ACCSCIENCE PUBLISHING

Design+

ARTICLE

Children's preferences for virtual agents with natural and synthetic voices: Exploring developmental differences in willingness to interact and perceived voice quality

Miriam Veneziano[1]*, Maria Sarno[1], Maria Santina Ler[1], Marialucia Cuciniello[1], Terry Amorese[1], Attilio Trusio[2], Olga Gordeeva[3], Gennaro Cordasco[4], and Anna Esposito[1,5,6]*

[1]Department of Psychology, Faculty of Psychology 2, University of Campania "Luigi Vanvitelli," Caserta, Caserta, Campania, Italy

[2]State Education Department III Carlo Collodi District, Pagani, Salerno, Italy

[3]Acapela Group SA, Mons, Hainaut, Belgium

[4]Department of Computer Science, Faculty of Computer Science, University of Salerno, Fisciano, Salerno, Italy

[5]Italian Society for Neural Networks, Vietri Sul Mare, Salerno, Italy

[6]Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest, Hungary

***Corresponding authors:**
Miriam Veneziano
(miriam.veneziano@unicampania.it)
Anna Esposito
(anna.esposito@unicampania.it)

Citation: Veneziano M, Sarno M, Ler MS, *et al*. Children's preferences for virtual agents with natural and synthetic voices: Exploring developmental differences in willingness to interact and perceived voice quality. *Design+*. 2026;3(1):025370040. doi: 10.36922/DP025370040

Received: September 12, 2025

Revised: October 24, 2025

Accepted: January 19, 2026

Published online: March 27, 2026

Abstract

Investigations into children's perception and acceptance of new technologies, including voice chatbots, robots, and virtual agents, represent an emerging and rapidly expanding field of academic research. These interactive dialogue systems offer new opportunities in various settings (e.g., educational, therapeutic, and health settings) by providing innovative tools to improve the effectiveness of interactions and overall well-being. This study investigates children's preferences for eight virtual agents: four with natural human voices (one adult male, one adult female, one boy, and one girl) and four with synthetic voices matching the same age and gender distribution. A sample of 132 children aged between 4 and 10 years (excluding 6-year-olds) was divided into four age groups to assess whether preferences change with development. Results suggest that younger children showed no clear preference between natural and synthetic voices in terms of interaction. In contrast, older children displayed a stronger preference for natural voices, likely due to their engaging qualities. While no significant differences in voice quality were noted among younger children, older children favored natural adult voices, which they perceived as clearer and more trustworthy.

Keywords: Children; Virtual agents; Acceptance; Synthetic and natural voice

1. Introduction

From the earliest months of life, infants begin to distinguish between animate and inanimate objects based on a combination of visual, motor, and functional characteristics. This ability relies on cues such as movement, which is a key element in the distinction. Animate objects, such as people or animals, move in unpredictable and self-regulated ways, whereas the movement of inanimate objects is typically predictable and caused by external forces.[1] This distinction undergoes a gradual development during childhood, moving from an understanding based on perceptual features to a more complex one based on biological and intentional concepts.[2]

Neuropsychological studies have demonstrated that specific brain regions, such as the temporal lobe, are implicated in the categorization of animate and inanimate objects, suggesting distinct neural systems for these categories.[3] Language development is instrumental in facilitating children's categorization of objects and their understanding of the characteristics of living things, including the capacity to grow and reproduce. Between the ages of 4 and 10, this understanding evolves towards greater abstraction and sophistication, a process influenced by cognitive and social factors.[2] A growing body of research in human–computer interaction, human–robot interaction, and applied developmental psychology has explored how children perceive robotic and conversational agents.

These devices are often described as occupying a "hybrid" space, emulating animate qualities through features such as language, movement, and responsiveness.[1] Children may establish relationships with these agents that resemble those formed with people.[4] Melson *et al.*[4] have demonstrated that children, particularly in early development, tend to ascribe animistic properties, such as intentionality or emotion, to these devices, especially when these objects exhibit interactive behaviors.

In recent years, research has increasingly focused on how agent design features—such as voice tone, anthropomorphic embodiment, and interactivity—shape children's social and emotional attributions. Studies have reported that these cues can influence empathy, trust, and moral reasoning, suggesting that interactions between children and artificial agents reflect complex socio-cognitive processes.[6–8] For instance, children may engage with social robots by ascribing intentions and emotional states, and may perceive them as ontologically distinct from other technological objects, such as computers. However, as children mature and accumulate experience, they become more adept at distinguishing between biological and mechanical capabilities.[5,9] Studies have also shown that while children recognize the absence of a human brain in robots, they may still attribute autonomous behavior to them, as if the robots possess an internal mechanism that guides their actions.[10] Interaction with reactive technological objects may therefore influence how children categorize animate and inanimate entities, creating opportunities and challenges for cognitive and social development.[11]

Perception of a conversational agent is not solely based on voice and interaction but also encompasses visual and personified elements. Users may ascribe human or living qualities to conversational agents, reflecting a blending of animate and inanimate characteristics. Visual perception has been demonstrated to exert a substantial influence on the way users engage with these agents, indicating that an agent's visual manifestation can shape the relationship and interactions that develop.[12] Recent studies have further confirmed that the visual and multimodal design of artificial intelligence-based agents plays a crucial role in modulating users' engagement and perceived animacy, particularly in children. For example, research on embodied conversational agents highlights how anthropomorphic and emotional cues can enhance children's learning experiences and moral reflection during interaction.[13,14]

Children may also develop communication strategies that bear similarity to those employed in human interactions, underscoring the notion of "cross-species" learning and socialization. This prompts the question of how children differentiate between animate and inanimate objects when interacting with technological devices that possess intelligent and interactive capabilities. Some children may place these agents somewhere between technological artifacts and living beings.[15]

To better shed light on the tendency for children to attribute animistic characteristics to inanimate objects, it is instructive to consider the Theory of Mind (ToM), defined as the capacity to infer others' emotional and mental states. Developmental research suggests that children begin to understand false beliefs and others' mental states between ages 3 and 5, when they acquire the ability to attribute intentions, emotions, and thoughts to other people.[16] This ability, developed through social interaction, may help explain why children attribute animistic characteristics to technological devices such as robots and voice-based agents, and engage with them in ways that resemble interaction with living beings.[17] Despite being aware that a robot is not a human being, many children demonstrate empathy toward it, asking whether it is "okay" or "tired." Furthermore, evidence suggests that children may develop moral norms toward robots, exhibiting respect or discomfort in responses to poor treatment.[9] However, the extent to which ToM processes are truly engaged during interactions with artificial agents remains an open question. While existing studies[17,18]—mostly focused on social or humanoid robots—suggest that ToM-like reasoning may be activated when agents display contingent and socially meaningful behaviors, the evidence remains limited. It is not yet clear whether humans, and particularly children, develop a genuine ToM toward mechanical or virtual agents that simulate intelligence and sociality rather than possessing them.

These observations suggest that interactions with technological devices may influence children's social and emotional development, prompting them to reflect on moral and relational concepts in a similar way as in human interactions.

Furthermore, recent studies[19–21] have explored the use of conversational agents as tools for educational support, therapeutic intervention, and skill development in children. Technological devices have been investigated as tutors to enhance learning, as supportive tools in interventions for children on the autism spectrum, and as instruments to improve social, cognitive, and emotional abilities. These applications demonstrate the practical relevance of studying how children perceive and interact with these devices, highlighting the importance of understanding the developmental and social factors that influence these interactions.

The present study aims to investigate how perception and interaction with technological devices, such as robots and speech agents, develop during development. The research was conducted on children aged between 4 and 10 years (6 years excluded) old attending preschool and the first years of primary school to explore potential differences related to age and gender. The investigation focused on the transition from a concrete and immediate understanding, characteristic of younger children, to a more abstract and complex understanding, typical of older children. The study also examined how this transition affects the attribution of psychological and animistic qualities to technological devices.

This study constitutes a progression of a preliminary pilot study (see 22) that targeted only one age group (7 to 9 years old) sharing the same objective. The current study's expanded sample allowed testing of the hypothesis that interaction with virtual agents changes over time, highlighting potential variations due to cognitive and emotional development.

2. Materials and methods

This study used a similar approach to that of Veneziano *et al.*[22], with the exception of the participants.

