500	0.208	0.011	0.028	0.018	0.088
Female	Functional	0.192	0.179	0.514	0.246	0.010	0.043	0.017	0.080
	Relational	0.175	0.159	0.558	0.233	0.010	0.033	0.006	0.024

Table 6 Means and standard deviations of β values on different brain ROIs

Voice gender	Communication styles	The left DLPFC		The right DLPFC		The left Broca's area		The right Broca's area		The FPC	
		M	SD	M	SD	M	SD	M	SD	M	SD
Male	Functional	0.0015	0.004	0.0005	0.004	0.0008	0.005	0.0015	0.006	-0.0019	0.005
	Relational	-0.0001	0.002	-0.0003	0.003	-0.0008	0.002	-0.0005	0.003	-0.0008	0.002
Female	Functional	-0.0012	0.003	-0.0009	0.004	-0.0010	0.005	-0.0002	0.004	-0.0025	0.003
	Relational	-0.0002	0.003	-0.0006	0.003	-0.0012	0.003	-0.0010	0.003	-0.0010	0.003

Fig. 6 Cerebral activity of users in Q&A interaction



automatically associate different voice genders with varying levels of warmth and competence, thereby influencing both conscious evaluation and implicit cognitive processing. By contrast, communication style only impacted subjective perceptions, which may indicate that communication style primarily shapes higher-level evaluative judgments rather than immediate attentional allocation or neural activation patterns. Furthermore, the absence of interaction effects suggests that voice gender and communication style may independently influence users' responses through distinct cognitive mechanisms. Below, further analysis of the results of users' subjective perception, visual attention, and cerebral activities, as well as the theoretical contributions and practical implications of these results, is addressed. Note that the conclusions of this study are from the perspective of promoting affective engagement, without considering about necessarily learning efficacy. Cognitive learning outcomes are out of this study's scope.

5.1 Effects on users' Subjective Perception

Regarding RQ1, the results on subjective perceptions of affective outcomes showed that the female voices of educational robots, compared to male voices, elicited higher learning effort and greater user preference for the robot. In other words, when learning with robots that used a female voice rather than a male voice, users exerted more effort

in the learning process and showed a stronger preference. While Dou et al. [30] reported that both adult female and male voices are more acceptable than child or synthetic voices in interactions with educational robots, they did not further clarify the difference between male and female synthetic voices. Our findings address this gap by highlighting the significance of the female synthetic voice for educational robots. Furthermore, our results extend the conclusion that users prefer female voices over male voices, a finding previously established in the context of home companion robots [29], to the field of education. Overall, our results align with the broader consensus that a robot's voice gender influences users' affective evaluations [9, 29, 30].

Secondly, the results indicated that educational robots with relational communication styles elicited higher perceived trust, learning effort, willingness to use, and preference compared to those with functional ones. This suggests that users prioritize building a positive relationship with educational robots over the efficiency of information transmission. These findings support the conclusion that relational communication styles are more effective than functional ones in fostering positive user attitudes [14, 35, 36]. Although some studies pointed out that this is not always the case [37, 38], our results can be explained from the following perspectives. First, previous research has shown that users with low competency prefer functional styles, while those with high competency favor relational

Table 7 Pearson's correlation coefficients between subjective and objective metrics

