Can a gender-ambiguous voice reduce gender stereotypes in human-robot interactions?

Ilaria Torre[1], Erik Lagerstedt[2], Nathaniel Dennler[3], Katie Seaborn[4], Iolanda Leite[1] & Éva Székely[1]

Abstract— **When deploying robots, its physical characteristics, role, and tasks are often fixed. Such factors can also be associated with gender stereotypes among humans, which then transfer to the robot. One factor that can induce gendering but is comparatively easy to change is the robot's voice. Designing voice in a way that interferes with fixed factors might therefore be a way reduce gender stereotypes in human-robot interaction contexts. To this end, we have conducted a video-based online study to investigate how factors that might inspire gendering of a robot interact. In particular, we investigated how giving the robot a gender-ambiguous voice can affect perception of the robot. We compared assessments (n=111) of videos in which a robot's body presentation and occupation mis/matched with human gender stereotypes. We found evidence that a gender-ambiguous voice can reduce gendering of a robot endowed with stereotypically feminine or masculine attributes. The results can inform more just robot design while opening new questions regarding the phenomenon of robot gendering.**

I. Introduction

Social robots are meant to be deployed in a variety of places to complete various tasks, such as helping in schools, the home, hospitals, factories, and so on. When investigating the effect of certain robot characteristics on human perception and behaviour, the factors of the *embodiment* of the robot, i.e., which robotic platform will be used, and its *assigned task and context of use* are very often fixed. For example, if we want to investigate whether a certain characteristic influences children's engagement in a learning activity, we must work with a *fixed robot* (whichever robot we have available or have deemed suitable for the investigation) in a *fixed context* (the learning activity).

The fixed factors of robot embodiment and context of use can affect how the robot is perceived, particularly if they are based in *human social identities* and/or related to *human gender stereotypes* [1]. Notably, previous work has found that people transfer gender stereotypes from the human world to artificial agents, including virtual avatars [2], computers [3], and robots [4]. For example, "male" robots were perceived as more suitable for stereotypically "male" tasks, e.g., repairing technical devices, guarding a house, while "female" robots were perceived as more suitable for stereotypically "female" tasks, e.g., tasks related to household and care services [4], [5]. This poses a dilemma. On the one hand, ensuring that robots are matched in terms of perceived gender and task suitability might reduce the potential for Uncanny Valley effects and increase trust [6]. On the other, explicitly mismatching gender cues might reduce gender stereotypes over time, i.e., the robot as a debiasing tool [7], [8].

Since direct experiences with social robots are still rare among the general public, we have an opportunity to act now and study how human attributes are (or could be) translated to robots so as to minimise deception while ensuring understandability and optimal user experience (UX) [9]. Any initial attributes ascribed to robots that are associated with human stereotypes can then be incrementally challenged as they are identified and before robots become more common used by a diversity of people. To this end, we must understand how human social identities are shaped during interactions with the robot and how different human social norms and cues interact with each other. However, this may not be straightforward. For example, people may perceive robots more favourably and be more willing to comply with robots when the robots behave appropriately for their task [10]. Moreover, even though stereotypes regarding gender and occupation can impact perceived competence and trust in human actors, these phenomena do not always transfer to gendered robots [11]. We must further study these phenomena with a variety of complexity to be able to capture higher order effects.

One characteristic relevant to gender attribution is *robot voice* [12], [13]. Unlike the robot platform, i.e., body and programmable actions, and the application context, voice is more "free" in the sense that it can be chosen for the robot independently of any hardware or application constraints. As before, this raises a dilemma. First, we must ensure that the voice is appropriate for the robot (see e.g. [8], [14]). This is sometimes assumed to mean that the voice should be congruent with other physical and social characteristics of the robot, such as its size and shape [8], [14]. However, voice could be used to challenge the gender stereotypes afforded by the aforementioned "fixed" robot factors [15]. As yet, little work in HRI has explored these tensions and potentialities.