2.1. Tool

A simplified and adapted version of the Virtual Agent Voice Acceptance Questionnaire (VAVAQ) was used to collect participants' preferences. This version was optimized specifically for children, using simple and more accessible language and a smaller number of items compared to the original. The original questionnaire was designed to assess the acceptance of virtual assistants by adults and seniors (for the original version of the questionnaire, see 23). Our questionnaire was administered individually in a quiet classroom, in Italian, by the experimenter, who read the questions aloud, repeating them when necessary to ensure comprehension.

The questionnaire is divided into six sections:

(i) Socio-demographic information: This section collects general data on participants. As part of the background questionnaire, children reported their familiarity with digital devices, including smartphones, tablets, voice assistants, and gaming consoles. Frequency of use was rated on a 3-point scale (1 = every day, 2 = often but not daily, 3 = never), and perceived ease of use on a 5-point scale (1 = very easy, 2 = easy, 3 = not sure, 4 = difficult, 5 = very difficult).

(ii) Willingness to interact (WI): This section assesses participants' interest in interacting with each agent using a single item per agent:

"On a scale of 1 (not at all) to 5 (very much), how much would you like to continue talking to [agent name] now that you have seen him/her?"

Response options: 1 = not at all, 2 = slightly, 3 = somewhat, 4 = quite a lot, 5 = very much.

(iii) Perception of the system: This section comprises two subsections, each with two items, rated on a 5-point Likert scale:

- Hedonic Qualities–Feeling (HQF):
  - (a) "I think talking to [agent name] could be fun."
  - (b) "I think talking to [agent name] could be scary."
- Useful (USE):
  - (a) "I think talking to [agent name] could teach me new things."
  - (b) "I think talking to [agent name] could be useless."

Response options: 1 = strongly disagree, 2 = disagree, 3 = I do not know, 4 = agree, 5 = strongly agree.

For "useless," scores were inverted so that higher values consistently reflected a more positive perception of the agent's voice.

(iv) Quality of voice (QoV): This section comprises two items per agent, rated on a 5-point Likert scale:

(a) "The voice of [agent name] is very clear and understandable."

(b) "[Agent name] has a strange way of speaking."

Response options: 1 = strongly disagree, 2 = disagree, 3 = I do not know, 4 = agree, 5 = strongly agree

For "strange way of speaking," scores were inverted so that higher values consistently reflected a more positive perception of the agent's voice.

(v) Perceived age: This section explores the influence of perceived age on participants' interest in interacting with each agent:

(a) "How old do you think [agent name] is?"

(b) "Does [agent name]'s age make you want to talk to him/her?"

Response options: Yes/No.

(c) "How old would you like [agent name] to be?"

Response options: Younger than me; the same age as me; older than me (if yes, please specify "about the same as my mom or dad, or about the same as my grandparents").

(vi) Tasks: This section includes three activities that explore participants' perceptions of the agent's suitability to assist with various tasks (school, homework, and games), rated on a 5-point scale:

"Below is a list of tasks that [agent name] could help you with. Please indicate how well you think he/she would perform each task."

Response options: 1 = not at all suitable, 2 = slightly suitable, 3 = neither slightly nor very suitable, 4 = quite suitable, 5 = very suitable.

2.2. Stimuli

Dynamic virtual agents were created within an augmented reality environment supported by the Android operating system. For this study, eight agents were used, equally balanced by age and gender. The natural-voice agents comprised two adults—a male (Mattia) and a female (Noemi)—and two children—a boy (Andrea) and a girl (Vittoria). The synthetic-voice agents followed the same age and gender distribution.

Each agent was introduced with the sentence: "Hello, my name is [agent name], would you like to play with me?" The sentence was presented in Italian, and the name corresponded to the specific agent's identity. The characteristics of the video stimuli, including agent type, age, gender, and voice condition, are summarized in Table 1.

The agents' voices were generated by Acapela Group, a company specializing in high-quality text-to-speech solutions, while the human voices were recorded using the free software Audacity 3.7.5 (http://www.audacityteam.org). All recordings were captured with the same parameters (22 kHz, 16-bit WAV) and normalized for loudness using Audacity's internal settings to ensure consistent playback levels across agents.

The fundamental frequencies (F_0) of the four analyzed natural voices were extracted using Parselmouth 0.4.7, a Python interface for Praat. The average, minimum, and maximum F_0 values were calculated for each file. The male child had an average F_0 frequency of 229.1 Hz, with minimum and maximum values of 193 Hz and 451.7 Hz, respectively. The female child had an average F_0 of 312.5 Hz, with minimum and maximum values of 151.2 Hz and 459.7 Hz, respectively, which are consistent with the higher pitch typically associated with a child's voice. The adult male had an average F_0 of 130.9 Hz, with minimum and maximum values of 92.7 Hz and 171.0 Hz, respectively, which are compatible with an adult male voice. The adult female had an average F_0 of 199.5 Hz, with minimum and maximum values of 109.2 Hz and 328.2 Hz, respectively, which are consistent with an adult female voice. These results confirm the expected differences in average F_0 and variability in minimum and maximum values between adult and child voices, and between male and female voices.

As for the synthetic voices, the average F_0 for the male child was 274.5 Hz, with minimum and maximum values of 218 Hz and 322.2 Hz, respectively, making it slightly higher-pitched than the natural voice. The synthetic voice of the female child had an average F_0 of 259.0 Hz, with minimum and maximum values of 123.9 Hz and 339.0 Hz, respectively. The average F_0 of the male adult's synthetic voice was 115.2 Hz, with minimum and maximum values of 81.6 Hz and 158.6 Hz, respectively. The female adult's synthetic voice had an average F_0 of 201.1 Hz, with minimum and maximum values of 153.4 Hz and 265.2 Hz, respectively. Overall, synthetic voices replicate the fundamental frequency characteristics of natural voices, with slight variations due to the synthesis process.

Speech rate was measured as the duration of each sentence from onset to offset. Sentences ranged from 4 to 6 s across agents, corresponding to 3.0–3.5 syllables per second. This pacing ensured that all agents—natural and synthetic—were presented with comparable temporal dynamics, maintaining ecological validity and avoiding artificial acceleration or slowing of speech. The recordings were made in a soundproof environment. The background noise ranged from −96.5 dBFS to −94 dBFS between voices, indicating very low background noise and excellent signal quality. Stimuli were delivered via MDR-ZX110 stereo on-ear headphones (Sony, Japan) connected to a laptop, with playback volume fixed at 80% of system output. Participants were seated about 50 cm from the computer screen (15″), allowing clear visual access to the agents while maintaining a comfortable posture. Adjustments to screen distance or headphone positioning were made as needed to accommodate individual children's requirements (e.g., visual acuity or comfort). The video clips, averaging around four seconds in length, showed the agents in full-body view, with lip movements synchronized to the audio.

2.3. Participants

The study involved a total of 132 children recruited from the State Education Department III Carlo Collodi District in Pagani, Salerno. The participants were divided into four groups according to the class they attended: 37 children from the preschool (21 boys, mean [M] age = 4.51 years; SD ±0.55), 35 from the second class (17 boys, M age = 7.14 years; SD = ±0.35), 30 from the third class (19 boys, M age = 8.13 years; SD = ±0.34) and another 30 from the fourth class (16 boys, M age = 9.13 years; SD = ±0.47).

Age groups were defined according to school grades to capture developmental changes in children's preferences over time. The nursery school group included children aged 4 and 5, excluding 6-year-olds due to their transitional status between nursery school and primary school. Similarly, children in the fifth year of primary school, who were transitioning to secondary school, were excluded. The subsequent groups corresponded to second, third, and fourth grades, representing ages 7, 8, and 9, respectively. Although the initial plan was to include children up to age 9, some fourth-class participants had already turned 10; these children were retained in the analyses to preserve the grouping based on school grade and ensure ecological validity. This approach allows the study to explore developmental trends while maintaining a clear link to educational stages.

2.4. Procedure

Following the acquisition of informed consent from parents and assent from children, participants were invited into a quiet room that had been specially established for the experiment. The experimenter proceeded to introduce himself and explain the planned activities, emphasizing that the children were at liberty to withdraw at any time. The children then donned headphones and initiated the experiment, which consisted of the observations of eight video clips featuring the virtual agents (Figure S1). The order of the video clips and the sections of the questionnaire were randomized individually using a Google-based application. The randomization of the sequence was adopted to prevent potential distortions related to fatigue, learning effects, or habituation to a specific sequence. Following the observation of each video, participants were administered the VAVAQ corresponding to the agent they had just observed.