Pearson's correlations		Trust	Learning effort	Willingness to use	Preferen- ce	AOI-1	AOI-2	AOI-3	AOI-4	The Left DLPFC	The right DLPFC	The left Broca's area	The right Broca's area	The FPC
Trust	r	1.000	0.695 ^{***}	0.695 ^{***}	0.713 ^{***}	-0.072	0.148	0.144	-0.031	0.165	-0.037	-0.079	-0.006	0.033
	Sig.		<0.001	<0.001	<0.001	0.445	0.114	0.123	0.742	0.076	0.693	0.400	0.949	0.727
Learning effort	r		1.000	0.813 ^{***}	0.817 ^{***}	-0.041	0.119	0.258 ^{**}	0.012	0.039	-0.073	-0.167	-0.169	0.163
	Sig.			<0.001	<0.001	0.660	0.204	0.005	0.902	0.679	0.438	0.073	0.071	0.080
Willingness to use	r			1.000	0.760 ^{***}	-0.075	0.107	0.218 [*]	-0.024	0.161	-0.124	-0.018	-0.034	0.099
	Sig.				<0.001	0.425	0.254	0.019	0.798	0.084	0.187	0.846	0.717	0.291
Preference	r				1.000	-0.133	0.139	0.199 [*]	0.029	0.011	-0.068	-0.077	-0.133	0.095
	Sig.					0.153	0.138	0.032	0.761	0.910	0.469	0.410	0.153	0.311
AOI-1	r					1.000	-0.628 ^{***}	-0.128	-0.084	0.046	0.046	0.089	0.039	0.036
	Sig.						<0.001	0.171	0.369	0.624	0.624	0.343	0.681	0.703
AOI-2	r						1.000	0.021	-0.323 ^{***}	-0.006	-0.015	-0.130	-0.058	0.021
	Sig.							0.820	<0.001	0.952	0.870	0.165	0.539	0.822
AOI-3	r							1.000	-0.057	-0.010	-0.078	-0.023	-0.014	0.102
	Sig.								0.545	0.913	0.403	0.807	0.879	0.274
AOI-4	r								1.000	-0.033	0.054	0.023	0.039	0.024
	Sig.									0.726	0.563	0.806	0.681	0.795
The Left DLPFC	r									1.000	0.225 [*]	0.237 [*]	0.171	0.151
	Sig.										0.015	0.010	0.066	0.105
The right DLPFC	r										1.000	0.256 ^{**}	0.413 ^{***}	0.649 ^{***}
	Sig.											0.006	<0.001	<0.001
The left Broca's area	r											1.000	0.710 ^{***}	0.179
	Sig.												<0.001	0.055
The right Broca's area	r												1.000	0.351 ^{***}
	Sig.													<0.001
The FPC	r													1.000
	Sig.													

Note: * depicts $0.01 < p < 0.05$, ** depicts $0.001 < p < 0.01$, *** depicts $p < 0.001$. AOI-1 represents the fixation duration ratio on robot's head, AOI-2 represents the fixation duration ratio on robot's screen, AOI-3 represents the fixation duration ratio on robot's arms, AOI-4 represents the fixation duration ratio on robot's trunk

styles when interacting with AI-based assistants [37]. The participants in this study, college students, are generally considered to have high competency. Second, robots with relational communication styles are commonly perceived as having “high warmth and low competence” [37]. The role of teaching assistant of robots may perceptually compensate for the “low competence” impression brought by relational communication styles. In summary, based on subjective results, educational robots with female voices and relational communication styles induce more positive affective outcomes during the learning process.

Finally, our results revealed significant positive correlations among trust, learning effort, willingness to use, and preference. In other words, when users perceive high levels of trust in educational robots, they tend to demonstrate greater learning effort, increased willingness to use the robot, and higher preference.

5.2 Effects on users' Visual Attention

Considering RQ2, the results indicated that the voice gender of educational robots significantly impacts users' visual attention. Compared to interactions with educational robots with male voices, users had a higher fixation duration ratio on the screen (AOI-2) when interacting with robots that used female voices. In other words, users paid more attention to the robot's screen, which displayed the text questions, when the robots used female voices. This visual behavior corroborates the subjective result that users exert more learning effort when studying with educational robots that use female voices compared to those that use male voices. In addition, the results showed that users focused primarily on the robot's screen (AOI-2), followed by its head (AOI-1), while the robot's arms (AOI-3) and trunk (AOI-4) received the least attention. The study of Li et al. [64] on users' visual and neural responses to humanoid robot appearances during free-viewing tasks suggested that the robot's head attracted the most visual attention. Since the test questions were displayed on the robot's screen in this study, it is unsurprising that users concentrated most on the screen to complete the Q&A task. Beyond the screen, the robot's head also captured significant visual attention, expanding the findings of Li et al. [64] to the interaction phase.

