Computers in Human Behavior 141 (2023) 107645

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Effects of robot gaze and voice human-likeness on users' subjective perception, visual attention, and cerebral activity in voice conversations

Mingming Li [a], Fu Guo [a,*], Xueshuang Wang [b], Jiahao Chen [a], Jaap Ham [c]

[a] Department of Industrial Engineering, School of Business Administration, Northeastern University, Shenyang, China
[b] School of Mechanical Engineering, Shenyang University of Technology, Shenyang, China
[c] Research Group of Human Technology Interaction, Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, the Netherlands

ARTICLE INFO

Handling editor: Bjorn de Koning

Keywords:
Robot gaze
Voice human-likeness
Subjective perception
Eye-tracking
fNIRS

ABSTRACT

Robot gaze and voice are essential anthropomorphic features to promote users' engagement in voice conversations. Earlier research chiefly examined how robot gaze and voice human-likeness separately influenced users' subjective perception. When implementing gaze on robots with different human-like voices, there has little evidence of their possible interaction effects, particularly on users' visual attention and cerebral activity, which could help to understand the perceptual and cognitive processing of anthropomorphic features. Therefore, a within-subject experiment of voice conversations with diverse robot gaze (gaze versus no gaze) and human-like voices (high human-like versus low human-like) using subjective reporting, eye-tracker, and fNIRS was conducted. The results showed that the robot with gaze or a high human-like voice evoked more pleasure, higher arousal, more perceived likability, and less negative attitudes. Robot gaze significantly increased users' average fixation durations and total fixation time, while voice human-likeness prolonged first fixation durations. Moreover, the robot with a high human-like voice (or gaze) induced increased activity in the left DLPFC and decreased activity in the right Broca's area than that had no gaze (or a low human-like voice). The results suggest that robot gaze might chiefly capture users' sustained attention, voice human-likeness might attract users' initial attention, and they might jointly influence users' perceptual processing of prosodic features and emotional processing.

1. Introduction

Voice conversation, one of the natural ways humans communicate with each other, has been broadly employed in human-robot interaction (HRI). In voice conversations with users, robots typically use multiple human-like manners, such as voice, gaze, gestures, and emotional expressions (Duffy, 2003). These anthropomorphic features are expected to convey social cues to users to drive unconscious utilization of social responses and behaviors in HRI (Tay, Jung, & Park, 2014; Vollmer, Read, Trippas, & Belpaeme, 2018) and promote users' attitudes and acceptance (Chang, Lu, & Yang, 2018). Numerous researchers and practitioners have designed various anthropomorphic features for robots (Liew & Tan, 2021; Mavridis, 2015; Roesler, Manzey, & Onnasch, 2021; Seaborn, Miyake, Pennefather, & Otake-Matsuura, 2021), including unimodal (Kuhne, Fischer, & Zhou, 2020; Li, Guo, Ren, & Duffy, 2022; Wei & Zhao, 2016) and multimodal (Tsiourti, Weiss, Wac, & Vincze, 2019; Xu, 2019).

However, a meta-analysis (Roesler et al., 2021) indicated that more explorations are needed to assess the effectiveness of these anthropomorphic features on robots, particularly the feature combinations that have been a topic of great interest in recent years (Babel et al., 2021; Ham, Cuijpers, & Cabibihan, 2015; Tatarian et al., 2021; Thepsoonthorn, Ogawa, & Miyake, 2021; Tsiourti et al., 2019). Moreover, users' perception of and response to these anthropomorphic features involve implicit and explicit expressions, including subjective perception, behavioral response, and physiological activity (Rapp, Curti, & Boldi, 2021; Roesler et al., 2021; Tiberio, Cesta, & Olivetti Belardinelli, 2013). Thus, multiple measures from different modalities might be promising to provide a comprehensive understanding of the effect of robot anthropomorphic features on users in HRI.

* Corresponding author. NO. 195, Chuangxin Road, Hunnan District, Shenyang, 110167, China.
E-mail address: fguo@mail.neu.edu.cn (F. Guo).

https://doi.org/10.1016/j.chb.2022.107645
Received 20 July 2022; Received in revised form 6 December 2022; Accepted 29 December 2022
Available online 30 December 2022

1.1. Robot gaze

As a fundamental signal in human face-to-face communication, eye gaze could regulate social intimacy (Kleinke, 1986), manage turn-taking (Degutyte & Astell, 2021), and convey individuals' mental states, intentions, and willingness to interact (Emery, 2000). In the field of robotics, gaze has received abundant attention from researchers and has been regarded as a crucial factor for facilitating interactions, including understanding interaction intentions (Baillon, Selim, & van Dolder, 2013), promoting the engagement of conversations (Mavridis, 2015), modulating decision-making processes (Belkaid, Kompatsiari, Tommaso, Zablith, & Wykowska, 2021), and so forth (Admoni & Scassellati, 2017). According to its functions in HRI, robot gaze could be categorized into mutual gaze, referential (deictic) gaze, joint attention, and gaze aversions (Admoni & Scassellati, 2017). Mutual gaze and gaze aversions are two basic types that usually occur in conversations between humans. Listeners often fixate on speakers' eyes or faces to signal their interest and involvement in speech contents, and the speakers will glance at the listeners to get feedback about their speech (Argyle, 1972). The gaze aversions could happen in an embarrassing situation or cognitive effort (Andrist, Tan, Gleicher, & Mutlu, 2014). These two gaze types are investigated in human-robot conversation widely (Admoni & Scassellati, 2017; Babel et al., 2021; Winkle, Caleb-Solly, Turton, & Bremner, 2019; R. Zhang, Saran, et al., 2020). Nevertheless, earlier studies observed less gaze aversion in human-robot conversation and posited that people might avert eye contact with robots less because they may feel less embarrassed to contact eyes with a mindless robot compared to humans (Bartneck, Bleeker, Bun, Fens, & Riet, 2010; Wood et al., 2013). Thereby, this study focuses on mutual gaze primarily, rather than gaze aversion.

Robot gaze could leave users with a feeling of "being looked at" in conversations and improve users' subjective perceptions such as engagement (Kompatsiari, Ciardo, Tikhanoff, Metta, & Wykowska, 2021; Yang et al., 2018), conversational fluidity (Mavridis, 2015), likability (Mutlu, Kanda, Forlizzi, Hodgins, & Ishiguro, 2012), trust (Morillo-Mendez, Schrooten, Loutfi, & Mozos, 2021), and acceptance (Babel et al., 2021). It is assumed that robot gaze in conversations enables users to perceive more social attributes, enhancing the perceived human-like nature of the robot and making it more acceptable to the user (Babel et al., 2021).

Beyond subjective evaluations, a large corpus of research provided evidence that robot gaze might influence users' behaviors, particularly visual attention (Kompatsiari et al., 2017, 2019; Zhang, Beskow, & Kjellström, 2017). Visual attention refers to the cognitive operations of the selective focus via central fovea areas, visual information retrieval, encoding, and processing (Corbetta & Shulman, 2002). It could reveal what users pay attention to (Husic-Mehmedovic, Omeragic, Batagelj, & Kolar, 2017). Recently, Thepsoonthorn et al. (2021) examined the mutual gaze of a NAO robot in the task of listening to robot speech and indicated that robot mutual gaze, including head nodding, would attract more visual attention from users and result in a longer total fixation time on the robot. Perugia, Paetzel-Prüsmann, Alanenpää, and Castellano (2021) compared users' total fixation time toward human-like, mechanical, or morph robot faces with mutual gaze and found that users'total fixation time could predict engagement in voice conversations. Users' visual attention toward the human gaze versus the robot gaze with different parameters also has been explored by watching videos (Ghiglino et al., 2020, 2021).

Notwithstanding, these studies primarily rely on single eye-tracking metrics. Different eye-tracking metrics might be associated with distinct facets of visual attention. The average fixation duration is related to the visual attention invested in each fixation and is considered a reflection of the processing depth of visual information (Just & Carpenter, 1976; Ramos Gameiro, Kaspar, König, Nordholt, & König, 2017; Unema, Pannasch, Joos, & Velichkovsky, 2005). The fixation count and total fixation time, measuring the accumulative number or time of each fixation, are associated with the total amount of visual attention invested in a specific area (Behe, Bae, Huddleston, & Sage, 2015; Wang, Antonenko, & Dawson, 2020). Furthermore, there is a notable lack of studies on users' visual attention toward robots with gaze in voice conversations, which is significant to understand the effect of robot gaze in interactions. Therefore, multiple eye-tracking metrics should be combined to probe users' visual attention in voice conversations comprehensively.

Earlier studies on the effect of robot gaze reached limited consensuses on users' cerebral activity (Belkaid et al., 2021; Kelley, Noah, Zhang, Scassellati, & Hirsch, 2021; Kompatsiari, Bossi, & Wykowska, 2021). Research on social interaction between human beings manifested that seeing gaze from other people could activate brain oscillations (Hietanen, Leppänen, Peltola, Linna-aho, & Ruuhiala, 2008; Leong et al., 2017). This effect is expected to work on robots. Thereby, scholars attempted to explore the impact of robot gaze on users' cerebral activity. The mutual gaze of an iCub robot in gaze cueing tasks was found to evoke a higher desynchronization of electroencephalographic (EEG) alpha rhythm, which signifies more engagement in interaction (Kompatsiari, Bossi, & Wykowska, 2021) and delays decision-making processes (Belkaid et al., 2021). However, Kelley et al. (2021) compared participants' brain activity when making eye contact with human and humanoid robot partners. The results showed that eye contact with humans elicited enhanced cerebral activity in the right temporoparietal joint area (rTPJ) and dorsolateral prefrontal cortex (DLPFC). In contrast, robots did not elicit enhanced cerebral activity in the rTPJ. The inconsistency might be due to the lower sense of social interaction in the static eye contact task with a 3D-printed humanoid robot bust (Kelley et al., 2021). Thus, the effect of robot gaze on users' cerebral activity needs more neuroimaging evidence in social interaction tasks. And there remains an open question of whether robot gaze impacts users' cerebral activity in voice conversations, which extensively occur in the social interactions between humans and robots.

Additionally, these earlier studies chiefly measured users' subjective perception (Babel et al., 2021; Morillo-Mendez et al., 2021; Mutlu et al., 2012) or combined with either visual attention (Ghiglino et al., 2020, 2021; Thepsoonthorn et al., 2021) or cerebral activity (Belkaid et al., 2021; Kompatsiari, Bossi, & Wykowska, 2021). Kelley et al. (2021) measured eye movements with brain signals and reported the proportion of time spent looking at robot eyes and cerebral activity. Following this perspective, more eye-tracking metrics and subjective perceptions should be included to provide a comprehensive understanding of the effect of robot gaze on users.

1.2. Voice human-likeness

As an expressive communication medium, voices contain rich implicit information about speakers, including gender, identity, emotional state, and personality (Belin, 2006; Nass & Brave, 2005). The implicit information could subconsciously facilitate using social patterns toward robots (Nass, Moon, Fogg, Reeves, & Dryer, 1995) and promote users' experiences, attitudes, and acceptance (Eyssel, Kuchenbrandt, Bobinger, De Ruiter, & Hegel, 2012; Walters, Syrdal, Koay, Dautenhahn, & Te Boekhorst, 2008). Voice human-likeness, referring to the level of naturalness and likeness to human voice, is one of the fundamental anthropomorphic features of robots. It could modulate users' social perception of robots and plays an essential role in the facilitation and promotion effects of voice. Along with the advanced development of Text-To-Speech (TTS) technologies and the achievement of intelligibility, synthesized voices are utilized extensively in social robots. Nevertheless, they might still differ from human voices in subtle expressions, for example, emotional rhythm (Baird et al., 2018). Voice human-likeness remains critical in the anthropomorphic design of robots and has been attracting enormous attention (Kuhne et al., 2020; Seaborn et al., 2021).

Previous studies chiefly focused on comparing users' subjective

perception of robots using human and synthesized voices and observed an overall preference for human voices. Specifically, Eyssel et al. (2012) compared a human-like voice and a mechanical voice on a robot and found that the human-like voice was preferred by most users. Xu (2019) compared a human voice and a synthesized voice on the Alpha humanoid robot and observed more preferences for the human voice. Dou, Wu, Linz, Gan, and Tseng (2020) compared a synthesized voice with three human voices in different robot applications, including shopping reception, home companion, and education, and found that the synthesized voice was accepted less than human voices. Several studies without robot embodiment found that voice human-likeness positively correlates with users' acceptance (Schreibelmayr & Mara, 2022), and voices with higher human-likeness might bring more pleasure and be preferred more than synthesized voices (Kuhne et al., 2020).

