Do Your Expectations Match? A Mixed-Methods Study on the Association between a Robot's Voice and Appearance

Citation for the original published paper (version of record):
De Cet, M., Cvajner, M., Torre, I. et al (2024). Do Your Expectations Match? A Mixed-Methods Study on the Association between a Robot's Voice
and Appearance. Proceedings of the 6th Conference on ACM Conversational User Interfaces, CUI 2024: 1-11. http://dx.doi.org/10.1145/3640794.3665551

Do Your Expectations Match? A Mixed-Methods Study on the Association Between a Robot's Voice and Appearance

Martina De Cet
demart@chalmers.se
Chalmers University of Technology
Gothenburg, Sweden

Martina Cvajner
martina.cvajner@unitn.it
University of Trento
Rovereto, Italy

Ilaria Torre
ilariat@chalmers.se
Chalmers University of Technology
Gothenburg, Sweden

Mohammad Obaid
mobaid@chalmers.se
Chalmers University of Technology
Gothenburg, Sweden

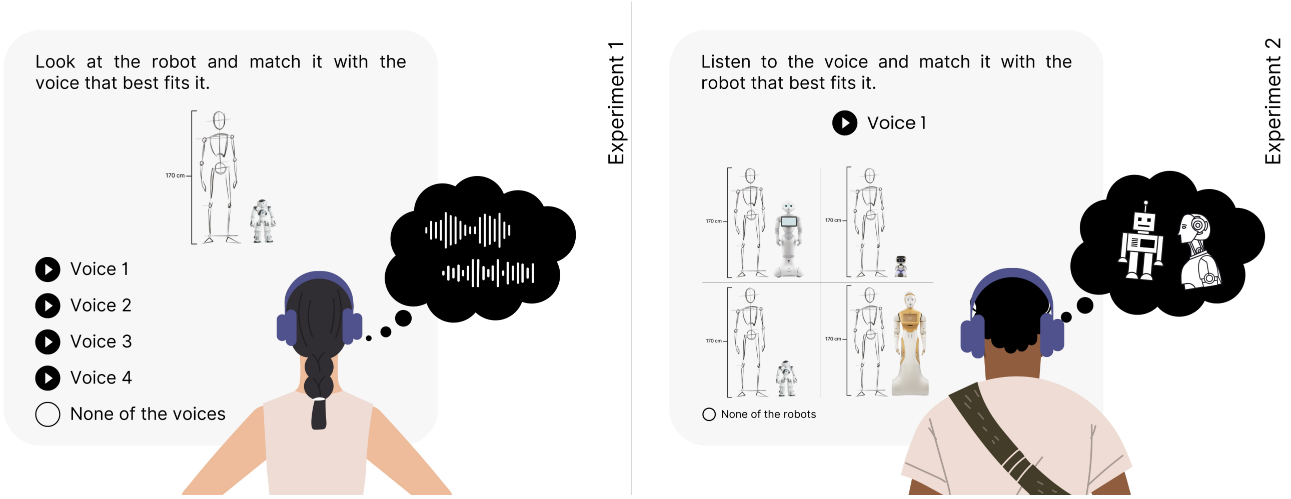
Look at the robot and match it with the voice that best fits it.
170 cm
Voice 1
Voice 2
Voice 3
Voice 4
None of the voices
Experiment 1
Listen to the voice and match it with the robot that best fits it.
Voice 1
170 cm
170 cm
170 cm
170 cm
None of the robots
Experiment 2

Figure 1: Visual representation of Experiment 1 and 2.

ABSTRACT

Both physical appearance and voice can elicit mental images of what someone and/or something should sound and look like. This is particularly relevant for human-robot interaction design and research since any voice can be added to a robot. Therefore, it is important to give robots voices that match users' expectations. In this paper, we examined the voice-appearance association by asking participants to match a robot image with a voice (Experiment 1, N = 24), and vice versa, a voice with a robot image (Experiment 2, N = 24), in two mixed-methods studies. We looked at participants' differences that could influence the voice-robot association (gender and nationality) and at voice and robot features that could influence participants' voice preferences (voice gender, pitch and robot's appearance). Results show that nationality influenced participants' association with a robot image after hearing its voice. Furthermore, a content analysis identified that when creating a voice mental image, participants looked at robots' gendered characteristics and height and they paid special attention to human-like and gender-specific cues in a voice when forming a mental image of a robot. Sociological differences also emerged, with Swedish participants suggesting the use of gender-neutral voices to avoid strengthening existing stereotypes, and Italians saying the opposite. Our work highlights the importance of individual differences in the robot voice-appearance association and the importance of involving the end user in designing the voice.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; *Empirical studies in HCI*; • **Applied computing** → *Psychology*.

KEYWORDS

Robot, Agent, Voice, Appearance

ACM Reference Format:
Martina De Cet, Martina Cvajner, Ilaria Torre, and Mohammad Obaid. 2024. Do Your Expectations Match? A Mixed-Methods Study on the Association

CUI '24, July 08–10, 2024, Luxembourg, Luxembourg
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0511-3/24/07
https://doi.org/10.1145/3640794.3665551

Between a Robot's Voice and Appearance. In *ACM Conversational User Interfaces 2024 (CUI '24), July 08–10, 2024, Luxembourg, Luxembourg*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3640794.3665551

1 INTRODUCTION

Modern society has begun to employ robots to perform an increasing number of activities, such as in education [1, 6], healthcare [8, 10, 15, 37], research, providing services and entertainment [11, 21, 31]. To be beneficial to human users, in many cases, robots should be able to communicate through spoken language, as this is the main communication method between humans, and thus it is essential in many Human-Robot Interaction (HRI) scenarios. Apart from linguistic content, the medium of spoken language communication, i.e. voice, also plays a role in how we perceive and accept information given to us by our interlocutors [24]. Despite this, robot designers and HRI practitioners tend to give robot voices barely any consideration [29]. For this reason, it is important to study in detail the vocal interaction between humans and robots to make this relationship as useful and fulfilling as possible.

Factors that influence users' perceptions of robot voices include pitch [32, 33, 36], whether physical robot characteristics match acoustic features of the voice [29, 40, 46], accent [3, 34, 47], and gender [13, 41]. Furthermore, existing literature in HRI suggests that factors such as human gender, cognitive style, or nationality might affect how people perceive robots [4, 26, 51].

To investigate this, we adopted the experimental paradigm used by McGinn and Torre [29], who asked people to listen to some voices (which differed in terms of gender, accent, and naturalness) and match them with the robot picture that best corresponded to them. In our current study, the main variables of interest were the robot voice's gender and pitch, the robot's appearance, and the participant's gender and nationality. Experiments were carried out in two different countries (Italy and Sweden). Participants were either asked to look at a robot image and match it with a voice from a set of voices, or the other way around (listen to a voice and choose the corresponding robot image from a set of images). Afterwards, participants were interviewed regarding which voice they would like a robot to have, with particular emphasis on any influence of the robot's gender and the context of interaction.

In this paper, we contribute with an investigation to examine 1) if people make consistent associations between robot images and voices, 2) if human-related factors (such as gender and nationality) influence the mental image people form of robots, 3) if voice-related factors (gender and pitch) influence participants' preferences for robot voices, and 4) if the stimulus presentation influences the voice-appearance association.