Reducing stereotypes does not necessarily mean adding the "opposite" attribute, e.g., a masculine voice for a feminine body. This risks rein/en/forcing a binary, cisnormative perspective of gender and the heterosexuality matrix [16], where there are only men and women who are "opposites," assumed to be heterosexual, and fixed in terms of sex[1], gender, and sexual identities. An alternative is mixing and

[1]School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Sweden [ilariat, iolanda, szekely]@kth.se
[2]School of Informatics, University of Skövde, Sweden erik.lagerstedt@his.se
[3]Department of Computer Science, University of Southern California, USA dennler@usc.edu
[4]Department of Industrial Engineering and Economics, Tokyo Institute of Technology, Japan seaborn.k.aa@m.titech.ac.jp

[1]We distinguish sex as the biological and gender as the social [17].

matching attributes to find something new, while avoiding Uncanny Valley territory [6]. Here, we explore one potential strategy: *gender-neutral voice* [13]. We selected a gender-ambiguous synthetic voice to serve as the voice of a robot mis/matched in terms of gender stereotypes in appearance and application context [4]. We considered whether the voice was perceived as gendered, ambiguous, and/or gender-neutral, as well as what factors of the robot embodiment—body or role within the context—were influential. We asked: *Does a robot with a gender-ambiguous voice influence perceptions of its gender, uncanniness, and trustworthiness in embodiments matched or mismatched according to human gender stereotypes?* We extend previous methodologies for gender and robot embodiment as well as contribute novel findings on the utility of gender-neutral voice.

II. Related Work

A. Gender Neutrality in Robots

The notion of gender neutrality in robots and other agents has a relatively long but complicated history in HRI [12]. Seaborn and Pennefather [12] conducted a rapid review of the robotics literature, finding that voice was a key factor in modulating perceptions of robot gender. Even so, the literature has so far been limited in important ways, notably insufficient theorizing on gender and designing for gender perceptions in robots; assumptions of gender neutrality and a lack of manipulation checks; and limitations in reporting, preventing a clear consensus on design and experiential findings related to voice, body, role, application context, methodology, user perceptions, and so on. This work highlights a need for increased rigour and a concerted focus on gender neutrality, which we have attempted to address.

Seaborn and Pennefather [12] and Sutton [18] also raised the question of how we frame and approach gender conceptually, methodologically, and practically. Notably, we can distinguish robots that are gender *ambiguous*, i.e., androgynous and/or having a mixture of masculine, feminine, or other gender characteristics, from those that are gender *neutral*, i.e., agender or genderless, effectively not stimulating gender perceptions. At present, the relationship between these two concepts for robot voice is unclear. We begin by asking **RQ1**: *Does a gender-ambiguous voice influence gendering of robots, or is it gender neutral?* Using a gender-ambiguous voice with a robot gendered in body and occupational context might reduce the overall gendering of the robot by people. Still, an "ambiguous" voice, i.e., one with a mixture of gender attributes drawn from real people and modified computationally, might instead lead to stronger gender attributions. As Sutton [18] points out, people are primed to gender even when voice is fundamentally, "objectively" genderless. With this in mind, we hypothesize **H1**: *A gender-ambiguous voice will influence the gendering of robots.*

B. Mis/matching Gender Stereotypes

We anchored our work on the research design of Eyssel et al. [4], discussed above. However, other work has considered whether and how we should align human stereotypes about gender with robotic embodiments. Tay et al. [5] had people interact with a robot in two different gender-stereotyped roles (healthcare as feminine and security as masculine). They found that people had more positive perceptions of the robot when its perceived gender matched the gender stereotype associated with the role. Aşkın et al. [19] showed that actions stereotypically perceived as feminine or masculine led to perceptions of a pilot-deemed gender-neutral robot as having the corresponding gender. Moreover, this effect was stronger when people were asked to consider society's view rather than their own. In contrast, Bryant et al. [11] found no effect of perceived robot gender on perceived occupational competency and trust towards the robot.