Beyond users' subjective perception, the human-likeness of robot voice might influence users' visual attention toward the robots in voice conversations. In face-to-face conversations, individuals would look at their partner's eyes, mouth, or face to retrieve information about the speech contents, emotions, and attitudes (Degutyte & Astell, 2021). With regard to robots, their speech could effectively catch people's visual attention to initiate communications (Okafuji et al., 2020). A publicly speaking Pepper robot could maintain the attention of audiences when the audiences are interested in the speech (Bourguet, Xu, Zhang, Urakami, & Venture, 2020). Moreover, the human-likeness of robot voices has been broadly manifested to impact users' subjective feelings (Schreibelmayr & Mara, 2022; Seaborn et al., 2021; Xu, 2019). A robot with a high human-like voice might bring more pleasure during voice conversations (Kuhne et al., 2020). However, there has been no literature examining whether the human-likeness of robot voice impacts users' visual attention in voice conversations, particularly for those humanoid robots without facial movements such as grinning, frowning, and closing eyes.

Besides, there is a critical gap between the human-likeness of robot voices and users' cerebral activity. Earlier studies comparing human voices and harmonic tones observed diverse neural activity in listeners (Charest et al., 2009; Levy, Granot, & Bentin, 2003). Those synthesized voices with low human-likeness might cause difficulty for users to percept and process the acoustic features. The pleasure evoked by human-like voices (Kuhne et al., 2020) might activate more cerebral activity. Nevertheless, whether and how the human-likeness of robot voice will impact users' cerebral activity remain unclear. Without this knowledge, it might be hard to deeply understand users' perceptual and cognitive processing of voice human-likeness in conversations, which is valuable for designing social robots.

1.3. Possible joint effect of robot gaze and voice human-likeness

When conversating with a robot, users simultaneously perceive the voice and gaze of a robot through audiovisual sensors and integrate them into a coherent percept. The inherent associations between robot voices and actions might jointly impact users' perception. For example, Tsiourti et al. (2019) observed that emotional voices and behaviors of robots expressing happiness, sadness, and surprise impact interactively on users' recognition.

Earlier research observed significant joint effects when examining different anthropomorphic features of robots (Roesler et al., 2021), including the appearance and voice (Sarigul, Saltik, Hokelek, & Urgen, 2020) and the appearance and communication modes (Klüber & Onnasch, 2022). Studies on voice conversation also observed apparent interaction effects between the anthropomorphic features of robots. Dou, Wu, Niu, and Pan (2021) found an interaction effect of robot voice type and head-light color on users' acceptance that children's voice and neutral light color are accepted more by users for home companion robots. Tatarian et al. (2021) investigated users' proxemics for social navigation, gaze kinesics, and social dialog of robot behavior. They indicated that multimodal behaviors influence users' perception of social intelligence collectively rather than separately. Thepsoonthorn et al. (2021) compared different combinations of robot behaviors in a speech listening task, including face tracking, gaze, head nodding, head nodding response, gestures, and Kinect. They advocated that the combinations of robot social behaviors would enhance users' perception of affinity.

In line with the above interaction effects of different anthropomorphic features, robot gaze and voice human-likeness, two fundamental anthropomorphic features occurring in human-robot voice interaction (Duffy, 2003; Roesler et al., 2021), possibly jointly impact users' subjective perception, visual attention, and cerebral activity. Nevertheless, to our knowledge, no attempts have been found to investigate the joint effect of robot gaze and voice human-likeness. Earlier research manifested that for a robot without gaze, displaying another social cue (gestures) diminishes the social responses of users, while for a robot with gaze, the display of an additional social cue (gestures) increases users' social responses (Ham et al., 2015). The robot gaze and voice human-likeness, which have been evidenced to individually impact users' subjective perception (Babel et al., 2021; Kuhne et al., 2020; Mutlu et al., 2012; Schreibelmayr & Mara, 2022), probably jointly impact users. Thereby, it will be imperative and promising to investigate whether and how robot gaze and voice human-likeness jointly impact users' subjective perception, visual attention, and cerebral activity.

1.4. Aims of the current study

To sum up, robot gaze and voice human-likeness are two fundamental anthropomorphic features in human-robot voice interaction (Roesler et al., 2021). Earlier studies separately examined robot gaze or voice human-likeness on users' subjective perception and indicated significant importance (Admoni & Scassellati, 2017; Kompatsiari, Ciardo, et al., 2021; Kuhne et al., 2020; Xu, 2019; Yang et al., 2018). Research on the effects of robot gaze and voice human-likeness on users' visual attention in voice conversations is scarce. Moreover, extant studies contributing to revealing the effect of robot gaze on users' cerebral activity reached limited consensus (Belkaid et al., 2021; Kelley et al., 2021; Kompatsiari, Bossi, & Wykowska, 2021), and more neuroimaging evidence from voice interactions are needed. As a result, there are several critical research gaps regarding the effects of robot gaze on users' visual attention and cerebral activity, the effects of voice human-likeness on users' visual attention and cerebral activity, and the possible joint effect of robot gaze and voice human-likeness on users' subjective perception, visual attention, and cerebral activity in voice conversations. Grounded on the aforementioned studies and these research gaps, we proposed three primary research questions. 1) Whether and how does robot gaze impact users' subjective perception, visual attention, and cerebral activity? 2) Whether and how does the voice human-likeness robot voice impact users' subjective perception, visual attention, and cerebral activity? 3) Whether and how do robot gaze and voice human-likeness joint impact users' subjective perception, visual attention, and cerebral activity?

Regarding the first research question, robot gaze has been extensively observed to increase users' subjective perception (Babel et al., 2021; Mutlu et al., 2012). It also attracted longer total fixation time in tasks of listening to robot speech (Thepsoonthorn et al., 2021), chatting with morph robot faces (Perugia et al., 2021), and watching videos of robot gaze (Ghiglino et al., 2020, 2021). Moreover, the gaze of an iCub robot evoked higher EEG activities in gaze cueing tasks (Belkaid et al., 2021; Kompatsiari, Bossi, & Wykowska, 2021). For these reasons, we hypothesized the effect of robot gaze in voice conversations as the following.

H1. Robot gaze will (a) increase users' positive emotions and perceived likability, decrease negative attitudes, (b) attract more visual attention, and (c) activate more brain activity on the DLPFC in voice conversations.

The human-likeness of robot voices could bring users more pleasure (Kuhne et al., 2020) and elevate users' acceptance (Schreibelmayr & Mara, 2022) in conversations and maintain audiences' visual attention in a speech listening task (Bourguet et al., 2020). A high human-like voice might attract users' initial attention at the beginning of a conversation and maintain more sustained attention during the conversation. Further, the human-likeness of synthetized voices is correlated with the prosody and naturalness of voice. Therefore, voice human-likeness might not only influence the activity of DLPFC, which is associated with emotions, interpersonal relationships, theory of mind, and comprehension of others' emotional states (Balconi, Fronda, & Bartolo, 2021), but also Broca's area, which is in charge of perceptual processing of prosodic features (Friederici & Alter, 2004), speech production (Opitz, Müller, & Friederici, 2003), and phonemic segmentation (Burton, Small, & Blumstein, 2000). In this regard, we hypothesized that.

H2. Robot with a high human-like voice will (a) increases users' positive emotions and perceived likability, decrease negative attitudes, (b) attract more visual attention, and (c) activate more brain activity in the DLPFC and Broca's area.

As two fundamental anthropomorphic features, robot gaze and voice human-likeness might have a match relationship with each other (Tatarian et al., 2021; Tsiourti et al., 2019). The high human-like voice, which sounds more natural, might induce more pleasure in users (Kuhne et al., 2020) and evoke more positive attitudes (Schreibelmayr & Mara, 2022). It may need to match a gaze rather than no gaze to reach a consistent level of anthropomorphism. That might facilitate users' perception and processing of robot gaze. Inversely, the low human-like voice might constitute an inconsistent anthropomorphism when paired with gaze and attenuate the effect of gaze. Thus, we proposed our hypothesis as the following.

H3. For the robot with a high human-like voice, its gaze will increase (a) users' positive emotions and perceived likability, decrease negative attitudes, (b) attract more visual attention, and (c) activate larger cerebral activity in the DLPFC and less cerebral activity in the Broca's area. For the robot with a low human-like voice, its gaze will (a) elicit similar positive emotions, perceived likability, and negative attitudes, (b) attract similar visual attention, and (c) activate similar cerebral activities in the DLPFC and Broca's area.

To test these hypotheses, we proposed a within-subject designed experiment combining subjective reporting, eye-tracking, and functional near-infrared spectroscopy (fNIRS) technologies to investigate the effect of robot gaze and voice human-likeness on users' subjective perception, visual attention, and cerebral activity in voice conversations. Eye-tracking technology is extensively used to capture and quantify users' visual attention through fixation-related metrics (Amso, Haas, & Markant, 2014; Kuo, Chen, Koyama, & Chang, 2021). And fNIRS is non-inversive and insensitive to motions and could provide a real-time and objective measurement of users' brain activity. Its comparatively high artifact resistance (Ferrari & Quaresima, 2012) has enabled its successful use in human-human social interaction (Jiang et al., 2012; Quaresima & Ferrari, 2019) and human-robot conversation (Keshmiri, Sumioka, Yamazaki, & Ishiguro, 2019a; 2019b). The combined method could unravel users' subjective perception, visual attention, and cerebral activities in real conversations without interruptions.

Then, we designed voice conversation tasks with different robot gaze (gaze versus without gaze) and human-like voices (low human-like versus high human-like) and recruited twenty-seven participants to complete the voice conversation tasks. We recorded participants' eye-tracking data and fNIRS signals during the voice conversations and collected subjective evaluations after each conversation to analyze the effects of robot gaze and voice human-likeness. The findings will contribute to understanding the nature of the effects of robot gaze and voice human-likeness on users. In addition, the findings will provide suggestions for the anthropomorphic design of robot gaze and voice, particularly the simultaneous utilization of gaze and voice human-likeness.

2. Method

The experiment was developed in a two × two within-subject design with robot gaze (gaze versus without gaze) and voice human-likeness (high human-like versus low human-like) as independent variables. Dependent variables were measured from three aspects: subjective perception, visual attention, and cerebral activity.

2.1. Participants

Twenty-seven students (13 males and 14 females) with an age range from 19 to 24 years old ($M = 21.7$, $SD = 1.41$) were recruited from a local university as participants. The required sample size was estimated ahead using G × power v3.1.9.7 (Faul, Erdfelder, Lang, & Buchner, 2007) 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 recommended by Cohen (1988). The sample size is 24, which means the 27 participants in this study were sufficient to detect a medium effect size. All participants were healthy and had normal or corrected-to-normal visions, and were free from hearing deficiency. They were unfamiliar with the robot used in this study and provided written informed consent before participation. Each participant received about 25 yuan (RMB) as compensation after the experiment. The experiment was approved by the local university ethical committee.

2.2. Voice conversation task

Voice conversation tasks (Babel et al., 2021; Dou et al., 2020; Zhang, Zhou, & Yuan, 2018) were adopted to explore the effect of robot gaze and voice human-likeness. The design of voice conversation tasks consisted of dialog contents, robot gaze, and robot voices.

2.2.1. Dialog contents

The dialog was conducted under a scenario of planning a trip to a domestic city. Participants were informed of a city name and requested to imagine that she/he needed to inquire about information from a robot to plan the trip. The inquiry included flights, weather, and hotels. Table 1 shows an example of the dialog scripts. The city was limited to domestic second-tier cities to avoid possible worries about the COVID-19 pandemic. Besides, alternative cities were prepared and used when participants had been to the city to control possible interference of familiarity. As a result, seven cities, including one practical, four formal,

Table 1
An example of human-robot voice dialog scripts.

Speakers	Dialog contents
Users:	Alpha, please check the flights to XXX next Monday.
Robot:	OK, the direct flights to XXX from Shenyang next Monday include China Air CA6451, which departs at 7:45 a.m. and arrives at 11:10 a.m. with a ticket price of 450 RMB, and Shenzhen Airlines ZH 2040, which departs at 1:25 p.m. and arrives at 5:10 p.m., with a ticket price of 430 RMB. Is there anything else I need to do?
Users:	Check the weather of XXX next Monday.
Robot:	OK, the weather of XXX next Monday will be cloudy with a temperature of 6–15 °C and have a westerly wind of 3–4 level and a good air index. Is there anything else I need to do?
Users:	Search hotels near the airport in XXX.
Robot:	OK, I have screened hotels near XXX airport as follows. Jinjiang Inn is 7 km from the airport, takes about 12 min by car, and has a price of 210 yuan per night at least. HanTing Hotel is 14 km from the airport, takes about 24 min by car, and has a price of 200 yuan per night at least. Home Inn is 10 km from the airport, takes about 17 min by car, and has a price of 230 yuan per night at least. Is there anything else I need to do?
Users:	No, thank you.
Robot:	OK, I will go to take a break.

and two alternatives, were selected for the experiment. The information of flights, weather, and hotels was designed similarly but differently to control the possible influence of dialog contents. The complete dialog scripts were provided in supplementary data.