Moreover, the results showed a significant negative correlation between the fixation duration ratio on the screen (AOI-2) and that on the robot's head (AOI-1) and the robot's trunk (AOI-4), suggesting that the robot's head (AOI-1) and trunk (AOI-4) may compete for users' visual attention with the screen where displayed the learning materials (test questions). Accordingly, the more visual attention users focused on the robot's head and trunk, the less visual attention they were likely to allocate to the learning tasks. Although voice

gender did not significantly affect fixation duration ratio on the robot's head (AOI-1), the descriptive pattern observed in Table 5 reveals a potentially meaningful tendency. Across both communication styles, users showed slightly higher fixation duration ratio on the robot's head (AOI-1) when interacting with the robot using male voices compared with female voices. Although this difference does not reach statistical significance, it may suggest that male voices draw more visual attention to the robot itself, whereas female voices may help users maintain greater focus on the learning task. This weak but noteworthy tendency aligns with our subjective findings, in which the robot with female voices was associated with higher perceived learning effort. Additionally, although the results showed positive correlations between subjective perceptions (learning effort, willingness to use, and preference) and fixation duration ratio on AOI-3 (robot's arms), it is indeed most likely a statistical artefact, especially since the fixation duration ratio on the robot's arms is very low (see Table 5). There are not sufficient grounds to speculate on the exact cause of this correlation.

5.3 Effects on users' Cerebral Activity

The results showed distinct cerebral activities of users when interacting with robots that employed different voice gender and different communication styles (RQ3). Specifically, educational robots with male voices and functional communication style elicited higher hemodynamic responses in the left DLPFC compared to those with male voices and relational communication style. Additionally, when robots used functional communication styles, male voices induced higher hemodynamic responses in the left DLPFC than female voices. Previous research suggests that positive emotions might activate the left DLPFC [41–43], indicating that male voices may be more compatible with functional communication styles. This provides physiological evidence supporting Liu et al. [8] conclusion that male voices are preferable over female voices in robots that use functional communication styles.

Moreover, the left DLPFC is also involved in off-task thoughts [44]. From this perspective, the higher activation in the left DLPFC when robots use male voices and functional communication styles may suggest that users experienced more off-task thoughts unrelated to learning. This aligns with the subjective findings that male voices combined with functional communication styles lead to lower learning effort.

Our results also showed that educational robots using relational communication styles induced higher activation in the FPC compared to those using functional communication styles. The FPC plays a crucial role in higher-order cognitive operations, including social cognition, reasoning tasks,

and mental state attribution [65]. Specifically, the FPC is associated with the specific aspects of social cognition, such as considering one's judgments in light of another's [45, 46]. In the context of this study, the activation of the FPC may reflect the cognitive process that participants comparing the educational robot's answer to their own during the process of listening to the robot's voice. Users may have a strong cognition of the social identity of the teaching assistant of the educational robot with relational communication styles, and it is easy to bring it into the learning task to judge whether their answers are consistent with the feedback of the robot. Thus, our findings suggest that relational communication styles of educational robots may enhance users' social cognition, consistent with the subjective results.

5.4 Implications

This study has several theoretical implications. Our findings contribute to the literature on voice gender [10, 11] and communication styles [14, 35–38] in HRI. First, we confirm the effect of voice gender and communication style on subjective perceptions, report the impact of voice gender on users' visual attention, and reveal the interaction effects of voice gender and communication styles on users' cerebral activities. Second, we focus on the affective outcomes, including perceived trust, learning effort, willingness to use, and preference, based on existing research in HRI within the education context [5]. Third, by combining subjective psychological measures with objective physiological assessments, this research highlights the significant effects of voice gender and communication styles, contrasting with literature that has primarily examined these factors from a subjective perception [8]. Furthermore, this study enhances our understanding of users' visual attention and cerebral activities when interacting with educational robots.

This study also provides several practical implications. First, the results provide guidelines for designing voice interactions for educational robots, emphasizing the importance of considering voice gender and communication styles. Based on our findings, female voices paired with relational communication styles are recommended to be utilized to foster positive user perceptions. Additionally, this study reveals that the robot's screen captures the most visual attention, followed by the head [64], showing that the robot's social function competes with the information displayed on the screen. Therefore, if an educational robot features an attractive head, it may divert users' visual attention away from the screen. To ensure that users remain focused on the screen, where learning materials are presented, the design of the robot's head should be less visually appealing.

6 Limitations and Future Research

First, this study employed a simulated Q&A experience to isolate the affective effects of the robot's voice gender and communication style without the interference from learning performance. Although the scripted interaction improved experimental control and ensured consistent delivery across conditions, it may not fully capture the dynamic and spontaneous nature of real-world educational interactions with robots. Therefore, caution is warranted when generalizing the findings to natural educational environments. Additionally, cognitive learning outcomes fall outside the scope of the present study, even though they represent another crucial dimension of HRI outcomes in educational robotics [5]. Although positive affective outcomes do not necessarily correspond to improved cognitive outcomes, affective outcomes remain an important aspect of robot-assisted learning [5, 66, 67]. Whether the observed affective advantages of the female voice and relational communication style in this study translate into improved cognitive outcomes needs further investigation in future work.