Thus, results so far are mixed. Additionally, we expect an influence of the voice, and we are not sure how ambiguous voices will be interpreted when other features of the robot's embodiment are gendered. We therefore asked **RQ2**: *Does a gender-ambiguous voice influence gendering of robots that are matched and mismatched in terms of gender body presentation and occupation stereotypes?* This leads us to two complementary hypotheses. First, we may expect the body and application context to influence gendering of the voice, even if the voice is gender neutral or ambiguous, e.g., [13], [19], or **H2a**: *Stereotypically matched body and occupation will lead to stronger attributions of gender to the voiced robot compared to stereotypically mismatching body and occupation, i.e., complementarity in gender role stereotypes and body gender perceptions.* Second, we would expect the robot to be gendered in the absence of a voice, e.g., [4], [5], or **H2b**: *Stereotypically matched body and task context will lead to stronger attributions of gender to the unvoiced robot compared to stereotypically mismatching body and task context, i.e., complementarity in gender role stereotypes and body gender perceptions.*

C. Voice Priming Effects

Voice gendering can greatly influence other gender perceptions [12], [13], [18]. Given this, we also wondered, mis/matching aside, **RQ3**: *Does listening to a gender-ambiguous voice lead to priming gendering effects?* As per the literature, we would expect **H3**: *The gendering of the voice alone will influence the gendering of the robot and voice combination.* To evaluate this, we included a voice-only session before the main study featuring a context irrelevant to those used in the main study, i.e., reading an encyclopedia entry. We then asked participants to evaluate their perceptions of the voice in terms of gender. This also acted as a post-pilot study manipulation check.

D. Uncanny Gendering

Mismatching gender cues and stereotypes can lead to feelings of uncanniness towards the robot [4], [6]. As Paetzel-Prüsmann et al. [6] found, robots perceived as gendered and assigned to roles matching the societal stereotypes associated with those genders rendered Uncanny Valley effects null. Still, we should avoid relying on stereotypes because of the risk for interactions with such robots to reify such limited and negative perspectives in people's minds [20], [21]. What remains understudied is whether and to what extent feelings of uncanniness occur on gender mismatching. We therefore asked **RQ4**: *Is there an Uncanny Valley effect when body and application context do not match?* Pointedly, we sought to confirm or refute **H4**: *Stereotyped mismatching of robot body and occupation will lead to higher ratings of uncanniness, trust, and generally less positive evaluations.*

While many of our questions and hypotheses link to an expectation of stereotyped responses, we remained hopeful that deploying a gender-ambiguous voice could positively disrupt and shift perceptions. The main focus of this work was to evaluate the potential gender neutrality of a gender-ambiguous voice for robotic embodiments that may (or may not) be gendered. We thus lean on and extend the literature on robot design and deployment for contexts in which participant gendering of the robot could occur.

III. Pilot study

We conducted a pilot study to choose a gender-ambiguous voice for the main study. The Text-To-Speech (TTS) community has recently started to develop voices that could be considered gender neutral or gender ambiguous [22]–[24]. We selected an industry-standard voice from a set of 6 English-speaking voices for Amazon Polly [25]. We used "Kendra" with a lowered pitch such that the average fundamental frequency (f0) was ∼135Hz. We determined that all Polly voices were highly intelligible and "Kendra" at this pitch was perceived as ambiguously gendered. We also generated 3 novel voices based on state-of-the-art TTS technologies. A multi-speaker version of the sequence-to-sequence neural TTS engine Tacotron 2 [26] was used to train a voice on two large TTS corpora (female [27], male [28]). A gender-ambiguous timbre can be achieved by using a weighted average of speaker embeddings at inference. Three different combinations of weights (f0.3/m0.3; f0.45/m0.4; f0.65/m0.55) were selected for evaluation using a Speech Gender Recognition (SGR) network, similar to the one in [29][2].