2.2.2. Robot gaze

An Alpha2 robot (UBTECH Inc. In Shenzhen, China) was adopted to conversate with and gaze at users in this study. It has 20 degrees of freedom and can locate users through sonar sensors. It had been used successfully in investigating emotional behaviors in an earlier study (Guo, Li, Qu, & Duffy, 2019). The gaze action was implemented by controlling the degree of freedom in the robot's neck and the LEDs in the eyes. Specifically, the robot raised its head at an angle to orient the participants' faces, as shown in Fig. 1. Meanwhile, the robot shinned eye LEDs to simulate blinks. In order to investigate the effect of robot gaze, a condition without gaze was designed as a contrast. The detailed configurations of gaze conditions are shown in Table 2.

2.2.3. Robot speech

In order to obtain robot voices with distinct human-likeness, 23 female voices in Chinese Mandarin were synthesized through the Text-To-Speech system developed by iFlytek CO. LTD (Hefei, China) and initially screened. Then, another sixty-one university students (thirty-one females) aged 20–30 years ($M = 24.1$, $SD = 2.41$) were recruited to rank the voice human-likeness in a pilot survey. After that, the voice (speaker ID: x2_yifei) had the highest average scores of human-likeness ($M = 5.787$, $SD = 1.368$), and the voice (speaker ID: aisxping) with the lowest average scores of human-likeness ($M = 2.098$, $SD = 1.248$) were selected as the formal voice stimuli. The human-likeness of the two voices differ significantly [F (1, 60) = 252.662, $p < 0.001$, $\eta_p^2 = 0.808$]. Then, the dialog scripts were synthesized into speeches using the two voice speakers. The loudness of each speech was unified to 70 dB using Praat software to eliminate the effect of loudness difference, and the durations were scaled to ensure the same speech rate (Zhang et al., 2020a).

2.3. Procedure

The experiment was conducted in a university laboratory with noise attenuated and light controlled. The participants were invited into the lab, informed of experimental contents, and voluntarily signed an informed consent form. Then, they were seated in a chair at the front of the robot, as shown in Fig. 2. After that, participants were equipped with the fNIRS device and eye-tracking glasses.

The formal experimental schema was programmed in E-prime professional 2.0 (Psychology Software Tools, Pittsburgh, PA, USA) and ran as Fig. 3. The fNIRS and eye-tracking devices were calibrated at first. Then participants completed a practice session and four formal sessions. The practice session used another robot voice and had no gaze action. The four formal sessions consisted of the 2 × 2 experimental conditions in a sequence of Latin Square design. Before each session, participants were given a city name and asked to imagine planning a flight trip to the city. In each session, the experiment program played a voice cue to instruct participants to keep quiet for 60 s to obtain a stable baseline of physiological signals. Then, another voice cue sounded to remind the beginning of the voice conversation. Participants started conversing with the robot to inquire about the flights, weather, and hotels, following the dialog scripts. After the conversation finished, participants were asked to evaluate their emotions, perceived likability, and attitudes toward the robot. There was a rest of 2 min between each session. The robot was controlled in the Wizard-of-Oz method. Another experimenter monitored and controlled the robot wirelessly through a mobile application. After the experiment, two participants were excluded because of equipment malfunction. Finally, compete data were obtained from twenty-five participants with an age range of 19–24 years old (fourteen females, mean age = 21.7 years, SD = 1.64). Participants were told the actual operations and thanked for their participation.

Table 2
Configurations of gaze and without gaze conditions.

Gaze conditions	States of the robot head
Without gaze	The robot kept its head down initially, then said OK and shined eye LEDs to stimulate blinks after being called. It kept in head down all the time and went back to the initial state after the dialog finished.
Gaze	The robot kept its head down initially, then said OK, raised its head to gaze at participants' eyes, and shined eye LEDs to simulate blinks after being called. The robot maintained gazing at the participants' eyes when listening and speaking. After the dialog finished, the robot went back to the initial state.

2.4. Measures

2.4.1. Questionnaires

To explore users' subjective perception of robot gaze and voice human-likeness in voice conversations, emotion, perceived likability, and attitudes were measured to reflect the psychological effects on users (Roesler et al., 2021). Users' emotional valence and arousal were assessed by the PAD scale (Mehrabian & Russell, 1974). The perceived likability was measured using the likability items of Cabral, Cowan, Zibrek, and McDonnell (2017). Besides, the Negative Attitudes towards Robots Scale (NARS) (Nomura, Suzuki, Kanda, & Kato, 2006) was

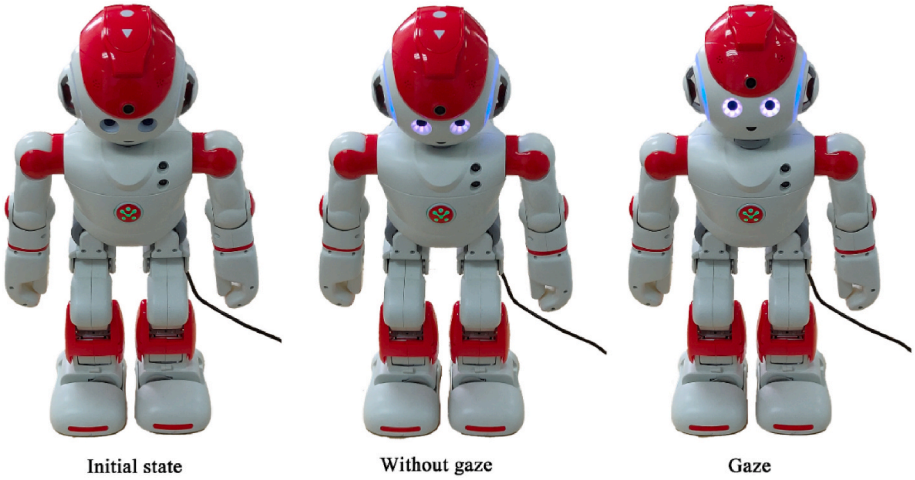

Fig. 1. Design schema of robot gaze actions.

Fig. 2. Experimental scene of human-robot voice conversation.

utilized to evaluate users' attitudes. All these subjective measures were rated using a seven-point Likert scale. Reliability analysis showed good Cronbach's alpha values of more than 0.8.

2.4.2. Eye-tracking measures

An SMI wireless Eye Tracking Glasses 2 (produced by Senso Motoric Instruments, Teltow, Germany) was utilized to measure users' visual attention during voice conversation tasks. The eye-tracker was calibrated in three points, with a tracking accuracy of 0.5° and a sampling rate of 60 Hz. Eye-tracking data were analyzed in Begaze 3.6 software. Referring to the recent findings that the robot head attracts the most visual attention from users when they look at robot appearances (Li et al., 2022), the area of interest (AOI) covering the robot head was defined frame by frame to extract eye-tracking metrics. The first fixation duration, which means the time duration of the first fixation on head area, was extracted to examine users' initial attention. Fixation count, average fixation duration, and total fixation time, which means the number of fixations, the average value of fixation time, and the total time of each fixation in the robot head, respectively, were extracted to examine users' sustained attention.

2.4.3. fNIRS measures

A 22-channel of a continuous-wave (CW) fNIRS system (LIGHTNIRS, Shimadzu Corp., Japan) was used to measure the cerebral activity of prefrontal cortex regions (Fig. 4), which are the primary area of cognitive and emotional processing (Glotzbach et al., 2011; Keshmiri et al., 2019a; 2019b; Wiese, Abubshait, Azarian, & Blumberg, 2019). The center of the below probe set row was placed at the FPZ point in the 10-10 international system. The absorption values of three wavelengths (780, 805, and 830 nm) of near-infrared light were measured at a sampling rate of 13.3 Hz. In addition, a 3D Fastrak digitizer was used to localize the spatial coordinates of optical probes with five reference points (NZ, CZ, IZ, AL, and RL).

fNIRS data were processed in the NIRS_KIT toolbox v2.0 (Hou et al., 2021) in MATLAB (R2020b, MathWorks, Inc.). The absorption values of near-infrared light were transformed into concentration changes of oxyhemoglobin (HbO_2), deoxyhemoglobin (HbR), and total hemoglobin (HbT) as per the modified Beer-Lambert law (Cope & Delpy, 1988). HbO_2 signals were adopted because it has a high signal-to-noise ratio and sensibility to task response (Hoshi, 2003). Then, the HbO_2 signals were de-drifted using 1st-order polynomial regression, motion artifact correction using correlation-based signal improvement (CBSI) method, and band-pass filtered with a third-order IIR filter of 0.01–0.08 Hz. After that, the mean value of hemoglobin concentration changes for each conversation task was calculated and baseline corrected by subtracting the average hemoglobin concentration changes during the quiet period of 60 s. According to the estimation result of MNI space (Singh, Okamoto, Dan, Jurcak, & Dan, 2005), four regions of interest, including the left and the right DLPFC and the left and the right Broca's area, were defined (Table 3). The baseline-corrected hemoglobin concentration

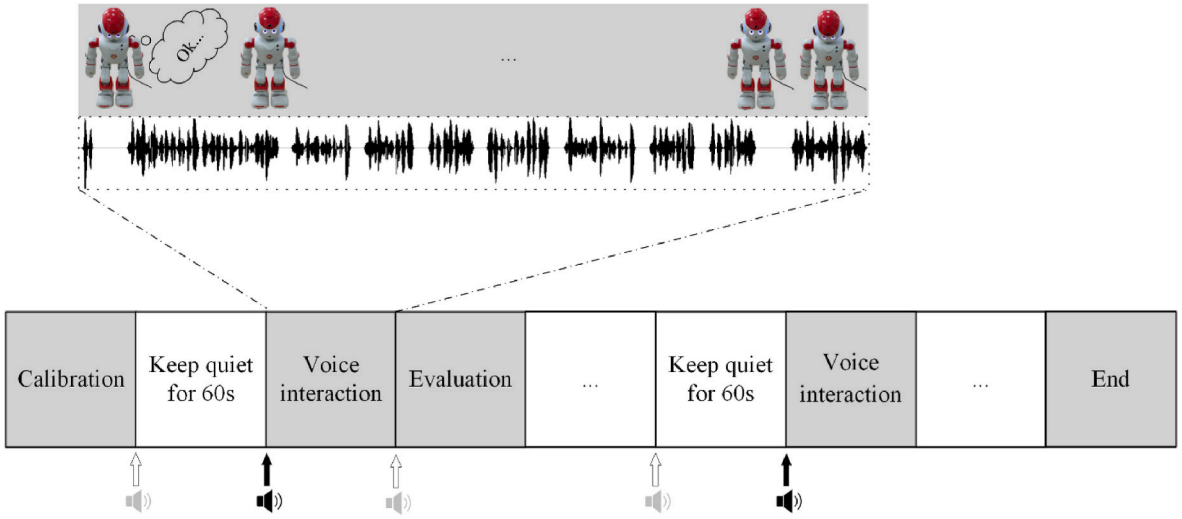

Fig. 3. Experimental schema. The gray arrows and speakers denote the voice cues of keeping quiet. The black ones denote the voice cues of the conversation beginning. After the conversation begins, participants start to talk with the robot as soon as possible.

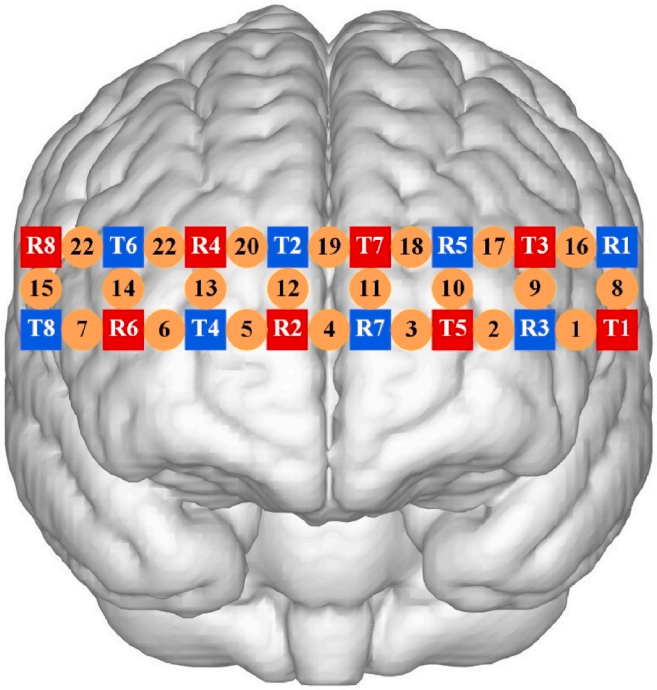

Fig. 4. Placement of optical probes and channels. The red squares represent emitters, the blue ones represent detectors, and the brown circles represent the channels between emitters and detectors.

changes of channels belonging to the same region were averaged for the following statistical analysis.