2 RELATED WORKS

Different acoustic features contribute to voice perception, such as volume, timbre, pitch, rhythm, articulation, fluency, and accent [35]. In human-robot interactions, for example, pitch has a significant influence on how people communicate and perceive a robot, but also on people's ability to retain information. Niculescu et al.[33] addressed the effect of voice pitch on the judgment of a female robot receptionist and found that the interaction quality was rated higher if the robot had a higher-pitched voice. Another study investigated how pitch and empathy/humour expression can influence the quality of interaction with a social robot receptionist [32]. The findings indicate that voice pitch had a significant impact on how users evaluated the quality of the whole interaction as well as the robot's attractiveness and general enjoyment. Furthermore, voice pitch appears to influence the memorisation of information. Pourfannan et al. [36] looked at how changing the robot's voice pitch and gender would impact individuals' ability to recall information in a noisy environment. A higher pitch was associated with considerably superior memory performance in the case of male voices. When a female voice was used, participants' memories were drastically improved by a lower pitch as opposed to a higher pitch.

The gender – of both the participant and the robot – can influence the perception of a robot. Siegel et al. [41] found that participants perceived the robot of the opposite gender to be more convincing, dependable, and interesting. On the other hand, Eyssel et al. [13] found that when participants perceived the robot as having their same gender, they held a more positive view of it and felt a stronger emotional bond. The same-gender robot was more highly anthropomorphized, but only if it spoke with a human-like voice.

While human-related factors such as gender have often been considered in previous research on robot voice, cultural or national origins are rarely investigated. Generally, it has been pointed out that people's nationality strongly influences how they perceive robots, the preconceptions they form and the attitudes they might have towards social robots [17, 26]. When it comes to robot voice perception, as far as we are aware, there are no studies that look at nationality as a possible influencing factor, yet a few studies have taken it into account when examining how people perceive robots in general [17, 26]. While a comprehensive cultural comparison is outside of the scope of the current work, we decided to recruit two sets of participants from two different cultural landscapes – Northern and Southern Europe – to gather at least initial evidence of any differences in robot voice perception.

By definition, robots are also physically embodied. Thus, the coherent design of a robot's appearance and voice is essential. Previous research has shown that a robot's physical appearance affects users' expectations of it [22]. According to Li et al. [25], personalising robots is fundamental to ensuring that a wide range of individuals can use them. Adaptability is a necessary design consideration for robots to meet the diverse needs of people. However, little research has been done to date on which robot voice matches which look. Mara et al. [27] were able to offer preliminary indications as to which physical characteristics are directly linked to humanlike voices (nose, hair, clothing) in contrast to less human-sounding voices (wheels). McGinn and Torre [29] investigated the mental images participants develop when they hear robots speak. Their findings reveal that even when the spoken words are incomprehensible, people still form a mental image of what the speaking robots look like, and suggest that giving a robot an inappropriate voice might negatively impact the interaction. Another study [46] investigated the missing link between deployment context, robot appearance and voice. Researchers asked participants to match a voice with a robot image in a specific context (e.g. a hospital or a school) and found out that people have an idea of what the robot's voice should sound like even before listening to the audio. They suggest that to form an impression of appropriateness, only two variables out of the three (voice features, context and robot features) are enough. The studies mentioned above have shown how people form expectations about the physical appearance of robots just by their voices. For this reason, appearance is another factor that needs to be taken into account when designing a robot voice [30, 48].

3 METHOD

Two experiments were conducted in Italy and Sweden to investigate robot-voice associations. To research this, we established the following research questions:

RQ1: Is there a link between voice and appearance in people's mental images of robots?

To answer this, two experiments were carried out. In Experiment 1, participants had to look at a picture of a robot and then later select which voice, if any, best matched the robot (see Figure 1). In Experiment 2, participants had to first listen to a voice and then connect it to one, or none, of four robot images (see Figure 1). Additionally, we also investigated whether participants created a mental image of a voice, after looking at a robot's image, or of a robot, after listening to a voice through a semi-structured interview performed at the end of the experiment. During the interview, we asked participants to reflect on the process behind their robot-voice/voice-robot associations during the experiments and to further comment on whether certain prominent robot characteristics (such as gender and human likeness) affected them.

We were also interested in understanding whether participant-related factors (participants' gender and nationality) and/or robot voice-related factors (gender and pitch) are involved in the robot-voice association, therefore we also asked:

RQ2: Do human-related factors (gender and nationality) influence the mental image that people form of a robot?

To answer this, we recruited an equal sample of men and women, Swedish and Italian participants.

Furthermore, we were interested in whether voice-related factors influence participants' preferences for robot voices, for this reason, the third research question arose:

RQ3: Do voice-related factors (gender and pitch) influence participants' preferences for robot voices?

We answered this by asking people to reflect on what characteristics they would like a robot voice to have during the post-experiment semi-structured interview.

Lastly, as one of our goals was to determine whether the stimulus presentation, image first or voice first, influenced the association done during the experiment, we formulated the following research question:

RQ4: Does the stimulus presentation influence the voice-robot association?

To answer this, we conducted two versions of our study (which we call Experiment 1 and Experiment 2), which present the stimuli in reverse order.

3.1 Stimuli preparation

As mentioned before, this study expands on the previous work of McGinn and Torre[29]. For this reason, with permission, the same audios from their experiment were adopted in this research. The sentences used to record the audios had already been tested as semantically neutral [39]. Since voice pitch and gender might be influencing factors when it comes to voice perception [13, 32, 33, 41], it was decided to have four versions for each of the sentences: 1) female high-pitched voice (Female High), 2) female low-pitched voice (Female Low), 3) male high-pitched voice (Male High), male low-pitched voice (Male Low). To do so, first, the recordings were normalized, which is the process of increasing the amplitude of a recording by a constant amount of gain to reach a norm decibel level. Then, the pitch of each audio was manipulated using the software Audacity [1]. The original average pitch was modified by +/- 1 semitone to have a higher pitch and lower pitch voice version for each audio files. For the female recordings, the original average pitch was 212 Hz (Female High = 224 Hz, Female Low = 203). For the male recordings, the original average pitch was 105 Hz (Male High = 113 Hz, Male Low = 100 Hz)[2]. For the robot image stimuli, we chose 4 of the most currently used robots in HRI studies: Misty, Nao, Pepper and Ari. As they vary in terms of size, human likeness, and gender attribution (see Figure 2), they provide a good spread of the characteristics that are linked to voice perception. As shown in the figure, a human-sized sketched mannequin was placed next to each image to give people an estimate of the size of the robot.

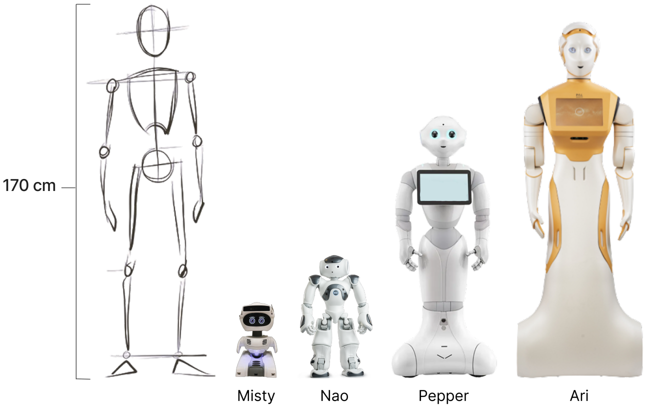

Figure 2: Robot images used in the experiments. A human-sized genderless sketched mannequin was placed next to each image to give people a size reference for the robots.