Second, the study relied exclusively on Chinese participants to assess subjective responses to the robot's voiced gender and communication style. Cultural norms and expectations can significantly influence perceptions of gendered voices and communication patterns, and as such, our findings may not generalize to participants from other cultural backgrounds. Future research should include more culturally diverse samples to examine the potential variability in these perceptions across different societies.

Third, the participant group consisted solely of college students, who generally possess a high level of education and technology acceptance. As a result, the generalizability of these findings to the broader population is limited. In studies of educational robots involving adult groups, participants are primarily aged between 20 and 30 years [5], which is similar to the demographic of our study. However, preferences for robot voice gender and communication style may vary across different demographic groups. Results obtained from adults cannot be generalized to other populations, such as children, adolescents, or older adults, underscoring the need for further research involving wider populations.

Finally, this study concentrated exclusively on the voice interactions of educational robots, specifically examining the interaction effects of voice gender and communication styles. Future studies might explore the impact of other voice characteristics and additional modes of communication (e.g., gaze behavior, gestures) to encourage designers to create educational robots that yield positive outcomes.

Appendix

Table A1 MNI coordinates of fNIRS channels and corresponding brain regions

AOI	Channel	Brodmann area (BA)	MNI coordinates			Overlap
			X	Y	Z	
Left DLPFC	CH06	46-Dorsolateral prefrontal cortex	-37.000	48.667	33.000	0.794
	CH13	46-Dorsolateral prefrontal cortex	-30.333	60.333	24.333	0.735
	CH21	46-Dorsolateral prefrontal cortex	-40.667	57.333	13.667	0.733
Right DLPFC	CH02	46-Dorsolateral prefrontal cortex	37.667	48.667	34.667	0.694
	CH10	46-Dorsolateral prefrontal cortex	32.333	60.333	25.333	0.678
	CH17	46-Dorsolateral prefrontal cortex	42.667	58.000	15.667	0.725
Left Broca's area	CH07	45-pars triangularis Broca's area	-50.667	28.667	34.333	0.567
	CH14	45-pars triangularis Broca's area	-47.000	44.667	23.000	0.766
	CH22	45-pars triangularis Broca's area	-54.667	35.333	11.667	1.000
Right Broca's area	CH01	45-pars triangularis Broca's area	51.667	29.667	37.333	0.458
	CH09	45-pars triangularis Broca's area	48.667	44.667	25.667	0.768
	CH16	45-pars triangularis Broca's area	57.000	35.667	16.000	1.000
Frontopolar area	CH11	10 - Frontopolar area	13.667	67.000	28.333	0.923
	CH12	10 - Frontopolar area	-11.000	65.000	29.667	0.810
	CH18	10 - Frontopolar area	24.000	69.333	18.333	0.982
	CH19	10 - Frontopolar area	3.333	68.000	19.667	1.000
	CH20	10 - Frontopolar area	-21.667	69.000	17.667	0.993

Acknowledgements The authors thank all the participants for their contribution to this study and the editor and the reviewers for their valuable comments.

Author Contribution Chen Fang: Conceptualization, Methodology, Investigation, Formal analysis, Writing - original draft, Writing - review & revising. Fu Guo: Conceptualization, Methodology, Funding acquisition, Supervision. Xuegong Bao: Conceptualization, Methodology, Investigation, Formal analysis. Zeyu Zhang: Conceptualization, Methodology. Tony Belpaeme: Writing - review & revising.

Funding This work was supported by the National Natural Science Foundation of China under Grant 72071035. The funder had no role in study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the article for publication. The work received partial funding from the Flanders AI Research Program and the China Scholarship Council.

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethical approval All procedures performed in studies involving participants were following the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all participants in the study.