2.5. Statistical analysis

With a 2 × 2 within-subject experimental design, two-way repeated measures analyses of variance (ANOVA) were adopted to test the effect of robot gaze and voice human-likeness on users' subjective perceptions (perceived likability, pleasure, arousal, and negative attitudes), eye-tracking metrics (first fixation duration, fixation count, average fixation duration, and total fixation time), and fNIRS data (HbO_2 concentration changes in the left DLPFC, right DLPFC, left Broca's area, and right Broca's area). The statistical analyses were conducted in SPSS 19.0 with an α level of 0.05. The partial eta squared (η_p^2) was computed and reported.

Table 3
MIN coordinates of fNIRS channels and corresponding brain regions.

AOI	Channel	Brodmann area (BA)	MIN coordinates			Overlap
			X	Y	Z	
Left DLPFC	CH18	BA9 - Dorsolateral prefrontal cortex	41	41	41	0.578
Right DLPFC	CH20	BA9 - Dorsolateral prefrontal cortex	−57	28	17	0.987
Left Broca's area	CH01	BA45 - pars triangularis, part of Broca's area	−57	28	17	0.885
	CH09	BA45 - pars triangularis, part of Broca's area	−50	36	28	0.967
	CH08	BA44 - pars opercularis Broca's area	−62	8	28	0.606
	CH16	BA44 - pars opercularis Broca's area	0	55	42	0.963
Right Broca's area	CH07	BA45 - pars triangularis, part of Broca's area	59	30	19	0.973
	CH14	BA45 - pars triangularis, part of Broca's area	−40	40	39	0.401
	CH15	BA44 - pars opercularis Broca's area	−20	54	40	0.874
	CH22	BA44 - pars opercularis Broca's area	−24	67	21	0.487

3. Results

3.1. Subjective perception

Fig. 5 shows the mean and standard error of subjective perception measures (The supplementary data shows them in tables). The results of repeated measures ANOVA (Table 4) showed that robot gaze significantly impacted users' pleasure [F (1, 24) = 13.838, p = 0.001, η_p^2 = 0.366] and arousal [F (1, 24) = 28.588, p < 0.0005, η_p^2 = 0.544] (Table 4). Robot gaze brought more pleasure (M = 5.180, 1, SD = 0.94) and higher arousal (M = 4.640, SD = 1.425) to users than no gaze, which brought less pleasure (M = 4.553, SD = 0.908) and lower arousal (M = 3.758, SD = 1.157). Voice human-likeness significantly impacted users' pleasure [F (1, 24) = 31.689, p < 0.0005, η_p^2 = 0.569] and arousal [F (1, 24) = 14.703, p = 0.001, η_p^2 = 0.380] as well. Users' pleasure (M = 5.280, SD = 0.904) and arousal (M = 4.600, SD = 1.278) brought by the high human-like voice were higher than the pleasure (M = 4.440, SD = 0.861) and arousal (M = 3.780, SD = 1.345) brought by the low human-like voice. However, no significant interaction effect of voice human-likeness and gaze was found on users' pleasure [F (1, 24) < 0.0005, p = 1.000, η_p^2 < 0.0005] and arousal [F (1, 24) = 0.016, p = 0.901, η_p^2 = 0.001].

The results of perceived likability showed the main effects of robot gaze [F (1, 24) = 24.923, p < 0.0005, η_p^2 = 0.509] and voice human-likeness [F (1, 24) = 35.030, p < 0.0005, η_p^2 = 0.593], while insignificant interaction of them [F (1, 24) = 0.025, p = 0.877, η_p^2 = 0.001]. The robot with gaze (M = 5.120, SD = 1.172) or the high human-like voice (M = 5.240, SD = 1.061) was perceived as more likable by users than that without gaze (M = 4.238, SD = 0.975) or with the low human-like voice (M = 4.100, SD = 0.974).

Additionally, robot gaze [F (1, 24) = 10.763, p = 0.003, η_p^2 = 0.310] and voice human-likeness [F (1, 24) = 57.610, p < 0.0005, η_p^2 = 0.706] impacted users' negative attitudes. The robot with gaze (M = 2.798, SD = 1.156) or a high human-like voice (M = 2.324, SD = 0.795) induced fewer negative attitudes than that without gaze (M = 3.295, SD = 1.134) or with the low human-like voice (M = 3.778, SD = 1.021). But, no significant interaction effect was observed on users' negative attitudes [F (1, 24) = 0.978, p = 0.333, η_p^2 = 0.039].

3.2. Eye-tracking metrics

Fig. 6 shows the mean and standard error of eye-tracking metrics on the robot head. Repeated measures ANOVA (Table 5) showed that voice human-likeness significantly impacted users' first fixation durations [F (1, 24) = 4.352, p = 0.048, η_p^2 = 0.153]. The robot with the high human-like voice attracted longer first fixation durations (M = 391.840, SD = 347.867) than that with the low human-like voice (M = 275.740, SD = 200.059). However, the robot with gaze attracted similar first fixation durations (M = 322.660, SD = 244.598) as that without gaze (M = 344.484, SD = 328.377) [F (1, 24) = 0.414, p = 0.526, η_p^2 = 0.017]. No significant interaction effect was observed on users' first fixation durations [F (1, 24) = 1.313, p = 0.263, η_p^2 = 0.052].

Robot gaze significantly impacted users' average fixation durations [F (1, 24) = 6.776, p = 0.016, η_p^2 = 0.220]. The robot with gaze attracted longer average fixation durations (M = 381.447, SD = 138.073) than that without gaze (M = 354.762, SD = 115.611). The robot with the high human-like voice attracted similar average fixation durations (M = 364.004, SD = 131.339) compared to that with the low human-like voice (M = 371.671, SD = 124.612) [F (1, 24) = 0.293, p = 0.593, η_p^2 = 0.012]. No significant interaction effect with robot gaze was found [F

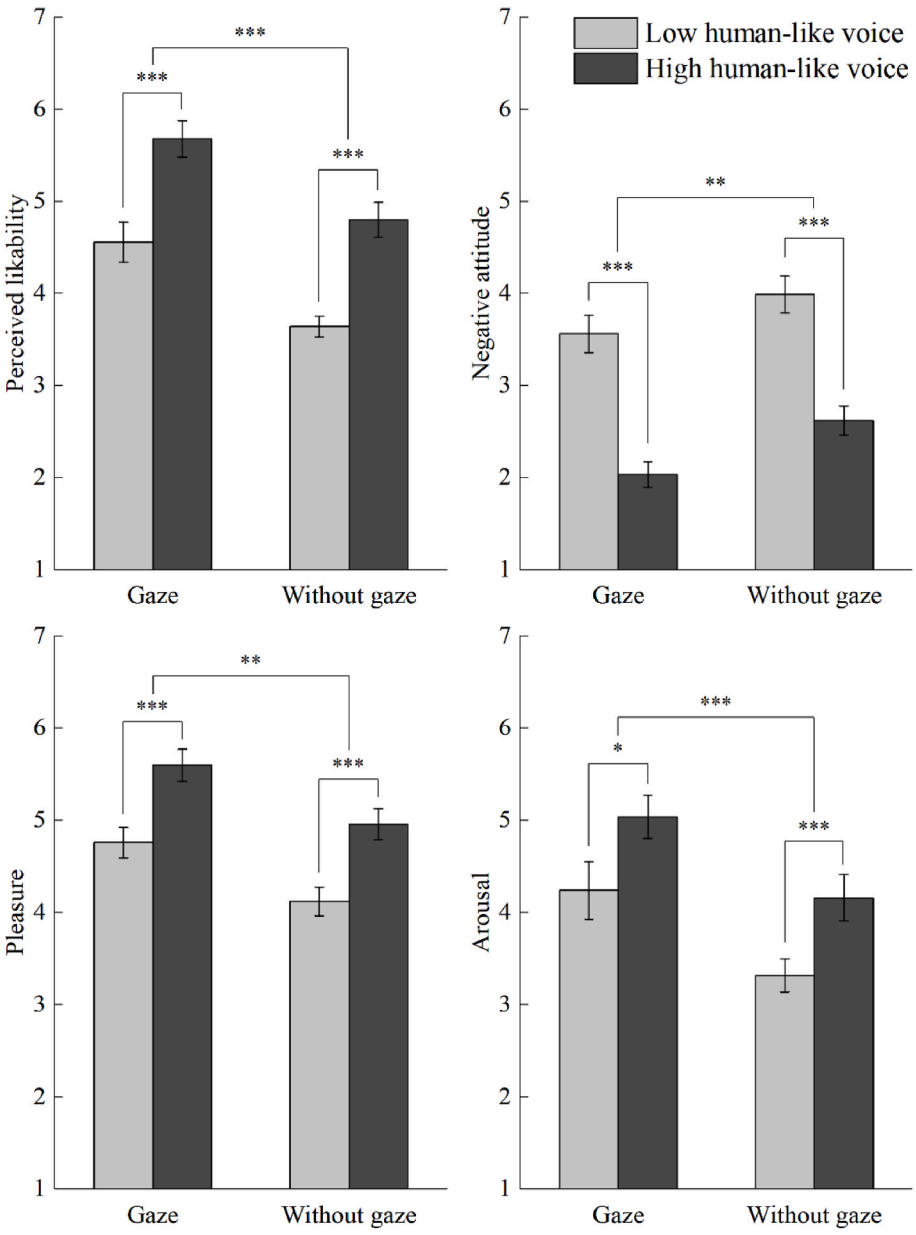

Fig. 5. Users' subjective perceptions in voice conversations. Error bars indicate *SEM*. * depicts $p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

(1, 24) = 0.017, $p = 0.896$, $\eta_p^2 = 0.001$].

In addition, robot gaze significantly impacted users' total fixation time [F (1, 24) = 4.385, $p = 0.047$, $\eta_p^2 = 0.154$], and the robot with gaze attracted higher total fixation time (M = 41,988.234, SD = 18,606.1241) from users than that without gaze (M = 38,315.414, SD = 16,764.873). Likewise, no significant difference in total fixation time was attracted by the robot with high (M = 40,405.440, SD = 19,076.283) and low human-like voice (M = 39,824.752, SD = 16,441.798) [F (1, 24) = 0.084, $p = 0.775$, $\eta_p^2 = 0.003$]. No significant interaction effect [F (1, 24) = 1.338, $p = 0.259$, $\eta_p^2 = 0.053$] was found on users' total fixation time. The results of fixation count showed neither significant main effects of robot gaze [F (1, 24) = 0.133, $p = 0.719$, $\eta_p^2 = 0.006$] and voice human-likeness [F (1, 24) = 0.027, $p = 0.870$, $\eta_p^2 = 0.001$] nor interaction effect [F (1, 24) = 2.437, $p = 0.132$, $\eta_p^2 = 0.092$].

3.3. fNIRS data

Fig. 7 illustrates the concentration changes of HbO_2. Repeated measures ANOVA (Table 6) showed that robot gaze and voice human-likeness significantly interactively affected users' HbO_2 concentration changes in the left DLPFC [F (1, 24) = 4.283, $p = 0.049$, $\eta_p^2 = 0.151$]. The simple effect of robot gaze showed that the robot with gaze induced a higher concentration increase of HbO_2 in users' left DLPFC ($M = 0.0015$, $SD = 0.0021$) than that without gaze ($M = 0.0002$, $SD = 0.0018$) when they had a low human-like voice ($p = 0.046$). When the robot had a high human-like voice, the difference was insignificant between the robot with gaze ($M = 0.0007$, $SD = 0.0015$) and without gaze ($M = 0.0011$, $SD = 0.0017$) ($p = 0.365$). The simple effect of voice human-likeness showed that when the robot had no gaze, the high human-like voice induced more concentration increases of HbO_2 in users' left DLPFC than the low human-like voice ($p = 0.047$). The difference in the HbO_2 concentration changes between the high and low voices was insignificant when the robot had gaze ($p = 0.146$).

The results of HbO_2 concentration changes in the right DLPFC showed neither significant main effects of robot gaze [F (1, 24) = 0.314, $p = 0.581$, $\eta_p^2 = 0.013$] and voice human-likeness [F (1, 24) = 0.492, $p = 0.490$, $\eta_p^2 = 0.020$] nor interaction effect [F (1, 24) = 0.108, $p = 0.746$, $\eta_p^2 = 0.004$]. Likewise, the results of HbO_2 concentration changes in the

Table 4
Repeated measures ANOVA results of pleasure, arousal, perceived likability, and negative attitudes.