3.2 Participants

The experiments were conducted in English in Sweden and Italy. 48 adult participants were recruited for the two experiments. Participants were selected based on convenience sampling [42]. Participants were recruited via word of mouth and via advertisements on the campus of the Chalmers University of Technology, Sweden, and the University of Trento, Italy. In both experiments, participants were equally distributed by nationality and gender. Table 1 shows the distribution and demographics of participants.

[1] Software available here: https://audacityteam.org/

[2] See supplementary materials for the audios used during the experiments.

	Experiment 1 – image first		Experiment 2 – voice first	
Nationality	12 Swedish	12 Italians	12 Swedish	12 Italians
Gender Distribution	6 females, 6 males	6 females, 6 males	6 females, 6 males	6 females, 6 males
Age Distribution	mean = 25		mean = 25	
Level of Education	5 high school, 14 bachelor, 5 master		4 high school, 19 bachelor, 1 master	
Employment Status	20 students, 4 employees		24 students	

Table 1: Participants' demographics in Experiments 1 and 2.

3.3 Procedure

Participants were first welcomed into a quiet room, informed about the structure of the study and asked to sign a consent form[3]. Demographics were gathered: nationality, age, gender, level of education and employment status.

In Experiment 1 (see Figure 1), the participant was asked to observe the image of a robot and tell the experimenter when they were ready to listen to the four different voices. After they listened to the four audios, they could listen to the voices again. When the participant concluded the listening part, they had to choose the voice that best matched the image of the robot they were presented with. If the participant was not convinced by any of the audios, it was possible to click on the *none of the above* option. Since there were four robots under research, the procedure was repeated four times. To avoid any order effects, the robots' images were presented using a counterbalanced measures design, and the order of the voices was randomized.

In Experiment 2 (see Figure 1), the procedure was reversed: the participant first listened to a voice as many times as requested. When they finished the listening part, they were allowed to click on *next* to see the images of the four robots and select the image that most fit that voice (with an option to say that none of them were suitable). Since there were four voices, the procedure was repeated four times. The audios were presented using a counterbalanced measures design and the images of the robots were randomized.

Once this initial phase was concluded, participants of both experiments were invited to a semi-structured interview to better comprehend their thinking process during the experiment and to investigate their opinions about the robot voices and the interaction they would like to have with a robot. To do so, we asked questions to investigate 1) participants' thinking process during the robot-voice/voice-robot association, 2) what kind of voice they would like a robot to have, 3) if they thought robots could have a gender, 4) the interaction they would have with their robot if they had one and 5) how they would communicate with their robot [4].

The total duration of the experiment was approximately 20 minutes and about 15 minutes were for the interview.

3.4 Data Analysis

The data gathered through the experiments were analysed using a mixed-methods approach [19]. For the quantitative analysis, we first compiled contingency tables for how often a certain voice was selected upon seeing a robot image (Experiment 1, see Figure 3), and how often a robot was selected upon hearing a voice (Experiment 2, see Figure 4). Independent variables were robot appearance (Misty, Nao, Pepper, Ari), robot voice pitch (high and low), robot voice gender (female and male), participant nationality (Italian and Swedish) and participant gender (female and male). Chi-square tests were then carried out to find out if there was a significant relation between the robot-voice/voice-robot association, and participants' nationality or gender. Since the sample size was limited, we performed the chi-square tests twice for each experiment, first collapsing participants of the same nationality together (regardless of gender) and then collapsing across gender (regardless of nationality). Thus, we set the significance level to $\alpha = .025$ (.05/2, Bonferroni correction) to control the probability of Type I errors.

Subsequently, for the qualitative analysis, a content and thematic analysis were performed. At first, all the interviews were text-transcribed using clean verbatim transcription methods [49]. This method produces a transcript that is clearer and easier to read by eliminating filler words, stutters, and false beginnings while yet accurately recording every word said. Thereafter, two researchers decided on two main themes, *voice* and *interaction*, based on the topic investigated in the semi-structured interview, which encompass participants' preferences for robots' voices and their interactions with them. The transcriptions were coded by one researcher. The answers gathered through the first question, *can you describe to me your thinking process during the experiment?*, were analysed using content analysis. For the other questions, a thematic analysis was performed using an inductive approach. Therefore, a discussion and iteration phase started between the two researchers and the subthemes for each of the themes were generated. For the voice theme, the subthemes were: *human*, *prosody*, *gender*, *purpose*, and for the interaction theme: *gender*, *purpose*, *communication*, *support* and *privacy*. Furthermore, the data gathered through the second question of the semi-structured interview (which asks specifically about participants' thinking processes) were used to explain the significant result found through the chi-square test.

4 RESULTS

4.1 Experiment 1

Figure 3 shows the answers given by participants during Experiment 1. None of the chi-square tests were significant, thus we do not find evidence of a systematic pattern of matching a robot image to a voice.

Regarding the qualitative data collected through the semi structured interview, no patterns from participants were found based on their gender, but systematic differences emerged based on their nationality, which are highlighted in Table 2.

[3] The study was approved by the Ethics Committee of the University of Trento with code 2023-032.

[4] See supplementary materials for the questions asked during the semi-structured interview.

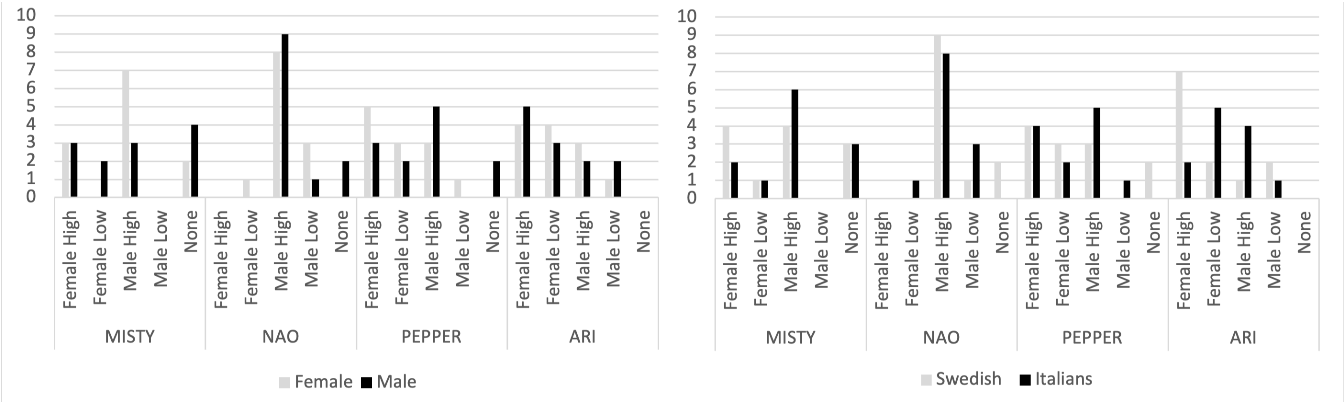

Figure 3: Experiment 1 - On the left, the robot-voice association is divided by gender, on the right, the robot-voice association is divided by nationality.