Table 1. Details of agents

Agent name	Category	Gender	Voice source	Clip duration (s)	Motion	Identical phrase	Language/Locale
Mattia	Adult	Male	Natural (human recording)	4	Mouth only	"*Ciao, sono Mattia, ti va di giocare con me*?"	Italian
Noemi	Adult	Female	Natural (human recording)	4	Mouth only	"*Ciao, sono Noemi, ti va di giocare con me*?"	Italian
Andrea	Child	Male	Natural (human recording)	4	Mouth only	"*Ciao, sono Andrea, ti va di giocare con me*?"	Italian
Vittoria	Child	Female	Natural (human recording)	4	Mouth only	"*Ciao, sono Vittoria, ti va di giocare con me*?"	Italian
Mattia-TTS	Adult	Male	Synthetic (TTS)	4	Mouth only	"*Ciao, sono Mattia, ti va di giocare con me*?"	Italian
Noemi-TTS	Adult	Female	Synthetic (TTS)	4	Mouth only	"*Ciao, sono Noemi, ti va di giocare con me*?"	Italian
Andrea-TTS	Child	Male	Synthetic (TTS)	4	Mouth only	"*Ciao, sono Andrea, ti va di giocare con me*?"	Italian
Vittoria-TTS	Child	Female	Synthetic (TTS)	4	Mouth only	"*Ciao, sono Vittoria, ti va di giocare con me*?"	Italian

Abbreviation: TTS: Text-to-speech.

2.5. Statistical analysis

The present study employed repeated-measures ANOVA to investigate children's preferences for the virtual agents. Scores from the two sections of the questionnaire relevant to this study—WI and QoV—were analyzed independently. The analyses included the type of agent (8 agents: 4 human-voiced, 4 synthetic-voiced) as a within-subjects factor, and participant gender and grade as between-subjects factors. Age of the agent was not included as a separate factor, as it is implicitly represented by the agent type (child vs. adult). Mauchly's test was conducted to assess sphericity, and Greenhouse–Geisser corrections were applied where violations were detected. Post-hoc comparisons were performed using Tukey's test to examine pairwise differences between agent types. The significance level was set at $\alpha = 0.05$. Partial eta squared (ηp^2) was reported as a measure of effect size.

The WI section consisted of a single-item measure, so reliability analyses were not applicable. Descriptive statistics (M, SD, min, max) were reported for each agent across the entire sample (Table S1).

Reliability analyses were performed for QoV, which comprised two items: clarity and comprehensibility of speech and perception of speech strangeness. The latter was reverse-scored so that higher scores consistently reflect a more positive rating, yielding a moderate Cronbach's alpha coefficient for all human agents ($\alpha = 0.503$; Table S2). The two items were moderately correlated, supporting the consistency of the section.

Marginal means plots (Figures S2–S4) were generated to illustrate the effects of agent voice type (natural or synthetic) across grades (preschool, second, third, and fourth grades) for both WI and QoV. Estimated marginal descriptive tables (M ± SD; Tables S3–S11) were generated to illustrate the effects of agent type (voice, age, and gender) and participant grade (preschool, second, third, and fourth grades). These tables include values only for significant interactions, offering a detailed overview of the observed differences without interrupting the narrative flow. To facilitate interpretation, two acronyms were adopted to distinguish between the two voice types: NV for natural voice and SV for synthetic voice.

Descriptive results on children's technological experience and perceived ease of use of devices were also reported, providing context for their interaction with the agents and helping to interpret developmental differences across age groups.

3. Results

3.1. Frequency of use and perceived ease of technological devices

Participants reported daily use of technological devices, and their perceived ease of use was assessed across four age groups: preschool (kindergarten), second, third, and fourth grades. In the preschool group (Figure 1), smartphones were used every day by 56.76% of children, often by 32.43%, and never by 10.81%. Tablets were used daily by 54.05%, often by 18.92%, and never by 27.03%. Voice assistants were used daily by 21.62%, often by 24.32%, and never by 54.05%, while gaming consoles were used daily by 50%, often by 11.11%, and never by 38.89%. In terms of perceived ease of use (Figure 2), smartphones were rated very easy by 75%, easy by 5.56%, not sure by 2.78%, difficult by 16.67%, and very difficult by 0%. Tablets were rated very easy by 59.46%, easy by 18.92%, not sure by 13.51%, difficult by 5.41%, and very difficult by 2.70%. Voice assistants were rated very easy by 62.16%, easy by 8.10%, not sure by 5.40%, difficult by 5.40%, and very difficult by 18.91%, while gaming consoles were considered very easy by 40.54%, easy by 5.41%, not sure by 8.11%, difficult by 32.43%, and very difficult by 13.51%.

In the second-grade group (Figure 3), daily smartphone use was 17.14%, often use was 71.43%, and never use was 11.43%. Tablets were used daily by 11.43%, often by 28.57%, and never by 60.00%. Voice assistants were used daily by 2.86%, often by 57.14%, and never by 40.00%, while gaming consoles were used daily by 34.29%, often by 51.43%, and never by 14.29%. For perceived ease of use (Figure 4), smartphones were very easy for 28.57%, easy for 40.00%, not sure for 28.57%, and difficult for 2.86%. Tablets were very easy for 34.29%, easy for 28.57%, not sure for 34.29%, and difficult for 2.86%. Voice assistants were very easy for 74.29%, easy for 2.86%, not sure for 17.14%, and difficult for 5.71%. Gaming consoles were very easy for 28.57%, easy for 20.00%, not sure for 37.14%, difficult for 11.43%, and very difficult for 2.86%.

Among third-grade participants (Figure 5), smartphones were used daily by 10.00%, often by 73.33%, and never by 16.67%. Tablets had the lowest daily use (6.67%), were often used by 20.00%, and were never used by 73.33%. Voice assistants were used daily by 16.67%, often by 50.00%, and never by 33.33%, while gaming consoles were used daily by 36.67%, often by 53.33%, and never by 10.00%. Smartphones were perceived as very easy by 46.67%, easy by 30.00%, not sure by 16.67%, and difficult by 6.67% (Figure 6). Tablets were very easy for

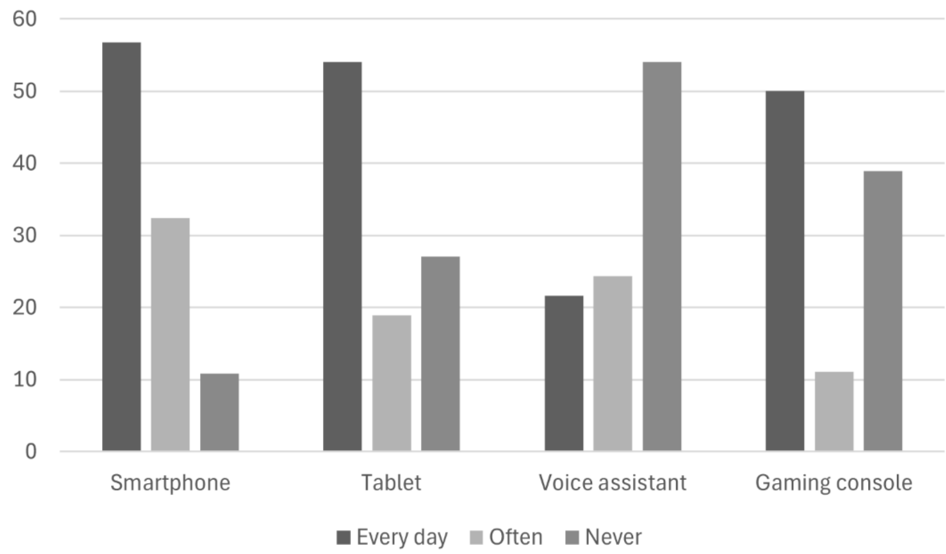

Figure 1. Frequency of use of technological devices reported by preschool participants