Subjective perceptions	Factors	F (1,24)	p	η_p^2
Pleasure	Gaze	13.838	**0.001**	0.366
	Voice human-likeness	31.689	**<0.001**	0.569
	Gaze × Voice human-likeness	<0.001	1.000	<0.001
Arousal	Gaze	28.588	**<0.001**	0.544
	Voice human-likeness	14.703	**0.001**	0.380
	Gaze × Voice human-likeness	0.016	0.901	0.001
Perceived likability	Gaze	24.923	**<0.001**	0.509
	Voice human-likeness	35.030	**<0.001**	0.593
	Gaze × Voice human-likeness	0.025	0.877	0.001
Negative attitudes	Gaze	10.763	**0.003**	0.310
	Voice human-likeness	57.610	**0.000**	0.706
	Gaze × Voice human-likeness	0.978	0.333	0.039

Table 5
Repeated measures ANOVA results of eye-tracking metrics.

Eye-tracking metrics	Factors	F (1,24)	p	η_p^2
First fixation duration	Gaze	0.414	0.526	0.017
	Voice human-likeness	4.352	**0.048**	0.153
	Gaze × Voice human-likeness	1.313	0.263	0.052
Average fixation duration	Gaze	6.776	**0.016**	0.220
	Voice human-likeness	0.293	0.593	0.012
	Gaze × Voice human-likeness	0.017	0.896	0.001
Fixation count	Gaze	0.133	0.719	0.006
	Voice human-likeness	0.027	0.870	0.001
	Gaze × Voice human-likeness	2.437	0.132	0.092
Total fixation time	Gaze	4.385	**0.047**	0.154
	Voice human-likeness	0.084	0.775	0.003
	Gaze × Voice human-likeness	1.338	0.259	0.053

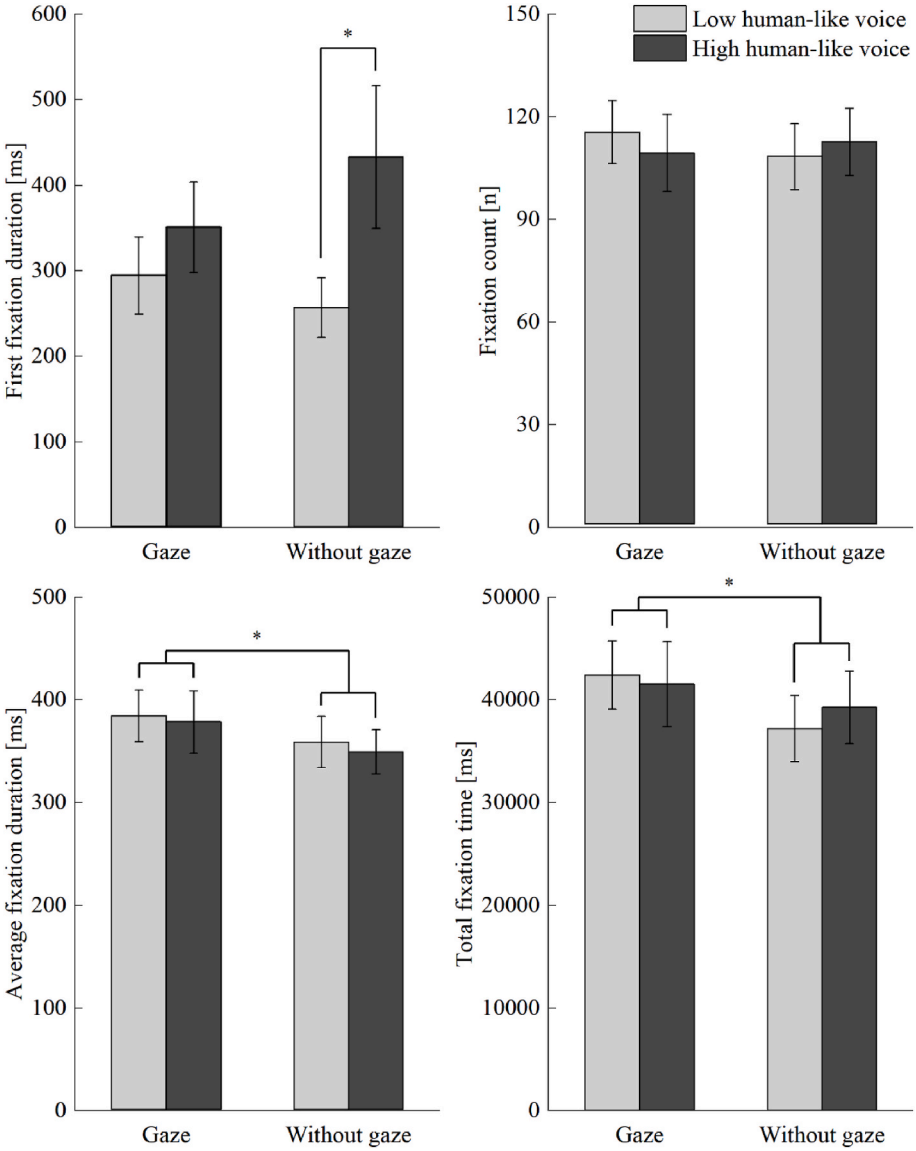

Fig. 6. Users' eye-tracking metrics. Error bars indicate *SEM*. * depicts $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

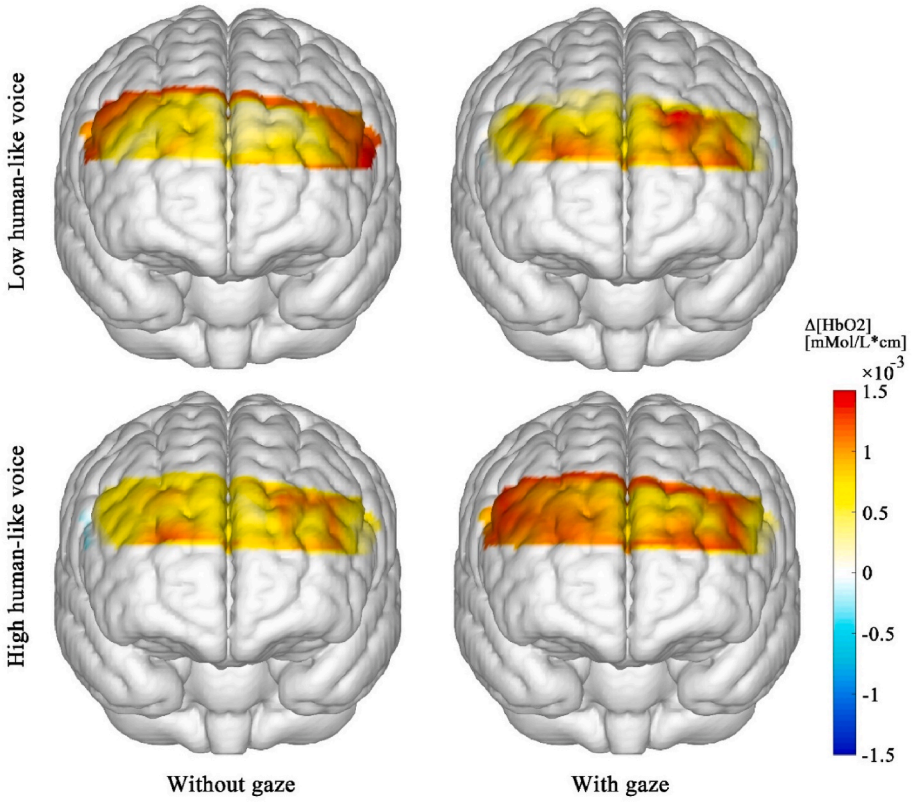

Fig. 7. Cerebral activity of users.

Table 6
Repeated measures ANOVA results of the concentration change of HbO_2.

Cerebral regions	Factors	F (1,24)	p	η_p^2
Left DLPFC	Gaze	1.223	0.280	0.048
	Voice human-likeness	0.092	0.765	0.004
	Gaze × Voice human-likeness	4.283	**0.049**	0.151
Right DLPFC	Gaze	0.314	0.581	0.013
	Voice human-likeness	0.492	0.490	0.020
	Gaze × Voice human-likeness	0.108	0.746	0.004
Left Broca's area	Gaze	2.390	0.135	0.091
	Voice human-likeness	0.440	0.513	0.018
	Gaze × Voice human-likeness	2.181	0.153	0.083
Right Broca's area	Gaze	0.006	0.937	<0.001
	Voice human-likeness	0.004	0.951	<0.001
	Gaze × Voice human-likeness	6.677	**0.016**	0.218

left Broca's area showed no significant main effect of robot gaze [F (1, 24) = 2.390, p = 0.135, η_p^2 = 0.091] and voice human-likeness [F (1, 24) = 0.440, p = 0.513, η_p^2 = 0.018] and their interaction effect [F (1, 24) = 2.181, p = 0.153, η_p^2 = 0.083].

The results of HbO_2 concentration changes in the right Broca's area showed that robot gaze and voice human-likeness had a significant interaction effect [F (1, 24) = 6.677, p = 0.016, η_p^2 = 0.218]. The simple effect of robot gaze showed that the robot with gaze induced a lower HbO_2 concentration increase in the right Broca's area (M = 0.0001, SD = 0.0012) compared with that without gaze (M = 0.0012, SD = 0.0020) when robots had the low human-like voice (p = 0.028). In contrast, the robot with gaze induced a similar HbO_2 concentration increase in the right Broca's area (M = 0.0012, SD = 0.0027) as that without gaze (M = 0.0000, SD = 0.0012) when the robot had the high human-like voice (p = 0.063). The simple effect of robot voice human-likeness showed that the robot with the high human-like voice induced a lower increase of HbO_2 concentration in the right Broca's area than that with the low human-like voice when robots had a gaze (p = 0.042). In contrast, the difference was insignificant when robots had no gaze (p = 0.059).

3.4. Correlations between measures

The correlation analysis was conducted to explore the relationships between the measures of users' subjective perception, visual attention, and cerebral activity (Table 7). The results showed that users' pleasure, arousal, perceived likability, and negative attitudes were significantly correlated with each other ($p < 0.001$). The average fixation duration was significantly correlated with arousal ($p = 0.030$), first fixation duration ($p = 0.002$), and total fixation duration ($p < 0.001$). No obvious significant correlations with the concentration changes of HbO_2 in the left DLPFC and right Broca's area were observed ($p > 0.05$).

4. Discussion

With the purpose of examining the effect of robot gaze and voice human-likeness on users' subjective perception, visual attention, and cerebral activity, we designed robot gaze, synthesized high or low human-like voices, and combined them into different voice conversational tasks. Eye-tracking metrics and fNIRS signals were recorded during these tasks, and subjective ratings were collected as well. The results showed that robot gaze and voice human-likeness had diverse effects on users' subjective perception, visual attention, and cerebral activity.

4.1. Effects on users' subjective perception

Our results showed that the robot with gaze or a high human-like voice separately brought users more pleasure, higher arousal, more perceived likability, and less negative attitudes. The results support H1a and H2a but decline H3a. When the robot gazed at users, it would raise its head to gaze at users and shin eye LEDs to stimulate blinks. These social cues might be perceived as responsive signals and induce more

Table 7
Pearson's correlation coefficient between measures.

Pearson's correlations		Pleasure	Arousal	Perceived likability	Negative attitudes	FFD	AFD	TFT	Left DLPFC	Right Broca
Pleasure	r	1	**0.638**	**0.788**	**−0.637**	−0.018	−0.075	−0.012	0.100	−0.045
	p		**<0.001**	**<0.001**	**<0.001**	0.858	0.459	0.906	0.320	0.654
Arousal	r		1	**0.689**	**−0.502**	−0.143	**−0.217**	−0.019	0.045	−0.123
	p			**<0.001**	**<0.001**	0.155	**0.030**	0.852	0.653	0.222
Perceived likability	r			1	**−0.641**	0.043	−0.053	−0.013	0.011	−0.122
	p				**<0.001**	0.672	0.602	0.895	0.912	0.225
Negative attitudes	r				1	−0.087	−0.004	0.007	−0.163	−0.110
	p					0.392	0.967	0.943	0.106	0.274
FFD	r					1	**0.310**	−0.046	−0.018	−0.046
	p						**0.002**	0.651	0.860	0.646
AFD	r						1	**0.532**	0.016	0.051
	p							**<0.001**	0.874	0.615
TFT	r							1	0.136	0.057
	p								0.178	0.572
Left DLPFC	r								1	0.128
	p									0.206
Right Broca	r									1

Note: FFD denotes the first fixation duration, AFD denotes the average fixation duration, and TFT denotes the total fixation time.

pleasure and higher arousal in users (Babel et al., 2021). The high human-like voice, which was pronounced more clearly, and naturally, might be more attractive and result in positive impressions in aesthetic appreciation (Kawasaki & Yamaguchi, 2012). These positive emotions may impact users' preferences and negative attitudes toward the robot. In a recent study without robot embodiment, voice human-likeness was manifested to impact users' pleasure, preference, and trust (Kuhne et al., 2020). Beyond the virtual robot, we used a physical humanoid robot in voice conversation tasks, and the results suggest that the robot gaze and voice human-likeness in voice conversations might separately induce more positive emotions, elevate users' preferences, and reduce negative attitudes toward the robot.