From the content analysis, it appears that participants from both nationalities, when looking at the robot, were searching for human attributes (face, eyes, mouth, nose, shoulders, hair, hips, clothing, body shape and height), trying to trace those details back to gender, thus already getting an initial idea of what that robot's voice might sound like. They tried to match their understanding of things in humans to robots. Afterwards, they focused on the robot's height to figure out the robot's possible age and thus the pitch of its voice. They reported that they were looking at robot images like they usually observe people.

Moving on to the thematic analysis, participants' statements for each subthemes are presented in Table 2. For the voice theme, the most recurring topics that participants brought up were:

1. Differentiation from humans: Participants reported preferring non-human voices to be able to distinguish robots from humans.
2. Voice prosody: Italian Participants (IP) and Swedish Participants (SP) had differing preferences for monotone or expressive voices.
3. Gender association: IP appreciated gender association with voices, while SP did not.
4. Voice relevance to robot's purpose: SP believed the voice should align with the robot's purpose.

Regarding the interaction theme:

1. Role of gender association: IP suggested gender association could aid interaction with robots.
2. Interaction dependent on purpose: SP believed interaction style should match the robot's purpose.
3. Communication length preferences: SP preferred shorter interactions, while IP preferred longer ones.
4. Importance of robot support: IP highly valued support from robots during interactions.
5. Privacy concerns: SP expressed worries about privacy in interactions with robots.

4.2 Experiment 2

In Experiment 2, we did not find any significant effects for the contingency tables divided by participant gender (see Figure 4), but

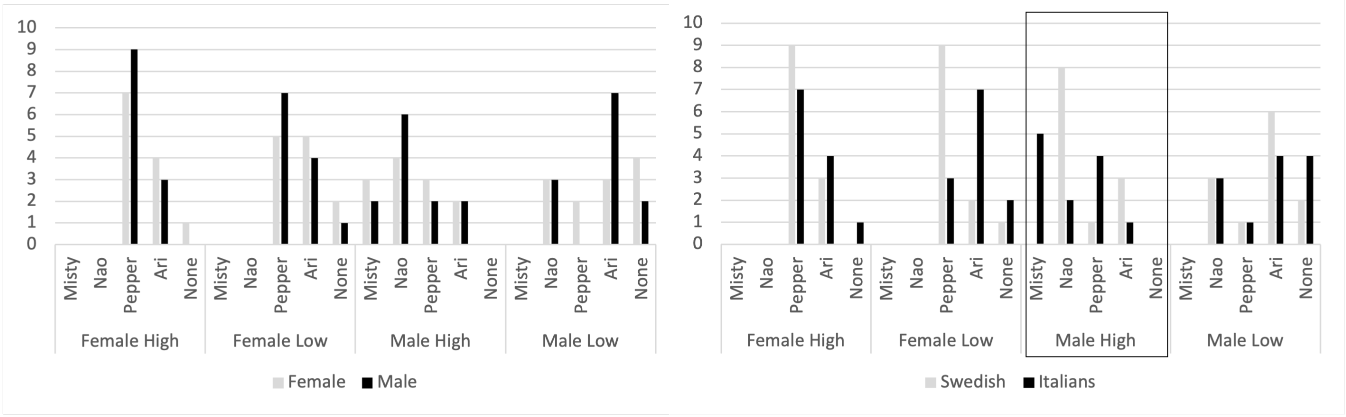

Figure 4: Experiment 2 - On the left, the voice-robot association is divided by gender, on the right, the voice-robot association is divided by nationality.

Themes	Sub-themes	Participants' statements
Voice	Human	**IP** and **SP** reported that since robots are supposed to live with them if they have a robotic voice, instead of a human one, it is easier to recognize who is talking. A human voice is scary, and the main reason is the idea that robots may be increasingly like humans and they are going to replace them. However, both preferred specific characteristics of the voice (calm, gentle and friendly) to make the user feel comfortable and not threatened.
	Prosody	**SP** commented that a monotone voice would be more appropriate for a robot. **IP** reported that expressive voice is fundamental for effective conversations.
	Gender	**IP** reported that when it comes to voice assistants, they are more used to listening to a female voice and that might influence them. According to them, it is perceived as kinder and more pleasant to hear and to be helped by. **SP** participants spoke out against the use of female voices for cleaning robots, as this reinforces existing gender stereotypes.
	Purpose	**SP** voice preferences change based on which area the robot is helping with and on the context; in a situation of emergency, a higher command voice would be preferable. No patterns emerged from **IP's** answers.
Interaction	Gender	**IP** said that assigning gender to a robot can help people feel more comfortable. No patterns emerged from **SP's** answers.
	Purpose	**SP** said the interaction depends on the robot's purpose: if it is a social companion, answering back is preferable, a cleaning assistant, does not have to answer back. No patterns emerged from **IP's** answers.
	Communication	**SP** reported they would prefer to have short conversations with the robot. According to them, robots should not talk on all occasions, they could use short answers, such as *yes* or *no*. **IP** said they would like more of a combination of talking and gestures.
	Support	**IP** said that, in the case of a problem, the robot should support the user by rationalising it, and by helping the person understand themselves and their emotions. They see robots as assistants, who help in daily activities. They should help with cleaning the house and managing the day when asked for suggestions or brainstorming about specific topics. No patterns emerged from **SP's** answers.
	Privacy	**SP** mentioned the fear that companies could violate users' privacy. There is a concern that robots will just mimic humans and their emotions. No patterns emerged from **IP's** answers.

Table 2: Themes and sub-themes found through the thematic analysis in Experiment 1.

we found an effect in the nationality ones. Specifically, we found that NAO was selected significantly more often for the Male High voice by Swedish than by Italians ($X^2(3, N = 24) = 11.400, p = .001$), as shown in Figure 4.

For the qualitative analysis, in particular, through the thematic analysis, we did not observe any gender-specific trends among the participants; however, systematic variations did arise concerning participants' nationality, which is why we have highlighted this element in Table 3 below.

From the content analysis, it emerged that at first, both Swedish and Italian participants tried to derive information about the robot's height from the pitch of the voice and the general appearance of the robot from the gender of the voice. Furthermore, the pitch was used to imagine the possible age of the person who was talking. Indeed, they used prior experience with humans to help them associate the voice with an image.

Moving to the thematic analysis, participants' statements for each subtheme are presented in Table 3. The voice theme is summarized as follows:

(1) Preferred human-like voice: Italian Participants (IP) and Swedish Participants (SP) preferred a slightly human-like voice.
(2) Voice prosody: Swedish Participants (SP) found expressive voices more suitable for robots.
(3) Gender association: both nationalities leaned towards a gendered voice, but SP shifted more towards preferring a "genderless"[5] voice.
(4) Voice adaptation to purpose: Both SP and IP agreed that a robot's voice should align with its purpose.

Concerning the interaction theme:

(1) Gender influence on interaction: IP believed assigning a gender to a robot might enhance human interaction.
(2) Interaction variation based on purpose: IP stated they would interact differently with a robot depending on its purpose.
(3) Communication length preferences: IP preferred longer conversations with robots, while SP preferred shorter interactions.
(4) Role of robot support: IP emphasized the importance of robots providing support to humans.
(5) Privacy concerns: Both IP and SP expressed concerns about privacy regulations regarding interactions with robots.

[5] There is a whole conversation to be had about whether something can be classified as "genderless", but it goes beyond the scope of this paper. Here, we are just reporting quotes from our participants.