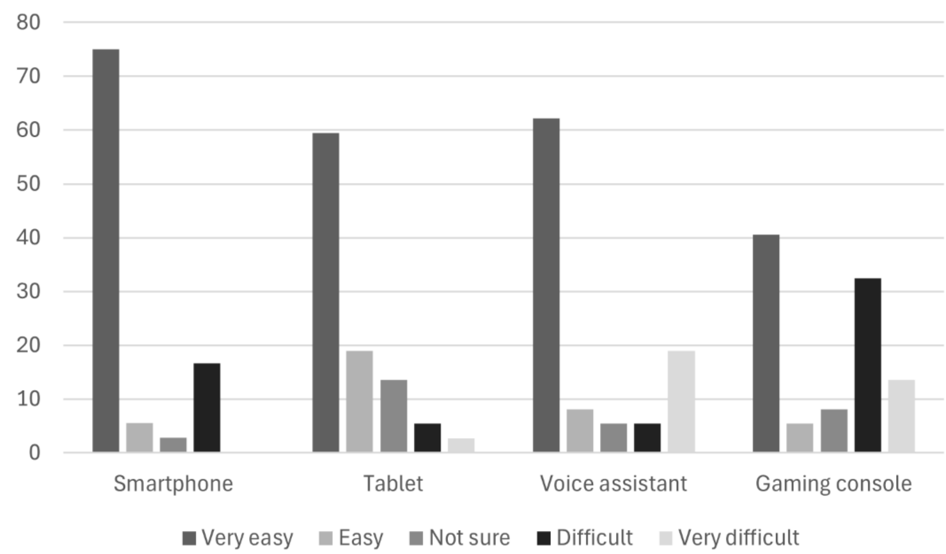

Figure 2. Perceived ease of use of technological devices reported by preschool participants

36.67%, easy for 33.33%, not sure for 20.00%, and difficult for 10.00%. Voice assistants were very easy for 63.33%, easy for 16.67%, not sure for 10.00%, and difficult for 10.00%. Gaming consoles were very easy for 17.95%, easy for 15.38%, not sure for 33.33%, and difficult for 33.33%.

In fourth-grade participants (Figure 7), smartphones were used daily by 31.03%, often by 62.07%, and never by 6.90%. Tablets were used daily by 16.67%, often by 30.00%, and never by 53.33%. Voice assistants were used daily by 20.00%, often by 26.67%, and never by 53.33%, whereas gaming consoles were used daily by 26.67%, often by 53.33%, and never by 20.00%. In terms of perceived ease of use (Figure 8), smartphones were very easy for 30.00%, easy for 43.33%, and not sure for 26.67%. Tablets were very easy for 30.00%, easy for 46.67%, and not sure for 23.33%. Voice assistants were very easy for 73.33%, easy for 16.67%, not sure for 6.67%, and difficult for 3.33%. Gaming consoles were very easy for 23.33%, easy for 40.00%, not sure for 30.00%, and difficult for 6.67%.

Overall, these results indicate that smartphones and

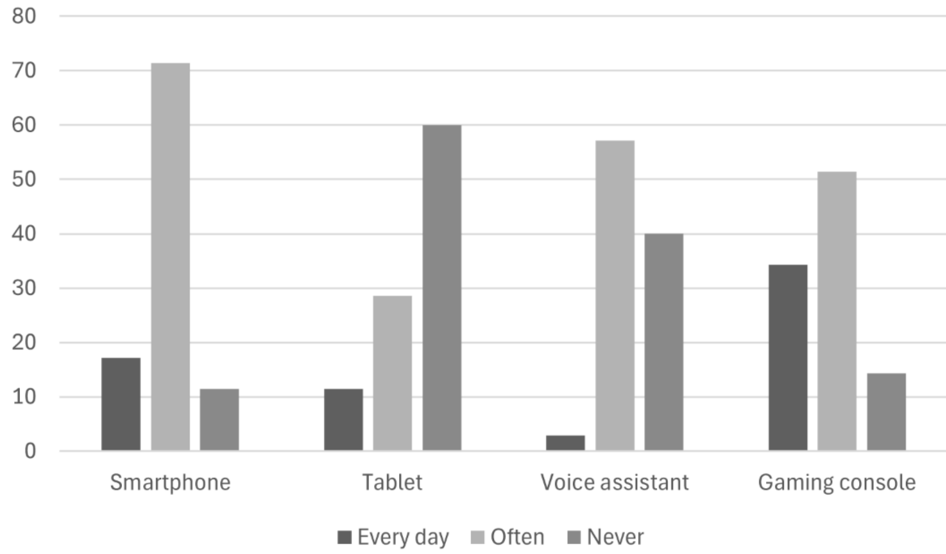

Figure 3. Frequency of use of technological devices reported by second-grade participants

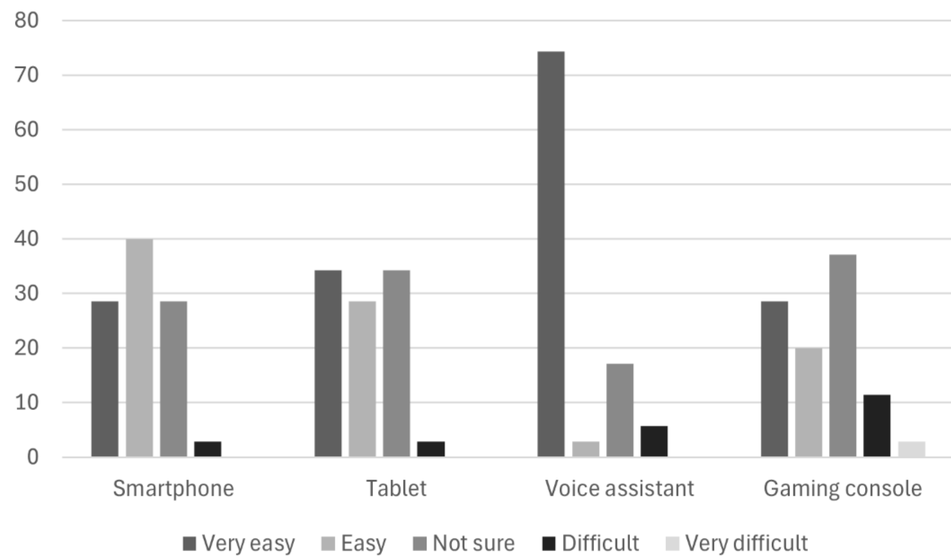

Figure 4. Perceived ease of use of technological devices reported by second-grade participants

gaming consoles were the most frequently used devices across age groups, whereas tablets and voice assistants showed lower daily usage, particularly among older participants. Voice assistants were consistently perceived as the easiest devices to use, while gaming consoles were generally rated as more challenging, especially among older participants.

3.2. Willingness to interact

Repeated measures ANOVA was conducted on the WI scores. The results, including main effects and interactions, are reported in Table 2. Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2[27] = 71.781$, $p < 0.001$, W = 0.554); therefore, Greenhouse–Geisser correction was applied ($\varepsilon = 0.873$).

The analysis revealed that a significant main effect of participants' gender was not observed ($F(1,124) = 0.598$, $p = 0.441$, $\eta p^2 = 0.005$). A significant main effect of participants' grade was observed ($F(3,124) = 3.474$, $p = 0.018$, $\eta p^2 = 0.078$). Tukey's post-hoc comparisons indicated that third-grade participants reported a significantly higher WI (M =

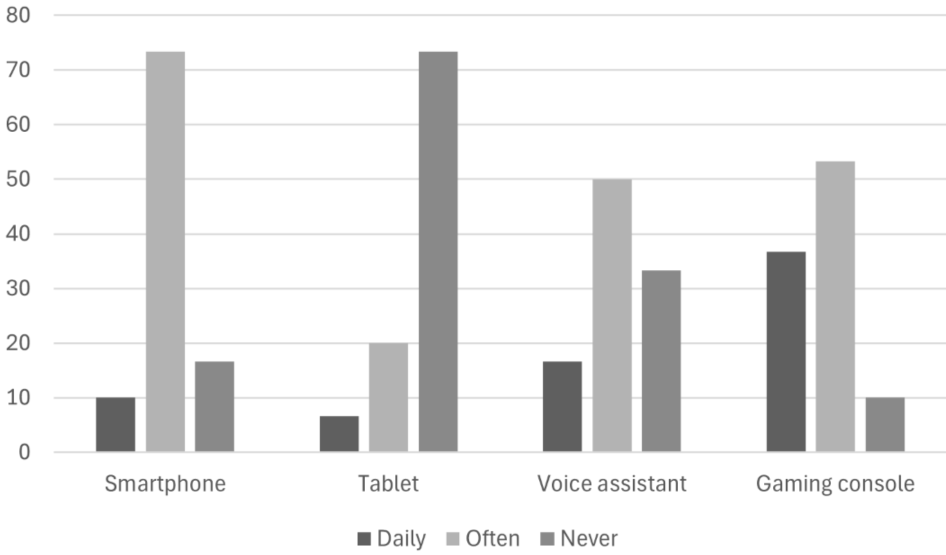

Figure 5. Frequency of use of technological devices reported by third-grade participants