Nevertheless, the decline of H3a seems inconsistent with the prior findings that emotional voices and behaviors interactively impact the multimodal emotional expression of robots (Tsiourti et al., 2019). The insignificant interaction effect on users' subjective perception might be attributed to the human-likeness level of the synthesized voices. In this study, we excluded fearful or eerier voices during the selection of stimuli because user-centered design pursues bringing users more positive experiences rather than fearful or negative ones. The high and low human-like voices might have no obvious incongruence with robot gaze and result in insignificant interaction effects on users' subjective perceptions. Consequently, the results signify that the human-likeness of robot voice might have no obvious modulations on the effect of robot gaze on users' subjective perceptions, including pleasure, arousal, perceived likability, and attitudes in voice conversations. The gaze effect on subjective perception might be related to the anthropomorphic level of eyes. Compared to highly anthropomorphic eyes in robots, for example, the iCub robot (Belkaid et al., 2021), the eyes of the Alpha 2 robot look less anthropomorphic and might be less intensive in impacting users' subjective perception interactively with voice human-likeness. Thus, robots with different anthropomorphic eyes might be needed to understand further whether and how robot gaze and voice human-likeness interactively impact users' subjective perception in voice conversations.

4.2. Effects on users' visual attention

Regarding eye-tracking metrics, the robot with gaze attracted longer average fixation durations and total fixation time. Average fixation duration is associated with the processing depth of visual information (Just & Carpenter, 1976; Ramos Gameiro et al., 2017; Unema et al., 2005), and fixation counts and total fixation time measure the total amount of invested visual attention (Behe et al., 2015; Wang et al., 2020). Thus, H1b is supported that robot gaze attracted more sustained attention. However, the insignificance of fixation counts indicates that the increment of total visual attention reflected by total fixation time might be attributed to the fact that robot gaze has features of shining eye LEDs to simulate blinks and fixating on users' eyes. These features in voice conversation might convey a strong sense of being responded to and focused on, which may trigger deeper visual processing and result in longer average fixation durations. Recently, several studies observed longer total fixation time for robot gaze in tasks of listening to robot speech (Thepsoonthorn et al., 2021), chatting with morph robot faces (Perugia et al., 2021), and watching videos of robot gaze (Ghiglino et al., 2020, 2021). According to our results from multiple eye-tracking metrics, the increase in total fixation time might be due to deeper visual processing. Our results of voice conversations between users and robots yield that robot gaze might trigger deeper visual processing and attract more amount of visual attention in human-robot voice conversations.

In addition, the result that voice human-likeness only significantly impacted users' first fixation durations partially supports H2b. First fixation duration is typically associated with initial attention and driven by stimulus salience (Garza, Heredia, & Cieslicka, 2016). In the present study, users devoted longer first fixation durations to the head of the robot with a high human-like voice than that with a low human-like voice. This result manifests that the high human-like voice might attract more interest from users at the beginning of the voice conversation task. The insignificant effect of voice human-likeness on users' sustained attention might be related to the conversation task. In this study, we designed conversations of inquiring information from the robot to plan trips. After an initial initiation of conversation, users spoke and listened to inquire about information. They might concentrate their attention on the information under a top-down modulation from the task goal. In future research, diverse types of conversation tasks, such as chatting, interviewing, and negotiation, should be designed to investigate how conversation tasks modulate the effect of voice human-likeness on users' sustained attention.

Nevertheless, we did not observe any significant interaction effects on the eye-tracking metrics. This result declines H3b. The result is speculated to be related to the conversation task. As discussed above, the conversation task of inquiring about information may modulate users' visual attention. Users might be initially attracted by the high human-like voice but allocate more attention to the vocal information rather than voice human-likeness. Overall, the results of eye-tracking metrics suggest that the human-likeness of robot voice might chiefly impact users' initial attention, and robot gaze might primarily capture sustained attention in voice conversations. At the same time, various types of voice conversations are needed to explore the joint effect of robot gaze and voice human-likeness on users' visual attention.

4.3. Effects on users' cerebral activity

Concerning the HbO_2 concentration changes, the robot with a high human-like voice and no gaze or with a low human-like voice and gaze induced increased activity in the left DLPFC. The results seem inverse to H3c. Previous studies on emotional processing manifested that positive emotions might activate the left DLPFC, and negative emotions might activate the right DLPFC (Gray, Braver, & Raichle, 2002; Herrington et al., 2005; Wheeler, Davidson, & Tomarken, 2007). In the present study, robot gaze and voice human-likeness might constitute a consistent design of anthropomorphism and increase users' emotions of pleasure and arousal significantly. That was hypothesized to activate users' left DLPFC and induce more significant increases of HbO_2 concentration in the left DLPFC when the robot had a low human-like voice or had no gaze. However, users showed no significant enhancement of HbO_2 concentration changes in the left DLPFC, although they reported higher pleasure and arousal when the robot had a high human-like voice with gaze. It is supposed that the elevation of users' pleasure and arousal might evoke no more increase in the activity of left DLPFC when the emotional arousal reaches a high level (Balconi, Grippa, & Vanutelli, 2015). Recently, Kompatsiari, Bossi, and Wykowska (2021) found that robot gaze could elevate users' interaction immersion and induce higher alpha rhythm desynchronization in a gaze-cueing task. Grounded on this finding, our results provide further brain imaging evidence that both robot gaze and voice human-likeness might activate more emotional processing in the left DLPFC when users are in a low level of emotional arousal during the human-robot voice conversations.

Some extant studies found changes in users' cerebral activity in the right DLPFC in human-computer interaction (Piva, Zhang, Noah, Chang, & Hirsch, 2017), human-automation system interaction (Huang, Choo, Pugh, & Nam, 2022), and the eye-contacts when listening to another person talking (Jiang, Borowiak, Tudge, Otto, & von Kriegstein, 2017). In contrast, we observed no significant difference in the cerebral activity of the right DLPFC. The right DLPFC is considered to be associated with motivation (Forbes & Grafman, 2010), negative emotions (Gray et al., 2002), and cognitive assessment (Lu et al., 2019; Murray & Ranganath, 2007). In the present study, users' pleasure reached no less than a neutral level, and there might be no oblivious negative emotions were elicited to activate the right DLPFC. A recent comparison between the human gaze and the robot gaze found that robot gaze elicited insignificant activity in the right DLPFC of viewers (Kelley et al., 2021). Although they used static gaze without interaction tasks, their results may provide partial support for our findings of dynamic gaze in voice conversations. The voice conversation tasks in this study differ from the collaboration tasks with computers (Piva et al., 2017) or automation (Huang et al., 2022) and chiefly contain vocal inquiry and information reception. Consequently, there might be no significant changes in motivation or cognitive assessment in the voice interaction tasks to modulate cerebral activity in the right DLPFC.

Moreover, the robot with the low human-like voice and no gaze induced greater increases in HbO_2 concentration in the right Broca's area than that with the high human-like voice and no gaze or with the low human-like voice and gaze. The left Broca's area plays a vital role in speech production (Opitz et al., 2003) and phonemic segmentation (Burton et al., 2000), while the right Broca's area might be activated in the perceptual processing of prosodic features (Friederici & Alter, 2004). In the current study, the low human-like voice, which sounds inarticulate, blunt, and less natural, might need more cognitive resources in the perceptual processing of prosodic features than the high human-like. The robot without gaze might give users a less feeling of being focused on and hinder users' immersion in conversational interaction. That may increase the difficulty of perceptual processing of the low human-like voice and result in higher activity in the right Broca's area. Conversely, the gaze, which present users with a sense of being responded to and focused on, might facilitate users' processing of the prosodic features of the low human-like voice. Thus, our results indicate that robot gaze and voice human-likeness might jointly influence users' perceptual processing of prosodic features, and the gaze might promote the perceptual processing of prosodic features of human-like voice.

4.4. Multimodal measurement of the anthropomorphic effect

Noticeably, we observed significant internal correlations within subjective perceptions or eye-tracking metrics but insignificant external correlations between measures of different modalities. The gaps between different modalities may explain the insignificant correlations. As Roesler et al. (2021) indicated in the effectiveness analysis of robot anthropomorphism, there is a gap between users' subjective perceptions and behavioral outcomes. In the current study, we found a significant interaction effect of robot gaze and voice human-likeness on users' cerebral activity but only significant main effects on users' subjective perception and visual attention. The results imply that robot gaze and voice human-likeness separately impact users' subjective perception and visual attention but jointly impact users' perceptual processing of acoustic features and emotional processing. Accordingly, there might also have gaps between users' subjective perception, visual attention, and cerebral activity.

Besides, measures of different modalities might have distinct sensitivities. For example, though subjective reporting is easy to access and straightforward, it is typically conducted after the interaction and hardly provides insight into users' real-time responses without interrupting users' engagement. Therefore, it might be reasonable that the measures of subjective perception, visual attention, and cerebral activity in this study showed insignificant correlations. The results imply that the effect of robot gaze and voice human-likeness involves users' multiple modalities, and its evaluation should combine multiple methods to provide a comprehensive understanding.

The method in this study provided a valid measure of users' visual attention, cerebral activity, and subjective perception in voice conversations with physical robots. Nevertheless, the combination of more methods, such as behavioral outcomes, Electrocardiogram (ECG), Electroencephalogram (EEG), and Functional magnetic resonance imaging (fMRI), might provide a higher ecological method to measure the effect of anthropomorphic design features in HRI. Additionally, Explainable artificial intelligence (XAI) probably unravels the complex relationships between different modalities and fuses them ecologically (Loh et al., 2022).

5. Implications, limitations, and future research

5.1. Implications

This study makes several significant theoretical contributions. Firstly, grounded on earlier findings on the effect of robot gaze on users' visual attention in tasks of listening to robot speech (Thepsoonthorn et al., 2021), chatting with morph robot faces (Perugia et al., 2021), and watching videos of robot gaze (Ghiglino et al., 2020, 2021), this study indicates that the human-likeness of robot voice might attract users' initial attention, and the robot gaze might trigger deeper visual processing and attract a larger total amount of visual attention in human-robot voice conversations. Secondly, it provides brain imaging evidence with a relatively high resolution for research on users' emotional processing of robot gaze. The results confirm the insignificant activity in the right DLPFC by robot gaze (Kelley et al., 2021) and further manifest the emotional activations by robot gaze in the left DLPFC. Moreover, the activity in the right Broca's area reveals the joint effects of robot gaze and voice human-likeness on users' perceptual processing of acoustic features in voice conversations. Finally, the study provides a combined method of subjective reporting, eye-tracking, and fNIRS technologies to probe users' subjective perception, visual attention, and cerebral activity in HRI.

This study provides practical implications for the design of human-

robot voice conversation as well. On the one hand, our results showed that the robot with gaze induced higher subjective perceptions and deeper visual processing than that without gaze. The findings signify that robot designers should consider incorporating gaze action in human-robot voice conversations to construct a feeling of being responded to and focused on. That would enhance users' positive emotions and result in more perception of likability and less negative attitudes in voice conversations. Incorporating robot gaze in voice conversations should accommodate conversation contents and convey explicit intentions to avoid ambiguous gaze intentions (Admoni & Scassellati, 2017). On the other hand, the robot with a high human-like voice brought more positive emotions and longer first fixation durations. The findings indicate that although the human-like voice of robots might not attract sustained attention in voice conversations, it would appeal to users' initial attention. Thereby, designers should establish or adopt the TTS system with a high quality of human-likeness. Additionally, although we observed no significant interaction effect of robot gaze and voice human-likeness on users' subjective perception, they impact users' initial and sustained attention individually and have an interaction effect on the activity in the left DLPFC and right Broca's area. Hence, the combination of robot gaze and voice human-likeness should be considered jointly. Robot gaze and high human-like voice could be used to attract users' initial and sustained attention to increase users' subjective perception. Particularly, when the robot has a low human-like voice or no gaze, gaze or high human-like voices should be utilized to reduce the difficulties in users' perceptual processing of prosodic features and promote users' subjective perception.