Themes	Sub-themes	Participants' statements
Voice	Human	**IP** and **SP** reported preferring something more human-like, but not exactly like a human voice; they find human voices on robots *scary* and *creepy*. They prefer features that make the voice pleasant to listen to, to feel that the robot is neither judging nor threatening.
	Prosody	**SP** preferred an expressive voice. If the robot says something important with a monotonous tone, it would be boring and would not encourage the user to get in contact with it. No patterns emerged from **IP's** answers.
	Gender	**IP** and **SP** said they would feel safer with a female voice, saying that a female voice is softer and brighter in voice tones. They also reported that they usually pay more attention to female voices. On the other hand, **SP** proposed to have a genderless voice: according to them, gender should not matter, the most important factor is that it is easy to understand and not annoying.
	Purpose	According to **SP**, a robot's voice depends on its purpose. They proposed to change voice based on which area the robot is helping with and on the context. **IP** preferred a human voice if the robot is a companionship but a robotic one for a medical robot. Furthermore, voice preference seems to be related to the robot's general appearance: if a robot has muscles and is tall, then a more masculine deep voice is preferable.
Interaction	Gender	**IP** prefer robots that fit people's mental models. They reported that since women are the ones who usually take care of people, in the case of a healthcare robot, they would prefer it to have a female aspect. They said that it is difficult to think about something genderless since their language is gender-oriented and robots are associated with males. No patterns emerged from **SP's** answers.
	Purpose	For **IP**, a robot that helps with cleaning does not need to have the same conversational skills that a healthcare robot should have. In some cases, it is acceptable if it behaves more naturally (i.e. more like a human being), while in others it is not. No patterns emerged from **SP's** answers.
	Communication	**SP** preferred short conversations and sounds suggesting that the robot could use sounds such as the classic *leap loop* that can be associated with *yes* or *no* answers. **IP** participants reported their willingness to talk to the robot in full sentences.
	Support	**IP** see robots as cleaning assistants but they want them to answer humanely. They think of the robot as a friend and want to have constructive feedback from it. No patterns emerged from **SP's** answers.
	Privacy	**IP** and **SP** nationalities have the feeling of being spied on by the robot companies.

Table 3: Themes and sub-themes found through the thematic analysis in Experiment 2.

5 DISCUSSION

With the studies presented here, we looked at whether people form a mental image of what a robot should sound and/or look like upon seeing its image/hearing its voice.

5.1 Matching robot voices and images

We found that there seems to be a link between voice and appearance in people's mental images of robots (RQ1). Based on Experiments 1 and 2, we found that: 1) simply looking at a robot's image gives people an idea of what a robot should sound like and 2) hearing a voice gives people an idea of what a robot should look like. The first result is a novel finding. Indeed, participants of Experiment 1, where they first saw a picture of a robot, reported during the interviews that while they were looking at a robot's image they were trying to match the appearance to some voice characteristics. From the content analysis of Experiment 1, it appears that participants initially searched for familiar human attributes, such as face, eyes, nose, hips, and size. Indeed, as found by Fussell et al. [14], robots are anthropomorphised when they have human characteristics. In addition, participants associated a gender with the robot based on physical characteristics that can be traced back to humans [9]. These attributes were used to determine the gender of the robot and then to choose a more feminine or masculine voice. Furthermore, people focused on the robot's height to figure out the age of the robot and thus the possible pitch of its voice.

Regarding the second result, in Experiment 2, while participants were listening to the voice, they were already thinking of how the robot should look. This finding is in agreement with what McGinn and Torre [29] found. From the content analysis it was possible to understand that voice gender was used for imagining the possible appearance of the robot and the voice pitch was fundamental for understanding the robot's height. Even though in speech studies there is not always a relationship between voice pitch and observed speaker's height [16], Yilmazyildiz et al. [52], found that the pitch is inversely proportional to the robot's height: a higher pitch corresponds to a smaller height, while a lower pitch corresponds to a bigger height.

5.2 Human-related factors influencing the mental images of a robot

Participant gender does not seem to play a role in whether certain voices are more suitable for certain robot images, or vice versa. In fact, the quantitative analyses (in the form of chi-square tests) did not reveal any significant associations in this regard, and no gender patterns emerged from the thematic analysis of the post-experiment interviews. This result is at odds with other studies which have instead demonstrated how the gender of the participant influences the perception one has of robots [13, 41]. A possible reason that can explain why gender seems to not have influenced participants in the current experiments might be that the sample size was insufficient to detect any existing underlying differences.

However, nationality-based differences emerged. From the thematic analysis of Experiment 1, Italian participants seemed to be more likely to prefer a voice with human characteristics, such as expressive prosody, while Swedish participants preferred a more monotonous voice. Regarding voice gender, Italians were more inclined towards a female voice, reporting that giving gender, especially feminine, to robots could help people. In contrast, Swedes expressed that they were not in favour of using female voices for cleaning robots because this would reinforce existing gender stereotypes. This reflects a known existing dilemma in the HRI and HCI communities, namely that the gendering of artificial agents can, on the one hand, allow for more familiar and efficient interactions [7, 44], but on the other hand, it can reinforce and propagate existing stereotypes, as highlighted by a 2019 UNESCO report [50]. While outside of the scope of the present work, scholars are trying to address this issue by e.g. considering giving gender-neutral voices to artificial agents such as robots [45].

In Experiment 2, results from the chi-square tests suggest that people's nationality influenced how they matched a voice to an image. Thanks to the answers given by the participant to the question *"...you listened to four different voices. Do you think there is a main feature that made you choose one robot over another? Which one?"*, it was possible to investigate how participants made the voice-robot associations that gave a significant result. As shown in Figure 4, Swedish participants tended to associate the male high-pitched voice with Nao (N=8), while there was no clear preference for Italian participants. Swedes explained their preference for Nao because of its size and masculine appearance. For Italians, there were no general patterns and the choice seemed highly subjective.

This influence was also confirmed by the thematic analysis. Those who preferred a genderless voice were Swedish; Italians, on the other hand, prefer robots that fit people's mental model of gender, complying with their expectations. Another difference is visible in the *interaction* theme, with only Italians being willing to have a robot as a friend and supporter. Furthermore, when it comes to communication, Swedes preferred short conversations while Italians preferred the robot to answer in full sentences. This might be due to cultural differences resulting in different conversational styles, e.g. Northern cultures tend to be more reserved and keep more personal space than Southern ones [5]. The fact that Swedish participants tended to prefer a genderless voice could also be explained by Swedish grammar. Grammatical gender in Swedish does correspond to social gender, which may make it easier for Swedish participants to think of robots and their voices as gender-neutral. On the other hand, Italian nouns have grammatical gender that corresponds to social gender (masculine or feminine), and the word "robot" is masculine. This might have influenced the ability of Italian participants to consider gender-neutral robots and voices. Even though all participants did the experiments in English, there is the possibility that participants' native language (Swedish or Italian) influences their perceptions and preferences for voices. This result is in line with research done by Roesler et al. [38]. They investigated if the language (German, a grammatically gendered language, or English, a natural gender language) influenced the perceived gender of the robot and found that a masculine perception of gender-neutral robots is often reinforced by masculine grammatical gender.