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Rosanda V, Bratko I, Gačnik M, Podpečan V, Istenič A (2025) Robot NAO integrated lesson vs. traditional lesson: measuring learning outcomes on the topic of “societal change” and the mediating effect of students’ attitudes. *Brit J Educ Tech* 56(1):435–450. <https://doi.org/10.1111/bjet.13501>
- Lemaignan S, Newbutt N, Rice L, Daly J (2024) “It’s important to think of pepper as a teaching aid or resource external to the classroom”: a social robot in a school for autistic children. *Int J Soc Robot* 16(6):1083–1104. <https://doi.org/10.1007/s12369-022-00928-4>
- Guggemos J, Seufert S, Sonderegger S (2020) Humanoid robots in higher education: evaluating the acceptance of pepper in the context of an academic writing course using the UTAUT. *Brit J Educ Tech* 51(5):1864–1883. <https://doi.org/10.1111/bjet.13006>
- Jia F, Sun D, Looi C-K (2024) Artificial intelligence in science Education (2013–2023): research trends in Ten years. *J Sci Educ Technol* 33(1):94–117. <https://doi.org/10.1007/s10956-023-1007-6>
- Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F (2018) Social robots for education: a review. *Sci Robot* 3(21):eaat5954. <https://doi.org/10.1126/scirobotics.aat5954>
- Gordon G, Spaulding S, Westlund JK, Lee JJ, Plummer L, Martinez M et al. Affective personalization of a social robot tutor for children’s second language skills. In: *Proceedings of the AAAI conference on artificial intelligence 2016*. <https://doi.org/10.1609/aaai.v30i1.9914>
- Song S, Baba J, Nakanishi J, Yoshikawa Y, Ishiguro H Mind the voice!: effect of robot voice pitch, robot voice gender, and user gender on user perception of teleoperated robots. In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp 1–8. <https://doi.org/10.1145/3334480.3382988>
- Liu X-X, Yin C-Y, Li M-R (2024) The power of voice! The impact of robot receptionists’ voice pitch and communication style on customer value cocreation intention. *Int J Hosp Manage* 122:103819. <https://doi.org/10.1016/j.ijhm.2024.103819>