5.2. Limitations and future research

The current study highlights whether and how robot gaze and voice human-likeness impact users' subjective perception, visual attention, and cerebral activity in human-robot voice conversations through an elaborated experiment combining subjective reporting, eye-tracking, and fNIRS. Several limitations should be noted when interpreting the findings of this study. A major limitation is that only robot gaze, which frequently occurs in human-robot voice conversations, and voice human-likeness, a critical characteristic of synthesized voice in the real robot, were investigated in this study. Beyond the two factors, robots might have more anthropomorphic features, such as greeting actions (Tatarian et al., 2021), hand gestures (Thepsoonthorn et al., 2021), emotional behaviors (Tsiourti et al., 2019), referential gaze (Manzi et al., 2020), and voice features of pitch, loudness, and gender. Therefore, future work should explore the effects of these features in human-robot voice conversations. Secondly, we recruited university students who generally have similar education levels and more access to new technologies (Burleigh, Schoenherr, & Lacroix, 2013; Kwak, Kim, & Choi, 2017). That might restrict the applicability of results to other populations. Further validations on groups with different ages, cultures, and personalities would extend the generalizability of our results. Last but not least, we utilized the Alpha 2 humanoid robot, which has no dynamic facial features, such as grinning, frowning, or closing eyes. Validations on different robot platforms and application areas (Dou et al., 2021) might extend the generalizability of our findings in the current study.

6. Conclusion

With the purpose of investigating the effect of robot gaze and voice human-likeness on users' subjective perception, visual attention, and cerebral activity, we combined subjective reporting, eye-tracking, and fNIRS technologies to conduct a within-subject design experiment of human-robot voice conversations with different robot gaze and human-like voices. The subjective results showed that the robot with gaze or a high human-like voice elicited more pleasure, higher arousal, more perceived likability, and less negative attitudes. The eye-tracking metrics of users showed longer first fixation durations for the robot with the high human-like voice than that with the low human-like voice and longer average fixation durations and total fixation time for the robot with gaze than without gaze. Moreover, the changes in HbO_2 concentration indicated that the robot with a high human-like voice and no gaze or with a low human-like voice and gaze induced increased activity in left DLPFC and decreased activity in right Broca' area. These results suggest that the human-likeness of robot voice might attract users' initial attention, and the robot gaze might trigger deeper visual processing in human-robot voice conversations. But they might jointly influence users' perceptual processing of prosodic features and emotional processing. Moreover, the effect of robot gaze and voice human-likeness involves users' multiple modalities. These findings provide theoretical contributions for understanding the nature of the effects of robot gaze and voice human-likeness and practical suggestions for the design of human-robot voice conversation.

Author statement

Mingming Li: conceptualization, methodology, software, investigation, writing - original draft, writing - review & editing. Fu Guo: conceptualization, writing - review & editing, supervision, project administration, funding acquisition. Xueshuang Wang: validation, data curation, writing - review & editing. Jiahao Chen: methodology, investigation, validation, data curation, writing - review & editing. Jaap Ham: conceptualization, methodology, writing - review & editing.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 72071035 and Grant No. 72171042). No conflict of interest exists in submitting this paper, and all authors approve it for publication. We are grateful to all the experimental participants for this study. Furthermore, we are genuinely pleased to extend our gratitude to editors and reviewers for their valuable work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2022.107645.

References

Admoni, H., & Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction, 6*(1), 25–63. https://doi.org/10.5898/JHRI.6.1.Admoni

Amso, D., Haas, S., & Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PLoS One, 9*(1), Article e85701. https://doi.org/10.1371/journal.pone.0085701

Andrist, S., Tan, X. Z., Gleicher, M., & Mutlu, B. (2014). *Conversational gaze aversion for humanlike robots. 2014 9th ACM/IEEE international conference on human-robot interaction (HRI)*.

Argyle, M. (1972). Non-verbal communication in human social interaction. In *Non-verbal communication*. Cambridge U. Press.

Babel, F., Kraus, J., Miller, L., Kraus, M., Wagner, N., Minker, W., et al. (2021). Small talk with a robot? The impact of dialog content, talk initiative, and gaze behavior of a social robot on trust, acceptance, and proximity. *International Journal of Social Robotics, 13*(6), 1485–1498. https://doi.org/10.1007/s12369-020-00730-0

Baillon, A., Selim, A., & van Dolder, D. (2013). On the social nature of eyes: The effect of social cues in interaction and individual choice tasks. *Evolution and Human Behavior, 34*(2), 146–154. https://doi.org/10.1016/j.evolhumbehav.2012.12.001

Baird, A., Parada-Cabaleiro, E., Hantke, S., Burkhardt, F., Cummins, N., & Schuller, B. (2018). *The Perception and Analysis of the Likeability and human Likeness of synthesized speech interspeech 2018*. India: Hyderabad.

Balconi, M., Fronda, G., & Bartolo, A. (2021). Affective, social, and informative gestures reproduction in human interaction: Hyperscanning and brain connectivity. *Journal of Motor Behavior, 53*(3), 296–315. https://doi.org/10.1080/00222895.2020.1774490

Balconi, M., Grippa, E., & Vanutelli, M. E. (2015). What hemodynamic (fNIRS), electrophysiological (EEG) and autonomic integrated measures can tell us about emotional processing. *Brain and Cognition, 95*, 67–76. https://doi.org/10.1016/j.bandc.2015.02.001

Bartneck, C., Bleeker, T., Bun, J., Fens, P., & Riet, L. (2010). The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn. Journal of Behavioral Robotics, 1*(2), 109–115. https://doi.org/10.2478/s13230-010-0011-3

Behe, B. K., Bae, M., Huddleston, P. T., & Sage, L. (2015). The effect of involvement on visual attention and product choice. *Journal of Retailing and Consumer Services, 24*, 10–21. https://doi.org/10.1016/j.jretconser.2015.01.002

Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1476), 2091–2107. https://doi.org/10.1098/rstb.2006.1933

Belkaid, M., Kompatsiari, K., Tommaso, D. D., Zablith, I., & Wykowska, A. (2021). Mutual gaze with a robot affects human neural activity and delays decision-making processes. *Science Robotics, 6*(58). https://doi.org/10.1126/scirobotics.abc5044. eabc5044.

Bourguet, M.-L., Xu, M., Zhang, S., Urakami, J., & Venture, G. (2020). The impact of a social robot public speaker on audience attention. *Proceedings of the 8th International Conference on Human-Agent Interaction*.

Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior, 29*(3), 759–771. https://doi.org/10.1016/j.chb.2012.11.021

Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience, 12*(4), 679–690. https://doi.org/10.1162/089892900562309

Cabral, J. P., Cowan, B. R., Zibrek, K., & McDonnell, R. (2017). The influence of synthetic voice on the evaluation of a virtual character. *19th Annual Conference of the International Speech Communication Association (Interspeech 2018), 1–6*.

Chang, R. C.-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. *Computers in Human Behavior, 84*, 194–210. https://doi.org/10.1016/j.chb.2018.02.025

Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., et al. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience, 10*(1), 127. https://doi.org/10.1186/1471-2202-10-127

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cope, M., & Delpy, D. T. (1988). System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Medical, & Biological Engineering & Computing, 26*(3), 289–294. https://doi.org/10.1007/bf02447083

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 201–215. https://doi.org/10.1038/nrn755

Degutyte, Z., & Astell, A. (2021). The role of eye gaze in regulating turn taking in conversations: A systematized review of methods and findings. *Frontiers in Psychology, 12*, 616471. https://doi.org/10.3389/fpsyg.2021.616471

Dou, X., Wu, C. F., Linz, K. C., Gan, S. Z., & Tseng, T. M. (2020). Effects of different types of social robot voices on affective evaluations in different application fields. *International Journal of Social Robotics, 13*(4), 615–628. https://doi.org/10.1007/s12369-020-00654-9

Dou, X., Wu, C. F., Niu, J., & Pan, K. R. (2021). Effect of voice type and head-light color in social robots for different applications. *International Journal of Social Robotics, 14*(1), 229–244. https://doi.org/10.1007/s12369-021-00782-w

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems, 42*(3–4), 177–190. https://doi.org/10.1016/s0921-8890(02)00374-3

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24*(6), 581–604. https://doi.org/10.1016/S0149-7634(00)00025-7

Eyssel, F., Kuchenbrandt, D., Bobinger, S., De Ruiter, L., & Hegel, F. (2012). If you sound like me, you must be more human. *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI, 12*.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage, 63*(2), 921–935. https://doi.org/10.1016/j.neuroimage.2012.03.049

Forbes, C. E., & Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annual Review of Neuroscience, 33*(1), 299–324. https://doi.org/10.1146/annurev-neuro-060909-153230

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language, 89*(2), 267–276. https://doi.org/10.1016/S0093-934X(03)00351-1

Garza, R., Heredia, R. R., & Cieslicka, A. B. (2016). Male and female perception of physical attractiveness: An eye movement study. *Evolutionary Psychology, 14*(1), 1–16. https://doi.org/10.1177/1474704916631614

Ghiglino, D., Willemse, C., De Tommaso, D., & Wykowska, A. (2021). Mind the eyes: Artificial agents' eye movements modulate attentional engagement and anthropomorphic attribution. *Frontiers in Robotics and AI, 8*, Article 642796. https://doi.org/10.3389/frobt.2021.642796

Ghiglino, D., Willemse, C., Tommaso, D. D., Bossi, F., & Wykowska, A. (2020). At first sight: Robots' subtle eye movement parameters affect human attentional engagement, spontaneous attunement and perceived human-likeness. *Paladyn. Journal of Behavioral Robotics, 11*(1), 31–39. https://doi.org/10.1515/pjbr-2020-0004

Glotzbach, E., Mühlberger, A., Gschwendtner, K., Fallgatter, A. J., Pauli, P., & Herrmann, M. J. (2011). Prefrontal brain activation during emotional processing: A functional near infrared spectroscopy study (fNIRS). *The Open Neuroimaging Journal, 5*, 33–39. https://doi.org/10.2174/1874440001105010033

Gray, J. R., Braver, T. S., & Raichle, M. E. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences, 99*(6), 4115–4120. https://doi.org/10.1073/pnas.062381899

Guo, F., Li, M., Qu, Q., & Duffy, V. G. (2019). The effect of a humanoid robot's emotional behaviors on users' emotional responses: Evidence from pupillometry and electroencephalography measures. *International Journal of Human-Computer Interaction, 35*(20), 1947–1959. https://doi.org/10.1080/10447318.2019.1587938

Ham, J., Cuijpers, R. H., & Cabibihan, J.-J. (2015). Combining robotic persuasive strategies: The persuasive power of a storytelling robot that uses gazing and gestures. *International Journal of Social Robotics, 7*(4), 479–487. https://doi.org/10.1007/s12369-015-0280-4

Herrington, J. D., Mohanty, A., Koven, N. S., Fisher, J. E., Stewart, J. L., Banich, M. T., et al. (2005). Emotion-Modulated performance and activity in left dorsolateral prefrontal cortex. *Emotion, 5*(2), 200–207. https://doi.org/10.1037/1528-3542.5.2.200

Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-aho, K., & Ruuhiala, H. J. (2008). Seeing direct and averted gaze activates the approach–avoidance motivational brain systems. *Neuropsychologia, 46*(9), 2423–2430. https://doi.org/10.1016/j.neuropsychologia.2008.02.029

Hoshi, Y. (2003). Functional near-infrared optical imaging: Utility and limitations in human brain mapping. *Psychophysiology, 40*(4), 511–520. https://doi.org/10.1111/1469-8986.00053

Hou, X., Zhang, Z., Zhao, C., Duan, L., Gong, Y., Li, Z., et al. (2021). NIRS-KIT: A MATLAB toolbox for both resting-state and task fNIRS data analysis. *Neurophotonics, 8*(1), Article 010802. https://doi.org/10.1117/1.NPh.8.1.010802

Huang, J., Choo, S., Pugh, Z. H., & Nam, C. S. (2022). Evaluating effective connectivity of trust in human-automation interaction: A dynamic causal modeling (dcm) study. *Human Factors, 64*(6), 1051–1069. https://doi.org/10.1177/0018720820987443

Husic-Mehmedovic, M., Omeragic, I., Batagelj, Z., & Kolar, T. (2017). Seeing is not necessarily liking: Advancing research on package design with eye-tracking. *Journal of Business Research, 80*, 145–154. https://doi.org/10.1016/j.jbusres.2017.04.019

Jiang, J., Borowiak, K., Tudge, L., Otto, C., & von Kriegstein, K. (2017). Neural mechanisms of eye contact when listening to another person talking. *Social Cognitive and Affective Neuroscience, 12*(2), 319–328. https://doi.org/10.1093/scan/nsw127

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience, 32*(45), 16064–16069. https://doi.org/10.1523/jneurosci.2926-12.2012

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Kawasaki, M., & Yamaguchi, Y. (2012). Individual visual working memory capacities and related brain oscillatory activities are modulated by color preferences. *Frontiers in Human Neuroscience, 6*, 318. https://doi.org/10.3389/fnhum.2012.00318