Even though differences based on nationalities were found via quantitative and qualitative analyses, a difference of opinions between Swedish participants in Experiments 1 and 2 was found regarding the sub-theme *prosody*. Indeed, in Experiment 1 the Swedish participants reported having a preference for a monotone voice while in Experiment 2 they reported preferring an expressive voice because a monotone voice would be boring and would not entice the user to interact with the robot. This apparent contradiction may derive from the fact that variations in personal experiences, beliefs, or attitudes could lead to divergent responses, even within the same cultural context.

From these results, we can suggest that human-related factors influence the mental images that people form of robots (RQ2). Specifically, nationality seems to play a role, while we did not find any evidence that gender does.

5.3 Voice-related factors influencing participants' preferences

According to the literature, pitch and gender of the voice influence people's perception and preferences of robots [13, 32, 33, 36, 41]. Extending previous findings, we investigated a new angle on the influence of pitch and gender attributes on participants from two different cultures and the way they match these attributes with robot images. To our knowledge, this had not been investigated previously.

Regarding voice gender, based on the thematic analysis, there is no clear preference towards a feminine or masculine voice. It is interesting to notice that Swedes mentioned the importance of not using female voices e.g. cleaning robots as they increase gender attribution to certain jobs, while Italians reported that giving gendered voices to robots helps to make people feel more comfortable while interacting with them or to comply with people's expectations. They reported that, if on the one hand, a robot is designed to interact with elderly people or in the medical field as a nurse, according to Italians, a female voice would be preferred. This is not only because of the stereotype that caregivers are usually women, but also because, according to some participants, female voices are more pleasant to hear, and give more confidence and security, because of the soft voice and brighter voice tones. If the robot is intended to help the user at home with basic tasks, the gender of the voice is not as important, and the most frequent feature mentioned is that the voice should be robotic. Italians tried to explain that maybe one of the reasons behind this way of thinking is because of their linguistic and cultural influence, as in Italian every noun is either grammatically masculine or feminine. Even though a lot of attention was put on gendered voices, the Swedish sided with the possibility of having a genderless voice.

One factor that influenced participants' preferences was that the voice should match the appearance (and vice versa), both in terms of pitch and gender. Indeed, participants reported that a higher pitch is preferable for small robots and a lower pitch for bigger ones. Furthermore, if the robot has purely male or female characteristics, these are also searched for in the voice, and a mismatch between the robot's appearance and its voice can confuse the user or make them reluctant to interact with it. This is in agreement with previous studies [29, 30].

Another aspect that emerged was voice naturalness. It seems that, while participants initially reported preferring a human voice for a humanoid-like robot and a robotic one for a mechanical one, when asked later what kind of voice their robot should have, they were more inclined towards a robotic voice, or at least one with some robotic features. As is also confirmed by Strait et al. [43], the main reason behind this choice might lie in the fear that these machines would replace humans, looking and sounding too much like them, and thus creating an uncanny feeling.

Participants also mentioned other voice characteristics that would be appropriate for robots: calm, peaceful and friendly voices to feel that the robots are neither judging nor threatening, expressive intonation for effective conversations and monotone voices to maintain some distance between humans and robots.

From these results, we can conclude that voice-related factors (specifically, pitch and gender) influence participants' preferences and mental images of robot voices (RQ3).

Although the focus of this research was on voice, investigating interaction preferences was crucial for understanding what the participants thought about robots and to better explain some of the responses related to voice preference. Indeed, some participants showed interest in verbal communication with the robot, while others preferred short responses or nonverbal sounds from it. The reason behind this may be both because of how people interact with voice assistants nowadays, through short and usually task-related interactions, but also because, as explained above, people are afraid of this type of technology and therefore cannot imagine verbal communication like they usually have with a human being [43]. Some of these results underline how important it is to do research *with* the user and not only *for* the user. By providing the user with flexibility, designers can support users' differing needs and desires within human-robot interactions [20].

5.4 Influences of stimulus order

We took inspiration from the work of McGinn and Torre to design the experiments reported here [29]. However, in this previous study, the order of stimulus presentation was always the same, specifically people always listened to a voice at a time and tried to match it to a set of robot images. However, the process of forming a mental image after hearing a voice might be different from the process of forming a mental image after seeing a picture. Therefore, in the present work, we conducted two experiments, changing the stimulus presentation order, to see if the process might be different. While we initially speculated that the results of the two experiments would be similar (i.e., that image-voice and voice-image association would be the same), this was not the case. One reason might be first impressions, i.e. the conclusions humans draw about someone after meeting them for the first time [2]. When attempting to make a quick and accurate initial impression of a stranger, clues connected with the face and voice are favourable since they are processed more quickly than other indicators such as a person's behavioural reactions and attire. Despite being favourable, these two clues do not contribute equally to the formation of a first impression [23]. Therefore, the first impression the participants created was probably influenced by whether they found themselves confronted with an image or a voice, which in turn influenced what they looked for when choosing a voice or a robot. To the best of our knowledge, there are no studies that looked at this difference in stimulus presentation before in this context. Our results suggest that stimulus presentation order does influence people's perception and should be investigated further in future works (RQ4).

5.5 Limitations and future works

Some limitations of the study should be acknowledged, as they could serve as starting points for future research.

The first limitation is the use of English as a language to conduct the overall study. None of the participants were native English speakers, which may have restricted their ability to express themselves clearly and, consequently, to provide more details, especially during the interviews. Secondly, we manipulated the voice gender as either masculine or feminine, and this could have reinforced the norm that robots should fall into a binary gender category. A suggestion for future works could be to experiment with a wider range of voices, some of them taken from studies that investigated gender-neutral/gender-ambiguous voices [12, 28, 45]. Furthermore, external limitations prevented us from recruiting more participants, and as such the sample size did not allow us to conduct more extensive statistical analyses.

Regarding the composition of the sample, it must be noted that the participants were chosen from university campuses for the two populations of interest. As a result, the sample might represent a younger and more educated population so it is possible that the results obtained are not indicative of a whole nation (see also the WEIRD population problem [18]).

Since we recruited participants from Sweden and Italy from convenience sampling, it is important to acknowledge that this may restrict the generalizability of our findings to other cultural contexts. Future research could explore the inclusion of participants from diverse cultural backgrounds to enrich our understanding of human-robot interactions across a broader spectrum of cultural dimensions. Indeed, given the fact that some differences were found between Swedes and Italians, it would be interesting to investigate how influential the nationality and culture of the participant are in human-robot interactions, perhaps by taking two, or more, cultures which differ more in values, such as Eastern and Western ones.

Another possible future development could be to carry out the same experiment with some of the robots physically present in the experiment's room, as some participants reported having difficulty imagining the size of the robots. Thus, it seems that our efforts to provide a reference human-size figure were not completely successful (see Figure 2). Since some participants showed a preference for short answers or nonverbal sounds as responses from the robot, it would be worth researching further when this type of interaction is preferred to e.g. verbal conversations.

In conclusion, future research should consider all the previously mentioned limitations and suggestions to provide a deeper comprehension of the issue from multiple perspectives. Furthermore, this work provides a starting point for industries to take robots' voices more into consideration when developing them, giving particular attention to users' perceptions, preferences and expectations for robot voices.