9. Niculescu A, Van Dijk B, Nijholt A, Li H, See SL (2013) Making social robots more attractive: the effects of voice pitch, humor and empathy. *Int J Soc Robot* 5(2):171–191. <https://doi.org/10.1007/s12369-012-0171-x>
10. McGinn C, Torre I (2019) Can you tell the robot by the voice? An exploratory study on the role of voice in the perception of robots. In 2019 14th ACM/IEEE international Conference on human-robot interaction, IEEE, HRI, 211–221
11. Eyssel F, Kuchenbrandt D, Bobinger S, De Ruiter L, Hegel F ‘If you sound like me, you must be more human’ on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In: Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction 2012, pp 125–126. <https://doi.org/10.1145/2157689.2157717>
12. Lee EJ, Nass C, Brave S Can computer-generated speech have gender? An experimental test of gender stereotype. CHI’00 extended abstracts on human factors in computing systems 2000. pp 289–290. <https://doi.org/10.1145/633292.633461>
13. Tay B, Jung Y, Park T (2014) When stereotypes meet robots: the double-edge sword of robot gender and personality in human-robot interaction. *Comput Hum Behav* 38:75–84. <https://doi.org/10.1016/j.chb.2014.05.014>
14. Kim J, Merrill JK, Xu K, Sellnow DD (2021) I like my relational machine teacher: an AI instructor’s communication styles and social presence in online education. *Int J Hum-Comput Interact* 37(18):1760–1770. <https://doi.org/10.1080/10447318.2021.1908671>
15. Reich-Stiebert N, Eyssel F (2015) Learning with educational companion robots? Toward attitudes on education robots, predictors of attitudes, and application potentials for education robots. *Int J Soc Robot* 7(5):875–888. <https://doi.org/10.1007/s12369-015-0308-9>
16. Qin F, Li K, Yan J (2020) Understanding user trust in artificial intelligence-based educational systems: evidence from China. *Brit J Educ Tech* 51(5):1693–1710. <https://doi.org/10.1111/bjet.12994>
17. Chen Z-H, Chou C-Y, Deng Y-C, Chan T-W (2007) Active open learner models as animal companions: motivating children to learn through interacting with my-pet and our-pet. *Int J Artif Intel Educ* 17(2):145–167. [https://doi.org/10.3233/IRG-2007-17\(2\)04](https://doi.org/10.3233/IRG-2007-17(2)04)
18. Li M, Guo F, Fang C, Li F (2023) Multisensory integration effect of humanoid robot appearance and voice on users’ affective preference and visual attention. *Behaviour Inf Technol* 42(14):2387–2406. <https://doi.org/10.1080/0144929X.2022.2125830>
19. de Haas M, Vogt P, Krahmer E (2021) When preschoolers interact with an educational robot, does robot feedback influence engagement? *Multimodal Technol Interact* 5(12):77. <https://doi.org/10.3390/mti5120077>
20. Wu J, Jiao D, Cai Y, Guo S, Chen Y, Li C, Yang H (2024) Personalized social reading: the effect of text markers and font size on the reading experience. *Int J Ind Ergon* 103:103626. <https://doi.org/10.1016/j.ergon.2024.103626>
21. Ren Z, Guo F, Li M, Hu M, Duffy VG (2023) The effect of Robot’s facial features on users’ perception: evidence from subjective, pupillometry, and electroencephalography measures. *Int J Hum-Comput Interact* 40(22):7375–7388. <https://doi.org/10.1080/10447318.2023.2265733>
22. Guo F, Fang C, Li M, Ren Z, Zhang Z (2024) How do robot touch characteristics impact users’ emotional responses: evidence from ECG and fNIRS. *Int J Soc Robot* 16(3):619–634. <https://doi.org/10.1007/s12369-024-01110-8>
23. Nass CI, Brave S (2005) *Wired for speech: how voice activates and advances the human-computer relationship*. MIT Press Cambridge