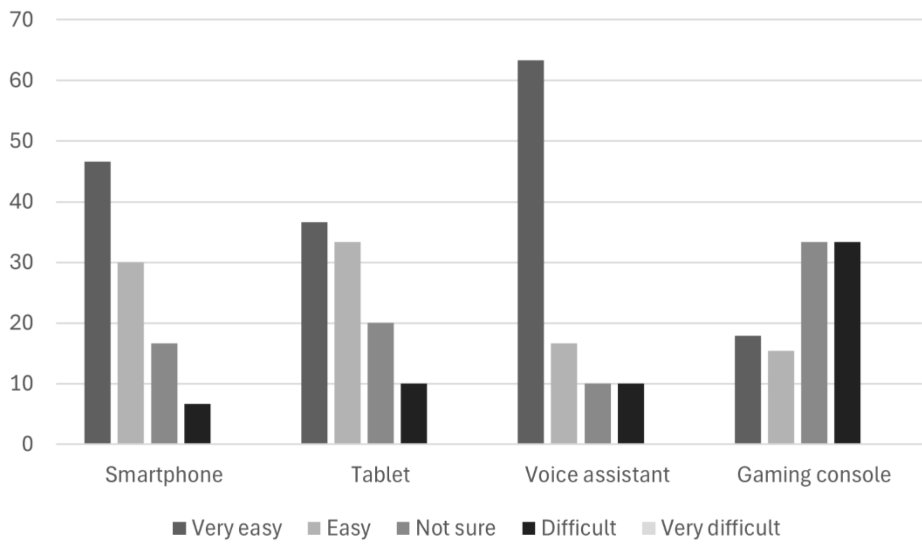

Figure 6. Perceived ease of use of technological devices reported by third-grade participants

4.534, SD = 0.150) than preschool participants (M = 3.890, SD = 0.132; p = 0.010).

A significant main effect of agent type emerged (F(6.11, 757.80) = 5.445, p < 0.001, ηp^2 = 0.042). Tukey's post-hoc comparisons for the main effect of agent type indicated that both adult NV agents and the female child NV agent were significantly preferred for WI compared to both adult SV agents. Specifically, the male adult NV agent (M = 4.370, SD = 0.104) scored higher on WI than the male (M = 3.820, SD = 0.139; p = 0.008) and female (M = 3.873, SD = 0.139; p = 0.045) adult SV agents. The female adult NV agent (M = 4.474, SD = 0.107) scored higher than the male (M = 3.820, SD = 0.139; p = 0.002) and female (M = 3.873, SD = 0.139; p = 0.003) adult SV agents. The female child NV agent (M = 4.389, SD = 0.107) reported higher WI scores than the male (M = 3.820, SD = 0.139; p = 0.008) and female (M = 3.873, SD = 0.139; p = 0.039) adult SV agents.

A significant interaction between agent type and participant gender was observed (F(6.11, 757.80) = 2.924, p = 0.008, ηp^2 = 0.023). Tukey's tests were performed for every factor (participants' gender and agent types),

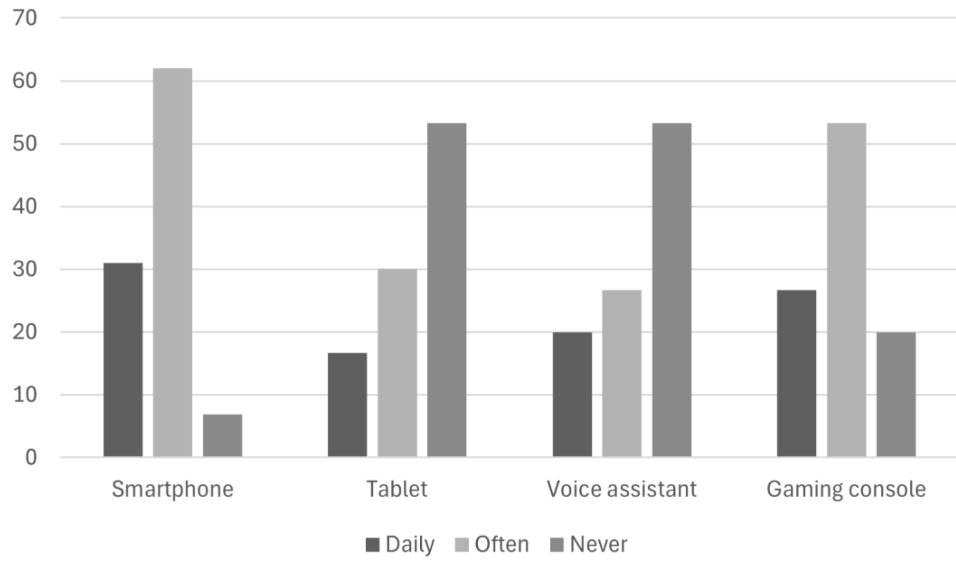

Figure 7. Frequency of use of technological devices reported by fourth-grade participants

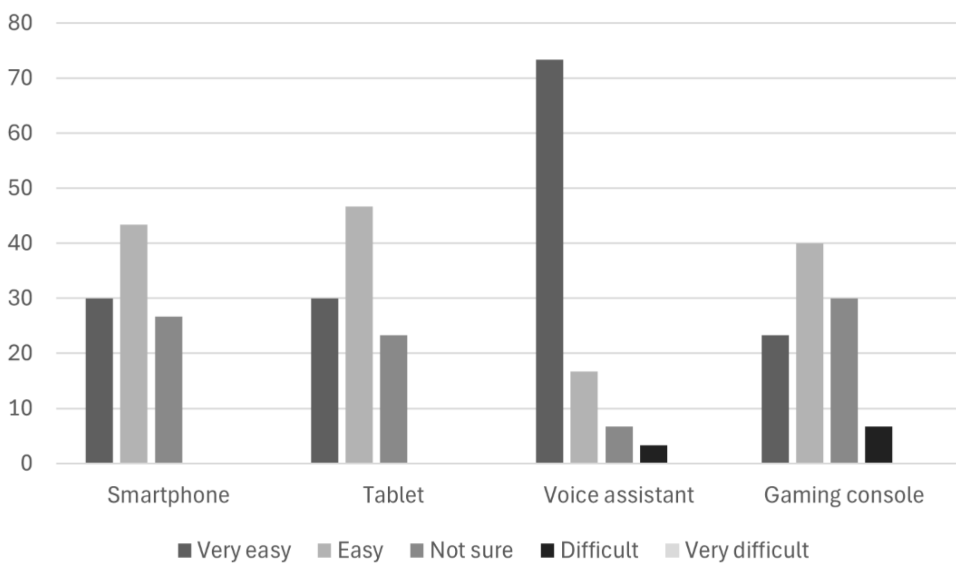

Figure 8. Perceived ease of use of technological devices reported by fourth-grade participants

revealing the following:

(i) Regarding participants' gender differences in attributing WI scores to the agent types, male participants (M = 4.573, SD = 0.160) showed more WI with male child NV than female participants (M = 4.004, SD = 0.180; p = 0.019). Similarly, male participants (M = 4.170, SD = 0.184) showed more WI with male adult SV than female participants (M = 3.470, SD = 0.207; p = 0.013).

(ii) Regarding the differences among agent types within each gender group in the assignment of WI scores, male participants showed more WI with male child NV (M = 4.573, SD = 0.160) than female adult SV (M = 3.844, SD = 0.185; p = 0.039). Female participants showed more WI with male adult NV (M = 4.331, SD = 0.156) than male adult SV (M = 3.470, SD = 0.207; p = 0.044). Furthermore, female adult NV (M = 4.625, SD = 0.159) was preferred over female adult SV (M =

3.902, SD = 0.208; $p = 0.043$), while female child NV (M = 4.391, SD = 0.159) reported higher scores than male adult SV (M = 3.470, SD = 0.207; $p = 0.002$).

No significant interaction was found between agent type and grade of participant ($F(18.334, 757.80) = 1.258$, $p = 0.207$, $\eta p^2 = 0.030$).

3.3. Quality of voice: Clear and understandable language

A repeated-measures ANOVA was conducted on the QoV (clear and understandable language) scores. Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2(27) = 67.586$, $p < 0.001$, W = 0.573); therefore, Greenhouse-Geisser correction was applied ($\varepsilon = 0.873$).