6 CONCLUSION

We conducted a novel experiment to understand what factors might influence the formation of people's mental images of robots, based on seeing a robot image or hearing a robot's voice. Such knowledge is fundamental to ensure that multimodal artificial agents, such as robots, are accepted and trusted by users, and avoid uncanny or uneasy feelings when a voice is perceived to not match its body. Our results highlight the importance of making sure that physical and auditory characteristics match, such as smaller robots needing a higher-pitched voice and, conversely, higher-pitched voices being associated with smaller robots. Our mixed-methods approach also allowed us to dig deeper into people's thinking processes when they formed these mental images, uncovering several interesting considerations, and covering topics such as gender, context and human-likeness of these artificial agents. All in all, our findings contribute to highlighting that user-centred design approaches are at the basis of meaningful human-robot interactions. By embracing the complexities of cultural, contextual, and individual preferences, we lay the groundwork for a more nuanced and inclusive approach to voice design in the field of Human-Robot Interaction.

REFERENCES

[1] Minoo Alemi, Ali Meghdari, and Maryam Ghazisaedy. 2014. Employing Humanoid Robots for Teaching English Language in Iranian Junior High-Schools. *International Journal of Humanoid Robotics* 11 (2014), 1450022–1. https://doi.org/10.1142/S0219843614500224

[2] Nalini Ambady and John J. Skowronski (Eds.). 2008. *First Impressions.* Guilford Publications, New York, NY, USA. 15–19 pages. https://books.google.se/books?id=poHGCvweVFsC

[3] Sean Andrist, Micheline Ziadee, Halim Boukaram, Bilge Mutlu, and Majd Sakr. 2015. Effects of Culture on the Credibility of Robot Speech: A Comparison between English and Arabic. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, Oregon, USA) *(HRI '15)*. Association for Computing Machinery, New York, NY, USA, 157–164. https://doi.org/10.1145/2696454.2696464

[4] Thomas Arnold and Matthias Scheutz. 2018. Observing robot touch in context: How does touch and attitude affect perceptions of a robot's social qualities?. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction.* ACM New York, NY, New York, NY, USA, 352–360. https://doi.org/10.1145/3171221.3171263

[5] Catherine Beaulieu. 2004. Intercultural study of personal space: A case study. *Journal of applied social psychology* 34, 4 (2004), 794–805. https://doi.org/10.1111/j.1559-1816.2004.tb02571.x

[6] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science Robotics* 3, 21 (2018), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

[7] Sheryl Brahnam and Antonella De Angeli. 2012. Gender affordances of conversational agents. *Interacting with Computers* 24, 3 (2012), 139–153. https://doi.org/10.1016/j.intcom.2012.05.001

[8] Joost Broekens, Marcel Heerink, and Henk Rosendal. 2009. Assistive social robots in elderly care: A review. *Gerontechnology* 8 (2009), 94–103. https://doi.org/10.4017/gt.2009.08.02.002.00

[9] De'Aira Bryant, Jason Borenstein, and Ayanna Howard. 2020. Why Should We Gender? The Effect of Robot Gendering and Occupational Stereotypes on Human Trust and Perceived Competency. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 13–21. https://doi.org/10.1145/3319502.3374778

[10] Felix Carros, Johanna Meurer, Diana Löffler, David Unbehaun, Sarah Matthies, Inga Koch, Rainer Wieching, Dave Randall, Marc Hassenzahl, and Volker Wulf. 2020. Exploring Human-Robot Interaction with the Elderly: Results from a Ten-Week Case Study in a Care Home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376402

[11] Sophie Curtis. 2016. Pizza Hut hires ROBOT waiters to take orders and process payments. http://www.mirror.co.uk/tech/who-needs-waiters-pizza-hut-8045172 Accessed on July 12, 2023.

[12] Andreea Danielescu, Sharone A Horowit-Hendler, Alexandria Pabst, Kenneth Michael Stewart, Eric M Gallo, and Matthew Peter Aylett. 2023. Creating Inclusive Voices for the 21st Century: A Non-Binary Text-to-Speech for Conversational Assistants. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 390, 17 pages. https://doi.org/10.1145/3544548.3581281

[13] Friederike Eyssel, Dieta Kuchenbrandt, Simon Bobinger, Laura de Ruiter, and Frank Hegel. 2012. 'If you sound like me, you must be more human': on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (Boston, Massachusetts, USA) *(HRI '12)*. Association for Computing Machinery, New York, NY, USA, 125–126. https://doi.org/10.1145/2157689.2157717

[14] Susan Fussell, Sara Kiesler, Leslie Setlock, and Victoria Yew. 2008. How people anthropomorphize robots, In Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction. *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots*, 145–152. https://doi.org/10.1145/1349822.1349842

[15] Daniel Hernández García, Pablo G. Esteban, Hee Rin Lee, Marta Romeo, Emmanuel Senft, and Erik Billing. 2020. Social robots in therapy and care. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI '19)*. IEEE Press, New York, NY, USA, 669–670.

[16] David Graddol and Joan Swann. 1983. Speaking Fundamental Frequency: Some Physical and Social Correlates. *Language and Speech* 26, 4 (1983), 351–366. https://doi.org/10.1177/002383098302600040

[17] Kerstin S. Haring, David Silvera-Tawil, Tomotaka Takahashi, Mari Velonaki, and Katsumi Watanabe. 2015. Perception of a humanoid robot: A cross-cultural comparison. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Kobe, Japan). IEEE, New York, NY, USA, 821–826. https://doi.org/10.1109/ROMAN.2015.7333613

[18] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and brain sciences* 33, 2-3 (2010), 61–83.

[19] Sharlene Nagy Hesse-Biber and Burke Johnson. 2015. *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry.* Oxford University Press, Oxford, United Kingdom, England. 3–20 pages. https://doi.org/10.1093/oxfordhb/9780199933624.001.0001

[20] Layne Hubbard, Shanli Ding, Vananh Le, Pilyoung Kim, and Tom Yeh. 2021. Voice Design to Support Young Children's Agency in Child-Agent Interaction. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 9, 10 pages. https://doi.org/10.1145/3469595.3469604

[21] Istituto Italiano di Tecnologia IIT. 2022. Art meets 5G: Digital experiments held at art museums in Turin. https://www.eurekalert.org/news-releases/954156 Accessed on July 12, 2023.

[22] Zuzanna Janeczko and Mary Ellen Foster. 2022. A Study on Human Interactions With Robots Based on Their Appearance and Behaviour. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 33, 6 pages. https://doi.org/10.1145/3543829.3544523

[23] Zhongqing Jiang, Dong Li, Zhao Li, Yi Yang, Yangtao Liu, Xin Yue, Qi Wu, Hong Yang, Xiaolin Cui, and Peng Xue. 2023. Comparison of Face-Based and Voice-Based First Impressions in a Chinese Sample. *British Journal of Psychology* 115, 1 (2023), 20–39. https://doi.org/10.1111/bjop.12675

[24] Katharina Kühne, Martin H Fischer, and Yuefang Zhou. 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 (2020), 105. https://doi.org/10.3389/fnbot.2020.593732

[25] Youdi Li, Eri Sato-Shimokawara, and Toru Yamaguchi. 2021. The Influence of Robot's Unexpected Behavior on Individual Cognitive Performance. In *30th*

IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE, New York, NY, USA, 1103–1109. https://doi.org/10.1109/RO-MAN50785.2021.9515317

[26] Velvetina Lim, Maki Rooksby, and Emily Cross. 2021. Social Robots on a Global Stage: Establishing a Role for Culture During Human-Robot Interaction. *International Journal of Social Robotics* 13, 6 (2021), 1307–1333. https://doi.org/10.1007/s12369-020-00710-4