A. Method

The study was conducted on the crowd-sourcing platform Prolific. Participants listened to 10 sentences synthesised with each of the 4 candidate voices in randomised order. They indicated their perception of voice gender using four Likert items, ranging from 1 (= strongly disagree) to 5 (= strongly agree): "Feminine," "Masculine," "Agender," and "Ambiguous". This was to allow for more nuanced ratings of gender, going beyond the binary (e.g., a male-female continuum), as one voice (or one robot, or one person, for that matter) can have masculine, feminine, and other attributes at the same time. We provided definitions for "Agender" and "Ambiguous", as follows: "By 'Ambiguous', we mean that the voice sounds neutral or androgynous; by 'Agender', we mean that the voice does not seem to have a gender at all". To rule out any priming effects due to presenting one scale before the other, we also showed half the participants the "Feminine" scale first, and half the participants the "Masculine" scale first. The sentences were taken from the TIMIT corpus [30], which consists of emotionally neutral utterances [31], and were passed through a gender bias detection tool to ensure they were free of textual bias[3]. The set of stimuli also included 4 attention checks, which were 2 unambiguously feminine voices (Amazon Polly "Kendra" with f0 set at 150 Hz and the Tacotron TTS using only the female speaker embedding) and 2 unambiguously masculine voices (Amazon Polly "Kendra" with f0 set at 120 Hz and the Tacotron TTS using only the male speaker embedding). Thus, each participant listened to 44 sentences in total.

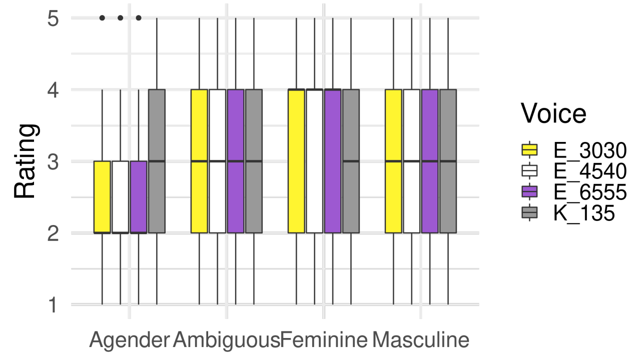

Fig. 1: Gendering of the candidate gender-ambiguous voices.

We recruited 62 participants (30 identified as female, 31 as male, none as non-binary or preferring to self-describe, and 1 preferred not to say), aged 19–48 (median=25). The experiment took ∼20 minutes and participants were paid £3.00. The study was run in accordance with the ethical guidelines from KTH Royal Institute of Technology.

B. Data Analysis and Results

Two participants failed 1+ attention check and were removed from the analyses, leading to a data set from 60 participants. Analyses were conducted in R (v4.2.1). As per Figure 1, there was more variation for the voices we designed (called "E_3030", "E_4540", "E_6555"), whereas the Amazon Polly voice ("K_135") was more consistently neutral across all scales. Thus, we selected this voice for the main study.

IV. Main study

After finding our "gender ambiguous" voice, we conducted a video-based study to see whether adding this voice to masculine and feminine robots that did or did not conform to human gender stereotypes (in terms of appearance and occupation) would influence the its perceived gender. Our protocol was registered before data collection on OSF[4].

[2]The full description of this novel approach is beyond the scope of this paper, and is under review elsewhere.

[3]`https://www.appcast.io/gender-bias-decoder`
[4]`https://osf.io/epy83`

A. Method

Fig. 2: Feminine (left) and masculine (right) robot conditions.