24. De Cet M, Obaid M, Torre I Breaking the binary: a systematic review of gender-ambiguous voices in human-computer interaction. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, pp 1–17. <https://doi.org/10.1145/3706598.3713608>
25. Macrae CN, Milne AB, Bodenhausen GV (1994) Stereotypes as energy-saving devices: a peek inside the cognitive toolbox. *J Pers Soc Psychol* 66(1):37. <https://doi.org/10.1037/0022-3514.66.1.37>
26. Siegel M, Breazeal C, Norton MI (2009) Persuasive robotics: the influence of robot gender on human behavior. In 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2563–2568
27. Traunmüller H, Eriksson A (1995) The frequency range of the voice fundamental in the speech of male and female adults. Unpublished Manuscr 11
28. Riding D, Lonsdale D, Brown B (2006) The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women. *J Nonverbal Behav* 30(2):55–61. <https://doi.org/10.1007/s10919-006-0005-3>
29. Chang RC-S, Lu H-P, Yang P (2018) Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in Taiwan. *Comput Hum Behav* 84:194–210. <https://doi.org/10.1016/j.chb.2018.02.025>
30. Dou X, Wu C-F, Lin K-C, Gan S, Tseng T-M (2021) Effects of different types of social robot voices on affective evaluations in different application fields. *Int J Soc Robot* 13(4):615–628. <https://doi.org/10.1007/s12369-020-00654-9>
31. Levy J, Wubbels T, Brekelmans M, Morganfield B (1997) Language and cultural factors in students’ perceptions of teacher communication style. *Int J Intercult Relat* 21(1):29–56. [https://doi.org/10.1016/S0147-1767\(96\)00005-3](https://doi.org/10.1016/S0147-1767(96)00005-3)
32. Hosek A, Houser M (2018) Introduction to instructional communication methodological approaches. *Handb Instructional Commun: Rhetorical Relational Perspectives* 207–208
33. Richmond VP, McCroskey JC, Mottet T (2015) *Handbook of instructional communication: rhetorical and relational perspectives*. Routledge
34. Frisby BN, Buckner MM (2017) Rapport in the instructional context. *Handbook of instructional communication*. Routledge, pp 126–137
35. Xu Y, Zhang J, Chi R, Deng G (2022) Enhancing customer satisfaction with chatbots: the influence of anthropomorphic communication styles and anthropomorphised roles. *Nankai Bus Rev Int* 14(2):249–271. <https://doi.org/10.1108/NBRI-06-2021-0041>
36. van Pinxteren MME, Pluymaekers M, Lemmink J, Krispin A (2023) Effects of communication style on relational outcomes in interactions between customers and embodied conversational agents. *Psychol Mark* 40(5):938–953. <https://doi.org/10.1002/mar.21792>
37. Xu Y, Niu N, Zhao Z (2023) Dissecting the mixed effects of human-customer service chatbot interaction on customer satisfaction: an explanation from temporal and conversational cues. *J Retail Consum Serv* 74:103417. <https://doi.org/10.1016/j.jretconser.2023.103417>
38. Chattaraman V, Kwon W-S, Gilbert JE, Ross K (2019) Should AI-Based, conversational digital assistants employ social-or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Comput Hum Behav* 90:315–330. <https://doi.org/10.1016/j.chb.2018.08.048>
39. Li M, Guo F, Li Z, Ma H, Duffy VG (2024) Interactive effects of users’ openness and robot reliability on trust: evidence from psychological intentions, task performance, visual behaviours, and cerebral activations. *Ergonomics* 67(11):1612–1632. <https://doi.org/10.1080/00140139.2024.2343954>
40. Li M, Guo F, Wang X, Chen J, Ham J (2023) Effects of robot gaze and voice human-likeness on users’ subjective perception, visual attention, and cerebral activity in voice conversations. *Comput*