The analysis revealed that a significant main effect of participants' gender was not observed ($F(1,124) < 0.001$, $p = 0.985$, $\eta p^2 < 0.001$). A significant main effect of participants' grade was observed ($F(3,124) = 4.599$, $p = 0.004$, $\eta p^2 = 0.100$). Tukey's post-hoc comparisons indicated that third-grade participants (M = 4.476, SD = 0.142) reported significantly higher perceptions of clarity and comprehensibility than preschool participants (M = 3.932, SD = 0.124; $p = 0.028$) and fourth-grade participants (M = 3.892, SD = 0.137; $p = 0.022$).

A significant main effect of agent type emerged ($F(6.11, 758.04) = 6.306$, $p < 0.001$, $\eta p^2 = 0.048$). Tukey's post-hoc comparisons indicated a significant perception of clarity and comprehensibility for female adult NV and female child NV over male adult SV. Male child NV received higher scores than male adult SV, female adult SV, and male child SV. Specifically, female adult NV (M = 4.336, SD = 0.105) scored higher than male adult SV (M = 3.770, SD = 0.138; $p = 0.016$). Female child NV (M = 4.373, SD = 0.102) scored higher than male adult SV (M = 3.770, SD = 0.138; $p = 0.003$). Male child NV (M = 4.532, SD = 0.090) reported higher scores than male adult SV (M = 3.770, SD = 0.138; $p < 0.001$), female adult SV (M = 3.948, SD = 0.122; $p = 0.007$), and male child SV (M = 4.001, SD = 0.111; $p < 0.001$).

A significant interaction between agent type and participant grade was observed ($F(18.34, 758.04) = 2.900$, $p < 0.001$, $\eta p^2 = 0.066$). Tukey tests were performed for each factor (participants' grade and agent types), revealing the following:

(i) Regarding participants' grade differences, female adult NV received higher ratings from second-grade participants (M = 4.688, SD = 0.201) than preschool participants (M = 3.821, SD = 0.197; $p = 0.016$), and from third-grade participants (M = 4.947, SD = 0.225) than fourth-grade participants (M = 3.888, SD = 0.218; $p = 0.006$). Third-grade participants (M = 4.947, SD = 0.225) also rated this agent higher than preschool participants (M = 3.821, SD = 0.197; $p = 0.002$). For female adult SV, preschool participants (M = 4.354, SD = 0.228) rated it higher than fourth-grade participants (M = 3.022, SD = 0.251; $p = 0.001$), and second-grade participants (M = 4.221, SD = 0.232) and third-grade participants (M = 4.194, SD = 0.260) rated it higher than fourth-grade participants (M = 3.022, SD = 0.251; $p = 0.004$ and $p = 0.009$, respectively).

(ii) Regarding differences among agent types within each grade, in the preschool group, male child NV (M = 4.238, SD = 0.168) scored higher than male child SV (M = 3.406, SD = 0.208; $p = 0.018$). Female child NV (M = 4.262, SD = 0.191) scored higher than male child SV (M = 3.406, SD = 0.208; $p = 0.016$). Female adult SV (M = 4.354, SD = 0.228) scored higher than male child SV (M = 3.406, SD = 0.208; $p = 0.032$).

In the fourth-grade group, male child NV (M = 4.737, SD = 0.186) received higher scores than male adult NV (M = 3.790, SD = 0.251; $p = 0.027$), female adult NV (M = 3.888, SD = 0.218; $p = 0.032$), male adult SV (M = 3.170, SD = 0.286; $p < 0.001$), and female adult SV (M = 3.022, SD = 0.251; $p < 0.001$). Female child NV (M = 4.232, SD = 0.211) scored higher than both male adult SV (M = 3.170, SD = 0.286; $p = 0.027$) and female adult SV (M = 3.022, SD = 0.251; $p = 0.017$). Female child SV (M = 4.259, SD = 0.246) scored higher than male adult SV (M = 3.170, SD = 0.286; $p = 0.013$) and female adult SV (M = 3.022, SD = 0.251; $p = 0.002$). Male child SV (M = 4.036, SD = 0.230) scored higher than female adult SV (M = 3.022, SD = 0.251; $p = 0.002$).

No significant interaction was found between agent type and gender ($F(6.11, 758.04) = 1.900$, $p = 0.077$, $\eta p^2 = 0.015$).

3.4. Quality of voice: Strange way of speaking

After inverting the coding of the "strange way of speaking" element, higher scores indicate a more natural and less unusual vocal quality, while lower scores indicate a stranger vocal quality.

Repeated measures ANOVA was conducted on the QoV (strange way of speaking) scores. Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2(27) = 98.602$, $p < 0.01$, W = 0.444); therefore, Greenhouse-Geisser correction was applied ($\varepsilon = 0.833$).

The analysis revealed that a significant main effect of participants' gender was not observed ($F(1,124) = 2.779$, $p = 0.098$, $\eta p^2 = 0.022$). A significant main effect of participants' grade was observed (F(3,124) = 3.352, p = 0.021, ηp^2 = 0.075). Tukey's post-hoc comparisons indicated that third-grade participants (M = 4.299, SD = 0.167) reported significantly lower perceptions of strangeness than preschool participants (M = 3.642, SD = 0.146; p = 0.021).

A significant main effect of agent type emerged (F(5.82, 722.72) = 15.007, p < 0.001, ηp^2 = 0.108). Tukey's post-hoc comparisons for the main effect of agent type indicated that, compared to the female adult NV, male child NV, and female child NV, all other SV agents were perceived as strangers. Specifically, male adult NV (M = 4.164, SD = 0.127), female adult NV (M = 4.359, SD = 0.113), male child NV (M = 4.367, SD = 0.112), and female child NV (M = 4.595, SD = 0.096) scored higher than both male and female adult SV agents (male: M = 3.430, SD = 0.150; female: M = 3.490, SD = 0.146; p < 0.001 for all comparisons). Female adult NV (M = 4.359, SD = 0.113) also scored higher than male child SV (M = 3.758, SD = 0.147; p < 0.001) and female child SV (M = 3.761, SD = 0.148; p = 0.016). Male child NV (M = 4.367, SD = 0.112) scored higher than male child SV (M = 3.758, SD = 0.147; p = 0.014) and female child SV (M = 3.761, SD = 0.148; p = 0.012). Female child NV (M = 4.595, SD = 0.096) scored higher than male child SV (M = 3.758, SD = 0.147; p < 0.001) and female child SV (M = 3.7 61, SD = 0.148; p < 0.001).

A significant interaction between agent type and participant grade was observed (F(17.48, 722.72) = 3.334, p < 0.001, ηp^2 = 0.075). Tukey tests were performed for each factor (participants' grade and agent types), revealing the following:

(i) Regarding participants' grade differences, male adult NV received significantly higher ratings from second-grade participants (M = 4.690, SD = 0.242) and third-grade participants (M = 4.699, SD = 0.272) compared to preschool participants (M = 3.330, SD = 0.238; p = 0.001 for both comparisons). Female adult NV was rated higher by second-grade (M = 4.456, SD = 0.216), third-grade (M = 4.842, SD = 0.242), and fourth-grade participants (M = 4.603, SD = 0.234) compared to preschool participants (M = 3.536, SD = 0.212; p = 0.018 vs. second grade, p = 0.001 vs. third grade, and p = 0.018 vs. fourth grade).

Male child NV received higher ratings from third-grade participants (M = 4.694, SD = 0.240) than preschool participants (M = 3.664, SD = 0.210; p = 0.009), and from fourth-grade participants (M = 4.786, SD = 0.231) than preschool participants (p = 0.003). Female adult SV received higher ratings from preschool participants (M = 4.101, SD = 0.274) than from fourth-grade participants (M = 2.879, SD = 0.302; p = 0.020).

(ii) Regarding differences among agent types within each grade, in the preschool group, female child NV (M = 4.320, SD = 0.180) received higher ratings than male adult NV (M = 3.330, SD = 0.238; p = 0.006), female adult NV (M = 3.536, SD = 0.212; p = 0.013), and male child SV (M = 3.265, SD = 0.276; p = 0.004).

In the second-grade group, male adult NV (M = 4.690, SD = 0.242) received higher ratings than male adult SV (M = 3.618, SD = 0.286; p = 0.016).