We created videos of the Furhat robot in different conditions. We chose this robot due to its manipulability, e.g., modifications of its face while keeping the rest of the body constant. For the feminine version of the robot, we chose the "Fedora" face and placed a long-haired brunette wig on the robot's head. For the masculine version, we chose the "Marty" face with a beanie hat (see Figure 2). Previous studies have shown that such cues can induce gendering of robots, even when shown on screen [32]. For the stereotypical occupations, we selected child tutor as stereotypically feminine, and security guard as stereotypically masculine (as per [4], [5]). We generated two utterances from the selected gender-ambiguous voice: *"Hi, my name is Sam. I was developed by a Swedish robotics company and my job is to tutor children after school"* and *"Hi, my name is Sam. I was developed by a Swedish robotics company and my job is to guard people's homes."* The robot's name was chosen as a gender-neutral English name. We mixed and matched the videos and utterances to create two sets of two videos that matched or mismatched based on human gender stereotypes Finally, we created voice-less versions of the videos with the sentences written as subtitles. Thus, we had 8 video stimuli.

Participants were recruited on Prolific. After agreeing to participate, they listened to the TTS voice "reading" a text adapted from the Encyclopaedia Britannica entry for "robot" and rated its genderedness with the pilot study scales. They then answered two attention check questions. Next, they watched the first block of video stimuli (either voiced or unvoiced) and rated each video using 9-point Likert items in terms of gender and Uncanny Valley-related perceptions [33]: likeability, trustworthiness, uncanniness, creepiness, scariness, eeriness. Then they answered two questions on their understanding and impressions of robots in general, which served as a distractor task. Then they watched and rated the second block of video stimuli. The study ended with a demographics questionnaire (gender and age) and a question on whether they had technical issues or any other comments. The experiment took ~20 minutes.

We recruited 120 people. 9 failed the attention checks or had technical issues and were removed. Of the remaining 111, aged 18–53 years old (median=23), 56 identified as male, 50 as female, 3 as non-binary, and 2 preferred not to say. They were paid £3.00. The study conformed to the ethical guidelines of KTH Royal Institute of Technology.

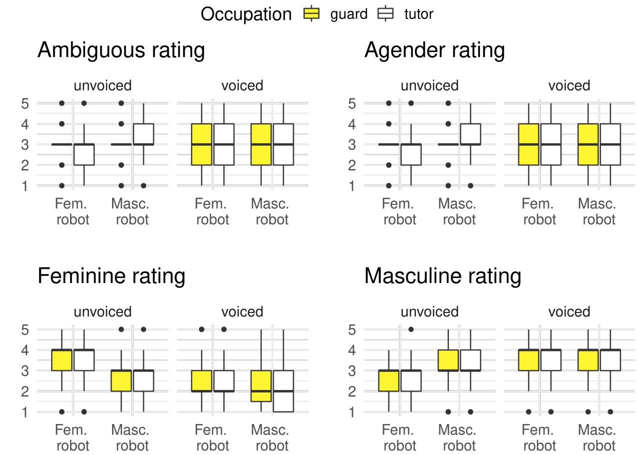

Fig. 3: Average gender ratings divided by voicing (presence or absence of the gender-ambiguous TTS voice) and robot gender manipulations (appearance and occupation).

B. Results

As a first manipulation check, we looked at how people perceived our gender-ambiguous voice, which they listened to prior to watching the video stimuli. A Kruskal-Wallis rank sum test showed that people perceived the voice differently in terms of gender ($\chi^2(3) = 125.11, p < .001$). Pairwise post-hoc comparisons using the Wilcoxon rank sum test with continuity correction showed that, on average, people rated the voice as being more "agender" than "ambiguous" ($p = .004$); more "feminine" than "agender" ($p < .001$), "ambiguous" ($p < .001$), and "masculine" ($p < .001$); and less "masculine" than "agender" ($p < .001$) and "ambiguous" ($p < .001$). Thus, it seems that, overall, people perceived the voice as more "feminine", even though the ratings centre around the "neutral" label in the Likert items (see summary statistics in Table I).