Kelley, M. S., Noah, J. A., Zhang, X., Scassellati, B., & Hirsch, J. (2021). Comparison of human social brain activity during eye-contact with another human and a humanoid robot. *Frontiers in Robotics and AI, 7*, Article 599581. https://doi.org/10.3389/frobt.2020.599581

Keshmiri, S., Sumioka, H., Yamazaki, R., & Ishiguro, H. (2019a). Decoding the perceived difficulty of communicated contents by older people: Toward conversational robot-assistive elderly care. *IEEE Robotics and Automation Letters, 4*(4), 3263–3269. https://doi.org/10.1109/lra.2019.2925732

Keshmiri, S., Sumioka, H., Yamazaki, R., & Ishiguro, H. (2019b). Older people prefrontal cortex activation estimates their perceived difficulty of a humanoid-mediated conversation. *IEEE Robotics and Automation Letters, 4*(4), 4108–4115. https://doi.org/10.1109/lra.2019.2930495

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*(1), 78–100. https://doi.org/10.1037/0033-2909.100.1.78

Klüber, K., & Onnasch, L. (2022). Appearance is not everything - preferred feature combinations for care robots. *Computers in Human Behavior, 128*, Article 107128. https://doi.org/10.1016/j.chb.2021.107128

Kompatsiari, K., Bossi, F., & Wykowska, A. (2021). Eye contact during joint attention with a humanoid robot modulates oscillatory brain activity. *Social Cognitive and Affective Neuroscience, 16*(4), 383–392. https://doi.org/10.1093/scan/nsab001

Kompatsiari, K., Ciardo, F., De Tommaso, D., & Wykowska, A. (2019). Measuring engagement elicited by eye contact in Human-Robot Interaction. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Kompatsiari, K., Ciardo, F., Tikhanoff, V., Metta, G., & Wykowska, A. (2021). It's in the eyes: The engaging role of eye contact in HRI. *International Journal of Social Robotics, 13*(3), 525–535. https://doi.org/10.1007/s12369-019-00565-4

Kompatsiari, K., Tikhanoff, V., Ciardo, F., Metta, G., & Wykowska, A. (2017). *The importance of mutual gaze in human-robot Interaction.Lecture notes in artificial intelligence [social robotics, icsr 2017]*. Tsukuba, JAPAN: 9th International Conference on Social Robotics (ICSR).

Kuhne, K., Fischer, M. H., & Zhou, Y. (2020). The human takes it all: Humanlike synthesized voices are perceived as less eerie and more likable. Evidence from a subjective ratings study. *Frontiers in Neurorobotics, 14*, Article 593732. https://doi.org/10.3389/fnbot.2020.593732

Kuo, J.-Y., Chen, C.-H., Koyama, S., & Chang, D. (2021). Investigating the relationship between users' eye movements and perceived product attributes in design concept evaluation. *Applied Ergonomics, 94*, Article 103393. https://doi.org/10.1016/j.apergo.2021.103393

Kwak, S. S., Kim, J. S., & Choi, J. J. (2017). The effects of organism- versus object-based robot design approaches on the consumer acceptance of domestic robots. *International Journal of Social Robotics, 9*(3), 359–377. https://doi.org/10.1007/s12369-016-0388-1

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., & Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. *Proceedings of the National Academy of Sciences of the United States of America, 114*(50), 13290–13295. https://doi.org/10.1073/pnas.1702493114

Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology, 40*(2), 291–305. https://doi.org/10.1111/1469-8986.00031

Liew, T. W., & Tan, S.-M. (2021). Social cues and implications for designing expert and competent artificial agents: A systematic review. *Telematics and Informatics, 65*, Article 101721. https://doi.org/10.1016/j.tele.2021.101721

Li, M., Guo, F., Ren, Z., & Duffy, V. G. (2022). A visual and neural evaluation of the affective impression on humanoid robot appearances in free viewing. *International Journal of Industrial Ergonomics, 88*, Article 103159. https://doi.org/10.1016/j.ergon.2021.103159

Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine, 226*, Article 107161. https://doi.org/10.1016/j.cmpb.2022.107161

Lu, X., Li, T., Xia, Z., Zhu, R., Wang, L., Luo, Y. J., et al. (2019). Connectome-based model predicts individual differences in propensity to trust. *Human Brain Mapping, 40*(6), 1942–1954. https://doi.org/10.1002/hbm.24503

Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., et al. (2020). The understanding of congruent and incongruent referential gaze in 17-month-old infants: An eye-tracking study comparing human and robot. *Scientific Reports, 10*(1), Article 11918. https://doi.org/10.1038/s41598-020-69140-6

Mavridis, N. (2015). A review of verbal and non-verbal human-robot interactive communication. *Robotics and Autonomous Systems, 63*, 22–35. https://doi.org/10.1016/j.robot.2014.09.031

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. The MIT Press.

Morillo-Mendez, L., Schrooten, M. G. S., Loutfi, A., & Mozos, O. M. (2021). Age-related differences in the perception of eye-gaze from a social robot. In *Social robotics* (pp. 350–361). Springer International Publishing. https://doi.org/10.1007/978-3-030-90525-5_30.

Murray, L. J., & Ranganath, C. (2007). The dorsolateral prefrontal cortex contributes to successful relational memory encoding. *Journal of Neuroscience, 27*(20), 5515–5522. https://doi.org/10.1523/jneurosci.0406-07.2007

Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., & Ishiguro, H. (2012). Conversational gaze mechanisms for humanlike robots. *Acm Transactions on Interactive Intelligent Systems, 1*(2), 1–33. https://doi.org/10.1145/2070719.2070725

Nass, C. I., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies, 43*(2), 223–239. https://doi.org/10.1006/ijhc.1995.1042

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies, 7*(3), 437–454. https://doi.org/10.1075/is.7.3.14nom

Okafuji, Y., Baba, J., Nakanishi, J., Kuramoto, I., Ogawa, K., Yoshikawa, Y., et al. (2020). Can a humanoid robot continue to draw attention in an office environment? *Advanced Robotics, 34*(14), 931–946. https://doi.org/10.1080/01691864.2020.1769724

Opitz, B., Müller, K., & Friederici, A. (2003). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Cognitive Brain Research, 16*(2), 285–296. https://doi.org/10.1016/S0926-6410(02)00284-7

Perugia, G., Paetzel-Prüsmann, M., Alanenpää, M., & Castellano, G. (2021). I can see it in your eyes: Gaze as an implicit cue of uncanniness and task performance in repeated interactions with robots. *Frontiers in Robotics and AI, 8*, Article 645956. https://doi.org/10.3389/frobt.2021.645956

Piva, M., Zhang, X., Noah, J. A., Chang, S. W. C., & Hirsch, J. (2017). Distributed neural activity patterns during human-to-human competition. *Frontiers in Human Neuroscience, 11*, 571. https://doi.org/10.3389/fnhum.2017.00571

Quaresima, V., & Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: A concise review. *Organizational Research Methods, 22*(1), 46–68. https://doi.org/10.1177/1094428116658959

Ramos Gameiro, R., Kaspar, K., König, S. U., Nordholt, S., & König, P. (2017). Exploration and exploitation in natural viewing behavior. *Scientific Reports, 7*(1), 2311. https://doi.org/10.1038/s41598-017-02526-1

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies, 151*, Article 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

Roesler, E., Manzey, D., & Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics, 6*(58), eabj5425. https://doi.org/10.1126/scirobotics.abj5425

Sarigul, B., Saltik, I., Hokelek, B., & Urgen, B. A. (2020). *Does the appearance of an agent affect how we perceive his/her voice? Companion of the 2020 ACM/IEEE international conference on human-robot interaction*.

Schreibelmayr, S., & Mara, M. (2022). Robot voices in daily life: Vocal human-likeness and application context as determinants of user acceptance. *Frontiers in Psychology, 13*, Article 787499. https://doi.org/10.3389/fpsyg.2022.787499

Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2021). Voice in human-agent interaction: A survey. *ACM Computing Surveys, 54*(4), 1–43. https://doi.org/10.1145/3386867

Singh, A. K., 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

Tatarian, K., Stower, R., Rudaz, D., Chamoux, M., Kappas, A., & Chetouani, M. (2021). How does modality matter? Investigating the synthesis and effects of multi-modal robot behavior on social intelligence. *International Journal of Social Robotics, 14*(4), 893–911. https://doi.org/10.1007/s12369-021-00839-w

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior, 38*, 75–84. https://doi.org/10.1016/j.chb.2014.05.014

Thepsoonthorn, C., Ogawa, K.-I., & Miyake, Y. (2021). The exploration of the uncanny valley from the viewpoint of the robot's nonverbal behaviour. *International Journal of Social Robotics, 13*(6), 1443–1455. https://doi.org/10.1007/s12369-020-00726-w

Tiberio, L., Cesta, A., & Olivetti Belardinelli, M. (2013). Psychophysiological methods to evaluate user's response in human robot interaction: A review and feasibility study. *Robotics, 2*(2), 92–121. https://doi.org/10.3390/robotics2020092

Tsiourti, C., Weiss, A., Wac, K., & Vincze, M. (2019). Multimodal integration of emotional signals from voice, body, and context: Effects of (In)Congruence on emotion recognition and attitudes towards robots. *International Journal of Social Robotics, 11*(4), 555–573. https://doi.org/10.1007/s12369-019-00524-z

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition, 12*(3), 473–494. https://doi.org/10.1080/13506280444000409

Vollmer, A.-L., Read, R., Trippas, D., & Belpaeme, T. (2018). Children conform, adults resist: A robot group induced peer pressure on normative social conformity. *Science Robotics, 3*(21). https://doi.org/10.1126/scirobotics.aat7111. Article UNSP eaat7111.

Walters, M. L., Syrdal, D. S., Koay, K. L., Dautenhahn, K., & Te Boekhorst, R. (2008). *Human approach distances to a mechanical-looking robot with different robot voice styles. RO-MAN 2008 - the 17th IEEE International Symposium on Robot and Human Interactive Communication*.

Wang, J., Antonenko, P., & Dawson, K. (2020). Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education, 146*, Article 103779. https://doi.org/10.1016/j.compedu.2019.103779

Wei, Y., & Zhao, J. (2016). Designing robot behavior in human robot interaction based on emotion expression. *Industrial Robot-an International Journal, 43*(4), 380–389. https://doi.org/10.1108/Ir-08-2015-0164

Wheeler, R. E., Davidson, R. J., & Tomarken, A. J. (2007). Frontal brain asymmetry and emotional reactivity: A biological substrate of affective style. *Psychophysiology, 30*(1), 82–89. https://doi.org/10.1111/j.1469-8986.1993.tb03207.x

Wiese, E., Abubshait, A., Azarian, B., & Blumberg, E. J. (2019). Brain stimulation to left prefrontal cortex modulates attentional orienting to gaze cues. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 374*(1771), Article 20180430. https://doi.org/10.1098/rstb.2018.0430

Winkle, K., Caleb-Solly, P., Turton, A., & Bremner, P. (2019). Mutual shaping in the design of socially assistive robots: A case study on social robots for therapy. *International Journal of Social Robotics, 12*(4), 847–866. https://doi.org/10.1007/s12369-019-00536-9

Wood, L. J., Dautenhahn, K., Rainer, A., Robins, B., Lehmann, H., & Syrdal, D. S. (2013). Robot-mediated interviews - how effective is a humanoid robot as a tool for interviewing young children? *PLoS One, 8*(3), Article e59448. https://doi.org/10.1371/journal.pone.0059448

Xu, K. (2019). First encounter with robot Alpha: How individual differences interact with vocal and kinetic cues in users' social responses. *New Media & Society, 21*(11–12), 2522–2547. https://doi.org/10.1177/1461444819851479

Yang, G. Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., et al. (2018). The grand challenges of Science Robotics. *Science Robotics, 3*(14), eaar7650. https://doi.org/10.1126/scirobotics.aar7650

Zhang, Y., Beskow, J., & Kjellström, H. (2017). Look but don't stare: Mutual gaze interaction in social robots. In *Social robotics* (pp. 556–566). Springer International Publishing. https://doi.org/10.1007/978-3-319-70022-9_55.

Zhang, H., Liu, M., Li, W., & Sommer, W. (2020a). Human voice attractiveness processing: Electrophysiological evidence. *Biological Psychology, 150*, Article 107827. https://doi.org/10.1016/j.biopsycho.2019.107827

Zhang, R., Saran, A., Liu, B., Zhu, Y., Guo, S., Niekum, S., et al. (2020b). Human gaze assisted artificial intelligence: A review. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*.

Zhang, D., Zhou, Y., & Yuan, J. (2018). Speech prosodies of different emotional categories activate different brain regions in adult cortex: An fNIRS study. *Scientific Reports, 8*(1), 218. https://doi.org/10.1038/s41598-017-18683-2