[27] Martina Mara, Simon Schreibelmayr, and Franz Berger. 2020. Hearing a Nose? User Expectations of Robot Appearance Induced by Different Robot Voices. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 355–356. https://doi.org/10.1145/3371382.3378285

[28] Konstantinos Markopoulos, Georgia Maniati, Georgios Vamvoukakis, Nikolaos Ellinas, Georgios Vardaxoglou, Panos Kakoulidis, Junkwang Oh, Gunu Jho, Inchul Hwang, Aimilios Chalamandaris, Pirros Tsiakoulis, and Spyros Raptis. 2023. Generating Multilingual Gender-Ambiguous Text-to-Speech Voices. In *Proc. INTERSPEECH 2023*. Dublin, Ireland, 621–625. https://doi.org/10.21437/Interspeech.2023-1467

[29] Conor McGinn and Ilaria Torre. 2020. Can you tell the robot by the voice? an exploratory study on the role of voice in the perception of robots. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction* (Daegu, Republic of Korea) *(HRI '19)*. IEEE Press, New York, NY, USA, 211–221.

[30] Roger K Moore. 2017. Appropriate voices for artefacts: some key insights. In *1st International workshop on vocal interactivity in-and-between humans, animals and robots*. Skovde, Sweden, 7–11.

[31] Clinton Nguyen. 2016. Restaurants in China are replacing waiters with robots. https://www.businessinsider.com/chinese-restaurant-robot-waiters-2016-7 Accessed on July 12, 2023.

[32] Andreea Niculescu, Betsy Dijk, Anton Nijholt, Haizhou Li, and Sl See. 2013. Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy. *International Journal of Social Robotics* 5 (2013), 171–191. https://doi.org/10.1007/s12369-012-0171-x

[33] Andreea Niculescu, Betsy van Dijk, Anton Nijholt, and Swee Lan See. 2011. The influence of voice pitch on the evaluation of a social robot receptionist. In *2011 International Conference on User Science and Engineering (i-USEr)*. IEEE, New York, NY, USA, 18–23. https://doi.org/10.1109/iUSEr.2011.6150529

[34] David Obremski, Paula Friedrich, Nora Haak, Philipp Schaper, and Birgit Lugrin. 2022. The impact of mixed-cultural speech on the stereotypical perception of a virtual robot. *Frontiers in Robotics and AI* 9 (2022). https://doi.org/10.3389/frobt.2022.983955

[35] Cyril R. Pernet and Pascal Belin. 2012. The Role of Pitch and Timbre in Voice Gender Categorization. *Frontiers in Psychology* 3 (2012), 23. https://doi.org/10.3389/fpsyg.2012.00023

[36] Hamed Pourfannan, Hamed Mahzoon, Yuichihiro Yoshikawa, and Hiroshi Ishiguro. 2022. Towards a simultaneously speaking bilingual robot: Primary study on the effect of gender and pitch of the robot's voice. *PLoS ONE* 17, 12 (2022), e0278852. https://doi.org/10.1371/journal.pone.0278852

[37] Nazerke Rakhymbayeva, Aida Amirova, and Anara Sandygulova. 2021. A Long-Term Engagement with a Social Robot for Autism Therapy. *Frontiers in Robotics and AI* 8 (2021), 669972. https://doi.org/10.3389/frobt.2021.669972

[38] Eileen Roesler, Maris Heuring, and Linda Onnasch. 2023. (Hu)man-Like Robots: The Impact of Anthropomorphism and Language on Perceived Robot Gender. *International Journal of Social Robotics* 15 (2023), 1–12. https://doi.org/10.1007/s12369-023-00975-5

[39] Jeff Russ, Ruben Gur, and Warren Bilker. 2008. Validation of affective and neutral sentence content for prosodic testing. *Behavior research methods* 40 (2008), 935–939. https://doi.org/10.3758/BRM.40.4.935

[40] Busra Sarigul, Imge Saltik, Batuhan Hokelek, and Burcu A. Urgen. 2020. Does the Appearance of an Agent Affect How We Perceive His/Her Voice? Audio-Visual Predictive Processes in Human-Robot Interaction. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 430–432. https://doi.org/10.1145/3371382.3378302

[41] Mikey Siegel, Cynthia Breazeal, and Michael I. Norton. 2009. Persuasive Robotics: The influence of robot gender on human behavior. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, New York, NY, USA, 2563–2568. https://doi.org/10.1109/IROS.2009.5354116

[42] Julia Simkus. 2023. Convenience Sampling: Definition, Method and Examples. https://www.simplypsychology.org/convenience-sampling.html Accessed on July 02, 2023.

[43] Megan K. Strait, Cynthia Aguillon, Virginia Contreras, and Noemi Garcia. 2017. The public's perception of humanlike robots: Online social commentary reflects an appearance-based uncanny valley, a general fear of a "Technology Takeover", and the unabashed sexualization of female-gendered robots. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, New York, NY, USA, 1418–1423. https://doi.org/10.1109/ROMAN.2017.8172490

[44] Benedict Tay, Younbo Jung, and Taezoon Park. 2014. When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior* 38 (2014), 75–84. https://doi.org/10.1016/j.chb.2014.05.014

[45] Ilaria Torre, Erik Lagerstedt, Nathaniel Dennler, Katie Seaborn, Iolanda Leite, and Éva Székely. 2023. Can a gender-ambiguous voice reduce gender stereotypes in human-robot interactions?. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, New York, NY, USA, 106–112. https://doi.org/10.1109/RO-MAN57019.2023.10309500

[46] Ilaria Torre, Adrian Benigno Latupeirissa, and Conor McGinn. 2020. How context shapes the appropriateness of a robot's voice. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, New York, NY, USA, 215–222. https://doi.org/10.1109/RO-MAN47096.2020.9223449

[47] Ilaria Torre and Sébastien Le Maguer. 2020. Should robots have accents?. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, IEEE, New York, NY, USA, 208–214. https://doi.org/10.1109/RO-MAN47096.2020.9223599

[48] Ilaria Torre and Laurence White. 2020. *Trust in Vocal Human–Robot Interaction: Implications for Robot Voice Design*. Springer, Singapore, 299–316. https://doi.org/10.1007/978-981-15-6627-1_16

[49] Verbit. 2023. How and when to use clean verbatim transcription. https://verbit.ai/how-and-when-to-use-clean-verbatim-transcription/ Accessed on February 10, 2024.

[50] Mark West, Rebecca Kraut, and Han Ei Chew. 2019. *I'd blush if I could: closing gender divides in digital skills through education*. Technical Report. https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1

[51] Katie Winkle, Erik Lagerstedt, Ilaria Torre, and Anna Offenwanger. 2023. 15 Years of (Who) man Robot Interaction: Reviewing the H in Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction* 12, 3 (2023), 1–28. https://doi.org/10.1145/3571718

[52] Selma Yilmazyildiz, Georgios Athanasopoulos, Georgios Patsis, Weiyi Wang, Meshia Oveneke, Lukas Latacz, Werner Verhelst, Hichem Sahli, David Henderickx, Bram Vanderborght, Eric Soetens, and Dirk Lefeber. 2013. Voice Modification for Wizard-of-Oz Experiments in Robot-Child Interaction. In *In Proceedings of the Workshop on Affective Social Speech Signals*.