In the third-grade group, male adult NV (M = 4.699, SD = 0.272), female adult NV (M = 4.842, SD = 0.242), male child NV (M = 4.694, SD = 0.240), and female child NV (M = 4.947, SD = 0.206) were rated higher than female adult SV (M = 3.304, SD = 0.312; p = 0.007 vs. both male adult and child NV, p = 0.001 vs. female adult NV, and p < 0.01 vs. female child NV). Female child NV was also rated higher than male adult SV (M = 3.818, SD = 0.320; p = 0.038).

In the fourth-grade group, male adult NV (M = 3.938, SD = 0.262) received higher ratings than male adult SV (M = 2.790, SD = 0.310; p = 0.018). Female adult NV (M = 4.603, SD = 0.234) received higher ratings than both male adult SV (M = 2.790, SD = 0.310; p <0.001) and female adult SV (M = 2.879, SD = 0.302; p < 0.001). Female child NV (M = 4.763, SD = 0.199) and male child NV (M = 4.786, SD = 0.231) were rated higher than male adult SV (M = 2.790, SD = 0.310; p <0.001 vs. female child NV and p < 0.001 vs. male child NV), female adult SV (M = 2.879, SD = 0.302; p < 0.001 for all comparisons), and female child SV (M = 3.652, SD = 0.306; p = 0.039 vs. male child NV and p = 0.013 vs. female child NV).

No significant interaction was found between agent type and participant gender, F(5.82, 722.72) = 1.624, p = 0.140, ηp^2 = 0.013.

4. Discussion

This study examined how children of different ages perceive and interact with voice assistants characterized by different vocal qualities. The results revealed two main aspects: the proclivity to interact and the clarity and intelligibility of the voice. These elements offer significant insights into children's cognitive and emotional development, as well as their expectations of speech technologies. While a preceding study[22] had only examined one age group, the current research confirms and extends the emerging evidence.

Willingness to interact appears to be primarily

Table 2. A summary of the repeated measures ANOVA results

Effect	df	*F*	*p*-value	ηp²
Willingness to interact				
Agent type	6.111, 757.80	5.445	<0.001	0.042
Participant grade	3, 124	3.474	0.018	0.078
Participant gender	1,124	0.598	0.441	0.005
Agent type × Participant gender	6.111, 757.80	2.924	0.008	0.023
Agent type × Participant grade	18.334, 757.80	1.258	0.207	0.030
Quality of voice: Clear and understandable language				
Agent type	6.111, 758.04	6.306	<0.001	0.048
Participant grade	3, 124	4.599	0.004	0.100
Participant gender	1, 124	<0.001	0.985	<0.001
Agent type × Participant gender	6.111, 758.04	1.900	0.077	0.015
Agent type × Participant grade	18.343, 758.04	2.900	<0.001	0.066
Quality of voice: Strange way of speaking				
Agent type	5.828, 722.72	15.007	<0.001	0.108
Participant grade	3, 124	3.352	0.021	0.075
Participant gender	1, 124	2.779	0.098	0.002
Agent type × Participant gender	5.828,722.72	1.624	0.140	0.013
Agent type × Participant grade	17.485, 722.72	3.334	<0.001	0.075

influenced by age and voice type, with no significant main effect of participant gender. Third-grade participants reported a greater WI than preschool participants, reflecting developmental progression in the ability to interact with technological agents. This age-related trend suggests that, as children grow older, they develop a greater interest in interacting with such agents. The consistent preference for agents with natural voices indicates that children find human-like voices more acceptable, understandable, and engaging than synthetic ones. These preferences are evident in both the desire to interact and the evaluation of voice clarity and strangeness, highlighting the importance of vocal naturalness in shaping children's responses. Interactions between agent type and participant gender further clarify these patterns. While gender had no significant effect, male participants were found to be more willing to interact with agents of the same gender. In contrast, female participants showed a marked preference for agents with natural voices, regardless of the agent's gender. These findings suggest that males may be more influenced by identification with peers or similar figures, whereas females may place greater importance on clarity, naturalness, and voice quality, evaluating the agent based on its communicative characteristics rather than its gender.

Therefore, children's preferences varied not only by gender but also by age. Older participants showed a clearer distinction in their evaluations of agent voices, regardless of the agent's age or gender. Overall, participants attributed greater clarity and less strangeness to natural voices, whereas synthetic voices were perceived as unfamiliar or strange, although this perception decreased with age. This is probably due to greater experience with digital voices and greater sensitivity to vocal differences. These results could be interpreted in various ways. As previously suggested in the context of ToM development, older children's enhanced capacity to attribute mental states and intentions to vocal agents may influence their evaluations of clarity and strangeness. However, the definitive link between ToM and preferences remains unknown.[16] In line with previous observations[24], children prefer agents that emit natural and socially engaging tones, which are often associated with human-like qualities and effective communication.

Additionally, the results may be consistent with broader theories of cognitive and social development. Piaget's[25] theory states that, during the preoperational stage (ages 2–7), children rely on direct perception and immediacy rather than abstract reasoning. As development progresses, however, more defined preferences emerge. In the concrete operational stage (ages 7–11), children acquire more advanced skills that lead them to appreciate the authenticity and naturalness of voices. Symbolic play and identification processes can also enhance positive interactions with synthetic voice agents in playful or emotionally salient contexts.

Furthermore, according to one theory, the attribution of psychological and animistic characteristics to voice agents changes with age.[5] Younger children tend to ascribe intentions and emotions to these agents, perceiving them as semi-living entities. This forms an intermediate ontological category, with these children considering voice agents as "hybrids" between living beings and technological objects. In contrast, older children, as demonstrated in the research by Severson and Carlson[26], come to recognize the artificial nature of these agents. This evolutionary transition is consistent with conceptual frameworks based on biological and intentional reasoning.[2] Previous studies[27] have shown that children aged 4 to 7 tend to attribute innate cognitive abilities to robots, while those aged 8 to 10 demonstrate greater technological awareness. This reflects a change in perspective that parallels the observed preference for more natural and realistic vocal characteristics.

Interestingly, third-grade participants stood out, showing significantly higher scores than fourth-grade participants in both WI and perception of voice clarity. This pattern may suggest that children at this age are particularly sensitive to vocal differences. Alternatively, the increased linguistic and social experience gained between ages 9 and 10 (fourth-grade participants) might make older children more accustomed to vocal variations, leading them to perceive differences without emphasizing them as younger children do.

5. Limitations

This study has several limitations that should be considered when interpreting the findings. First, the sample was recruited from a single Italian school, reducing its representativeness and limiting the generalizability of the conclusions to other educational contexts. Moreover, the cross-sectional nature of the research does not allow for the monitoring of changes in children's behaviors or perceptions over time, preventing causal inferences.

A further limitation concerns the absence of social and environmental variables, such as family context, which could have significantly affected the observed results. Although data on children's technological experience and perceived ease of use of digital devices were collected, these variables were not analyzed as predictors in the present study. Finally, some variables showed mean scores approaching the maximum of the scale, suggesting potential ceiling effects.

6. Conclusion

Despite the limitations, the present study provides unequivocal evidence that children's preferences and perceptions of voice agents are influenced by age, gender, and cognitive development. It has been demonstrated that older children, specifically those in third grade, exhibit heightened sensitivity about their propensity to engage in interaction, as well as their capacity to discern natural voices from synthetic ones. The present study found that male participants generally expressed a preference for agents of the same gender that was not influenced by voice quality. In contrast, female participants consistently demonstrated a preference for natural voices, irrespective of the gender of the agent. The perception of strangeness decreased with age, reflecting an increasing ability to adapt to vocal variations. The findings emphasize the necessity of designing voice agents that are age-appropriate, socially recognizable, and vocally natural, with the objective of enhancing engagement and comprehension in interactions between children and agents.

Several questions remain unresolved and thus require further investigation. For instance, it is crucial to investigate the developmental trajectories of children's perceptions of voice agents over time. Longitudinal research could reveal how preferences and interaction patterns evolve across different stages of development. Alternatively, consideration may be given to the way socio-cultural variations influence children's vocal perceptions and preferences. Cross-cultural studies have the potential to yield valuable insights.

Finally, the question arises as to whether integrating design features could enhance the adaptability of voice agents to varying stages of cognitive and emotional development. This could be a promising direction for future research. The potential benefits of such efforts include greater personalization and acceptance of voice agents, ultimately fostering more engaging interactions.

Acknowledgments

We would like to express our deep gratitude to the State Education Department III Carlo Collodi District and its headmaster and teachers for their indispensable support and cooperation during this project. We are particularly grateful to the referring teacher, Mariarosaria Pulsinelli, for her constant availability.