- Hum Behav 141:107645. <https://doi.org/10.1016/j.chb.2022.107645>
41. Ahern GL, Schwartz GE (1985) Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis. *Neuropsychologia* 23(6):745–755. [https://doi.org/10.1016/0028-3932\(85\)90081-8](https://doi.org/10.1016/0028-3932(85)90081-8)
 42. Herrington JD, Mohanty A, Koven NS, Fisher JE, Stewart JL, Banich MT, Webb AG, Miller GA, Heller W (2005) Emotion-modulated performance and activity in left dorsolateral prefrontal cortex. *Emotion* 5(2):200. <https://doi.org/10.1037/1528-3542.5.2.200>
 43. Etkin A, Egner T, Kalisch R (2011) Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn Sci* 15(2):85–93. <https://doi.org/10.1016/j.tics.2010.11.004>
 44. Turnbull A, Wang H, Murphy C, Ho N, Wang X, Sormaz M, Karapanagiotidis T, Leech RM, Bernhardt B, Margulies DS, Vatansever D, Jefferies E, Smallwood J (2019) Left dorsolateral prefrontal cortex supports context-dependent prioritisation of off-task thought. *Nat Commun* 10(1):3816. <https://doi.org/10.1038/s41467-019-11764-y>
 45. David N, Bewernick BH, Cohen MX, Newen A, Lux S, Fink GR, Shah NJ, Vogeley K (2006) Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *J Cognit Neurosci* 18(6):898–910. <https://doi.org/10.1162/jocn.2006.18.6.898>
 46. Mitchell JP, Ames DL, Jenkins AC, Banaji MR (2009) Neural correlates of stereotype application. *J Cognit Neurosci* 21(3):594–604. <https://doi.org/10.1162/jocn.2009.21033>
 47. Lei X, Rau P-LP (2023) Emotional responses to performance feedback in an educational game during cooperation and competition with a robot: evidence from fNIRS. *Comput Hum Behav* 138:107496. <https://doi.org/10.1016/j.chb.2022.107496>
 48. Paul F, Erdfelder E, Lang A-G, Buchner BAG (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39(2):175–191. <https://doi.org/10.3758/BF03193146>
 49. Bartlett ME, Edmunds C, Belpaeme T, Thill S (2022) Have I got the power? Analysing and reporting statistical power in HRI. *J Hum-Robot Interact* 11(2):1–16. <https://doi.org/10.1145/3495246>
 50. Tanaka F, Isshiki K, Takahashi F, Uekusa M, Sei R, Hayashi K (2015) Pepper learns together with children: development of an educational application. In 2015 IEEE-RAS 15th International Conference on Humanoid Robots, IEEE, Humanoids, 270–275). <https://doi.org/10.1109/HUMANOIDS.2015.7363546>
 51. You Z-J, Shen C, Chang C-W, Liu B-J, Chen G-D (2006) A robot as a teaching assistant in an English class. In Sixth IEEE international conference on advanced learning technologies (ICALT'06), IEEE, Kerkraide, 87–91). <https://doi.org/10.1109/ICALT.2006.1652373>
 52. Abubshait A, McDonald CG, Wiese E (2021) Examining social cognition with embodied robots: does prior experience with a robot impact feedback-associated learning in a gambling task? *J Cognition* 4(1):28. <https://doi.org/10.5334/joc.167>
 53. Gunes H, Churamani N (2023) Affective computing for human-robot interaction research: four critical lessons for the hitchhiker. In 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Busan: IEEE, Korea, Republic of, 1565–1572). <https://doi.org/10.1109/RO-MAN57019.2023.10309450>
 54. Zhu Y, Rodriguez-Paras C, Rhee J, Mehta RK (2020) Methodological approaches and recommendations for functional near-infrared spectroscopy applications in HF/E research. *Hum Factors* 62(4):613–642. <https://doi.org/10.1177/0018720819845275>
 55. Song Y, Luximon A, Luximon Y (2021) The effect of facial features on facial anthropomorphic trustworthiness in social robots. *Appl Ergon* 94:103420. <https://doi.org/10.1016/j.apergo.2021.103420>
 56. Choi JK, Ji YG (2015) Investigating the importance of trust on adopting an autonomous vehicle. *Int J Hum-Comput Interact* 31(10):692–702. <https://doi.org/10.1080/10447318.2015.1070549>
 57. Singh AK, Okamoto M, Dan H, Jurcak V, Dan I (2005) Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. *Neuroimage* 27(4):842–851. <https://doi.org/10.1016/j.neuroimage.2005.05.019>
 58. Hou X, Zhang Z, Zhao C, Duan L, Gong Y, Li Z, Zhu C (2021) NIRS-KIT: a MATLAB toolbox for both resting-state and task fNIRS data analysis. *Neurophoton*. 8(1):010802–. <https://doi.org/10.1117/1.nph.8.1.010802>
 59. Strangman G, Culver JP, Thompson JH, Boas DA (2002) A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *Neuroimage* 17(2):719–731. <https://doi.org/10.1006/nimg.2002.1227>
 60. Cope M, Delpy DT (1988) System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Med Biol Eng Comput* 26(3):289–294. <https://doi.org/10.1007/BF02447083>
 61. Cui X, Bray S, Reiss AL (2010) Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage* 49(4):3039–3046. <https://doi.org/10.1016/j.neuroimage.2009.11.050>
 62. Greenhouse SW, Geisser S (1959) On methods in the analysis of profile data. *Psychometrika* 24(2):95–112. <https://doi.org/10.1007/bf02289823>
 63. Girden ER (1992) ANOVA: repeated measures, vol 84. sage
 64. Li M, Guo F, Ren Z, Duffy VG (2022) A visual and neural evaluation of the affective impression on humanoid robot appearances in free viewing. *Int J Ind Ergon* 88:103159. <https://doi.org/10.1016/j.ergon.2021.103159>
 65. Raposo A, Vicens L, Clithero JA, Dobbins IG, Huettel SA (2011) Contributions of frontopolar cortex to judgments about self, others and relations. *Soc Cogn Affect Neurosci* 6(3):260–269. <https://doi.org/10.1093/scan/nsq033>
 66. Alam A (2022) Social robots in education for long-term human-robot interaction: socially supportive behaviour of robotic tutor for creating robo-tangible learning environment in a guided discovery learning interaction. *ECS Trans* 107(1):12389. <https://doi.org/10.1149/10701.12389ecst>
 67. Chen H, Park HW, Zhang X, Breazeal C Impact of interaction context on the student affect-learning relationship in child-robot interaction. In: Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction, pp 389–397. <https://doi.org/10.1145/3319502.3374822>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Chen Fang is currently a joint Ph.D. researcher at Northeastern University, China and Ghent University, Belgium. Her research focuses on human-robot interaction, human-AI interaction, and user experience, with particular interests in the cognitive and affective mechanisms underlying interaction with intelligent agents

Fu Guo is a professor of Industrial Engineering at School of Business Administration, Northeastern University, China. Her research focuses on human factors, Kansei engineering, user experience, and human–computer interaction.

Xuegong Bao received his M.S. degree in Industrial Engineering from Northeastern University, China, in 2024. He is currently with BYD. His research focuses on human–robot interaction.

Zeyu Zhang is currently a Ph.D. researcher at Northeastern University, China. His research focuses on human-robot interaction, human-agent interaction, and affective experiences.

Tony Belpaeme is currently a Professor at Ghent University, Belgium, and a researcher at IDLab–imec. His research focuses on cognitive robotics, human–robot interaction, social robotics, and artificial intelligence, with a particular focus on understanding and developing socially intelligent robotic systems.