Funding

This research was funded by the EU Horizon 2020 program (grant no.: 101182965 [CRYSTAL]), the NextGenerationEU PNRR Mission 4 Component 2 Investment 1.1 (D.D 1409 del 14-09-2022), the PRIN 2022 under the IRRESPECTIVE project (Code P20222MYKE; CUP: B53D23025980001), and the PNRR–MUR under AI-PATTERNS FAIR Project (CUP: E63C22002150007).

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Anna Esposito, Gennaro Cordasco
Formal analysis: Miriam Veneziano
Investigation: Miriam Veneziano, Maria Sarno, Maria Santina Ler, Attilio Trusio
Methodology: Anna Esposito, Gennaro Cordasco, Olga Gordeeva
Writing – original draft: Miriam Veneziano
Writing – review & editing: Marialucia Cuciniello, Terry Amorese

Ethics approval and data availability

The present study was conducted in full compliance with ethical principles regarding privacy and data confidentiality. Participation was entirely voluntary, with parents providing written informed consent and children giving verbal assent prior to the study. The consent form detailed all aspects of privacy and data protection, in accordance with Italian and European legislation (GDPR, EU Regulation 2016/679; Legislative Decree No. 196/2003).

The experiment was approved by the Ethics Committee of the University of Campania "Luigi Vanvitelli" (protocol no. 17/2024, approved on April 9, 2024), and data were collected between April 29 and June 7, 2024.

Personal information (e.g., names and surnames) was used solely for consent documentation and was not included in the research dataset. Each participant was identified by an anonymized alphanumeric code (e.g., AB_F_01). Consent forms were securely stored at the school facility, while anonymized data were kept on a password-protected university server. All procedures adhered to the ethical principles of the Declaration of Helsinki.

Consent for publication

Informed consent for data publication was obtained from the children's parents or legal guardians prior to the study. All data were collected and analyzed in anonymized form, and only aggregated results are reported in the manuscript, ensuring that no individual participant can be identified.

Availability of data

The anonymized dataset and study materials are available from the corresponding author upon reasonable request.

References

1. Rakison DH, Poulin-Dubois D. Developmental origin of the animate–inanimate distinction. *Psychol Bull.* 2001;127(2):209-228.

   doi: 10.1037/0033-2909.127.2.209

2. Opfer JE, Gelman SA. Development of the animate-inanimate distinction. In: Zelazo PD, editor. *The Wiley-Blackwell Handbook of Childhood Cognitive Development*. 2nd ed. Hoboken, NJ, USA: Wiley-Blackwell; 2011:213-238.

   doi: 10.1002/9781444325485.ch8

3. Caramazza A, Shelton JR. Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *J Cogn Neurosci.* 1998;10(1):1-34.

   doi: 10.1162/089892998563752

4. Tanaka F, Cicourel A, Movellan JR. Socialization between toddlers and robots at an early childhood education center. *Proc Natl Acad Sci USA.* 2007;104(46):17954-17958.

   doi: 10.1073/pnas.0707769104

5. Melson GF, Kahn PH Jr, Beck A, *et al.* Children's behavior toward and understanding of robotic and living dogs. *J Appl Dev Psychol.* 2009;30(2):92-102.

   doi: 10.1016/j.appdev.2008.10.011

6. Kory-Westlund JM, Breazeal C. Exploring the effects of a social robot's speech entrainment and backstory on young children's emotion, rapport, relationship, and learning. *Front Robot AI.* 2019;6:54.

   doi: 10.3389/frobt.2019.00054

7. Roesler E. Anthropomorphic framing and failure comprehensibility influence different facets of trust towards industrial robots. *Front Robot AI.* 2023;10:1235017.

   doi: 10.3389/frobt.2023.1235017

8. Festerling J, Siraj I. Anthropomorphizing technology: A conceptual review of anthropomorphism research and how it relates to children's engagements with digital voice assistants. *Integr Psychol Behav Sci.* 2022;56(3):709-738.

   doi: 10.1007/s12124-021-09668-y

9. Kahn PH Jr, Kanda T, Ishiguro H, *et al.* "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Dev Psychol.* 2012;48(2):303-314.

doi: 10.1037/a0027033

10. van Duuren M, Scaife M. Because a robot's brain hasn't got a brain, it just controls itself—Children's attributions of brain-related behaviour to intelligent artefacts. *Eur J Psychol Educ.* 1996;11:365-376.

    doi: 10.1007/BF03173278

11. Mikropoulos TA, Misailidi P, Bonoti F. Attributing human properties to computer artifacts: Developmental changes in children's understanding of the animate-inanimate distinction. *Psychol J Hell Psychol Soc.* 2003;10(2):123-140.

    doi: 10.12681/psy_hps.23951

12. Lee S, Kim S, Lee S. What does your agent look like? A drawing study to understand users' perceived persona of conversational agent. In: Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. ACM; 2019:1-6.

    doi: 10.1145/3290607.3312796

13. Kory-Westlund JM. Implications of children's social, emotional, and relational interactions with robots for human–robot empathy. In: Smith M, Jones L, editors. *Conversations on Empathy: Interdisciplinary Perspectives on Empathy, Imagination and Othering*. London, UK: Routledge; 2023:256-278.

    doi: 10.4324/9781003189978-17

14. Xu Y, Warschauer M. What are you talking to? Understanding children's perceptions of conversational agents. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. ACM; 2020:1-13.

    doi: 10.1145/3313831.3376416

15. Wellman HM, Cross D, Watson J. Meta-analysis of theory-of-mind development: The truth about false belief. *Child Dev.* 2001;72(3):655-684.

    doi: 10.1111/1467-8624.00304

16. Banks J. Theory of mind in social robots: Replication of five established human tests. *Int J Soc Robot.* 2020;12(2):403-414.

    doi: 10.1007/s12369-019-00588-x

17. Mao Y, Liu S, Ni Q, Lin X, He L. A review on machine theory of mind. *IEEE Trans Comput Soc Syst.* 2024;11(6):7114-7132.

    doi: 10.1109/TCSS.2024.3416707

18. Mou W, Ruocco M, Zanatto D, Cangelosi A. When would you trust a robot? A study on trust and theory of mind in human-robot interactions. In: Proceedings of the 2020 IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE; 2020:956-962.

    doi: 10.1109/RO-MAN47096.2020.9223551

19. Xu Y, Aubele J, Vigil V, *et al*. Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. *Child Dev.* 2022;93(2):e149-e167.

    doi: 10.1111/cdev.13708

20. Hiniker A, Wang A, Tran J, *et al*. Can conversational agents change the way children talk to people? In: Proceedings of the 20th Annual ACM Interaction Design and Children Conference. ACM; 2021:338-349.

    doi: 10.1145/3459990.3460695

21. Catania F, Spitale M, Garzotto F. Conversational agents in therapeutic interventions for neurodevelopmental disorders: A survey. *ACM Comput Surv.* 2023;55(10):1-34.

    doi: 10.1145/3564269

22. Veneziano M, Cuciniello M, Amorese T, *et al*. How willingness to interact and intelligibility of the voice affect children's interaction with an interactive dialogue system: A pilot study. In: Proceedings of the 2024 IEEE International Conference on Cognitive Infocommunications. IEEE; 2024:173-180.

    doi: 10.1109/CogInfoCom63007.2024.10894720

23. Esposito A, Amorese T, Cuciniello M, *et al*. Elder user's attitude toward assistive virtual agents: The role of voice and gender. *J Ambient Intell Humaniz Comput.* 2021;12(4):4429-4436.

    doi: 10.1007/s12652-019-01423-x

24. Kory-Westlund JM, Breazeal C. Assessing children's perceptions and acceptance of a social robot. In: Proceedings of the 18th ACM International Conference on Interaction Design and Children. ACM; 2019:38-50.

    doi: 10.1145/3311927.3323143

25. Barrouillet P. Theories of cognitive development: From Piaget to today. *Dev Rev.* 2015;38:1-12.

    doi: 10.1016/j.dr.2015.07.004

26. Severson RL, Carlson SM. Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. *Neural Netw.* 2010;23(8-9):1099-1103.

    doi: 10.1016/j.neunet.2010.08.014

27. Druga S, Williams R, Breazeal C, Resnick M. Hey Google, is it OK if I eat you? Initial explorations in child-agent interaction. In: Proceedings of the 2017 Conference on Interaction Design and Children. ACM; 2017:595-600.

    doi: 10.1145/3078072.3084330