Gender	Agender	Ambiguous	Feminine	Masculine
Voice	2.72 (1.09)	3.14 (1.10)	3.84 (0.93)	2.17 (0.88)
Fem. body	2.97 (1.00)	2.92 (0.86)	3.54 (1.00)	2.44 (0.90)
Masc. body	3.12 (0.95)	3.11 (0.83)	2.59 (0.88)	3.28 (0.91)
Fem. task	3.06 (1.01)	3.00 (0.87)	3.04 (1.04)	2.88 (0.99)
Masc. task	3.03 (0.94)	3.03 (0.83)	3.09 (1.06)	2.84 (1.01)

TABLE I: Manipulation check means and standard deviations (parentheses): voice only, feminine robot (unvoiced), masculine robot (unvoiced), stereotypically feminine task (unvoiced), and stereotypically masculine task (unvoiced).

We evaluated whether our robot appearance and task manipulations were successful using the unvoiced data. Kruskal-Wallis rank sum tests showed that the feminine robot was rated as more "feminine" than "masculine" ($\chi^2(1) = 122.54, p < .001$) and the masculine robot was rated as more "masculine" than "feminine" ($\chi^2(1) = 62.61, p < .001$). For the occupation, we find that the children tutor occupation was rated as equally "feminine" and "masculine" ($\chi^2(1) =$

$2.12, p = .15$), while the security guard occupation was rated as more "feminine" ($\chi^2(1) = 6.21, p = .012$). In short, only the appearance manipulation worked.

To answer RQ1, we used Kruskal-Wallis rank sum tests to test whether gender perceptions were affected by voicing (see Figure 3). For the "Agender" ratings, we found no effect of voicing ($\chi^2(1) = 2.79, p = .095$); for the "Ambiguous" ratings, we found no effect of voicing ($\chi^2(1) = 0.22, p = .639$); for the "Feminine" ratings, we found a main effect of voicing ($\chi^2(1) = 115.65, p < .001$), with higher ratings in the unvoiced (M = 3.07, sd = 1.05) compared to the voiced condition (M = 2.32, sd = 0.97); finally, for the "Masculine" ratings, we found a main effect of voicing ($\chi^2(1) = 154.87, p < .001$), with higher ratings in the voiced (M = 3.68, sd = 0.91) compared to the unvoiced condition (M = 2.86, sd = 1.00). Thus, we find support for H1.

To answer RQ2, we conducted a series of Kruskal-Wallis rank sum tests to see whether gender perception was affected by having gender-conforming robot appearance and task. For the "Agender" ratings, we find no effect of conformity, neither for the voiced ($\chi^2(1) = 0.82, p = .366$) nor unvoiced ($\chi^2(1) = 1.58, p = .209$) stimuli. For the "Ambiguous" ratings, we find no effect of conformity, neither for the voiced ($\chi^2(1) = 0.03, p = .863$) nor unvoiced ($\chi^2(1) = 1.26, p = .261$) stimuli. For the "Feminine" ratings, we also find no effect of conformity, neither for the voiced ($\chi^2(1) = 0.36, p = .551$) nor unvoiced ($\chi^2(1) = 0.25, p = .614$) stimuli. Finally, for the "Masculine" ratings, we also find no effect of conformity, neither for the voiced ($\chi^2(1) = 0.140, p = .708$) nor unvoiced ($\chi(1) = 0.04, p = .844$) stimuli. Thus, we reject H2a and H2b.

For RQ3, we used linear models on participants' ratings of the voice alone to explain how they rated the voiced video stimuli (see Figure 4). For the "Agender" ratings, we find a significant influence of the voice only ratings ($b = 0.27$, 95% CI $[0.15, 0.38]$, $t(109) = 4.72$, $p < .001$); we also find a significant effect for the "Ambiguous" ($b = 2.39$, 95% CI $[2.05, 2.73]$, $t(109) = 13.89$, $p < .001$) and "Masculine" ($b = 0.14$, 95% CI $[0.04, 0.24]$, $t(109) = 2.67$, $p = .009$) ratings; however, we find no effect for the "Feminine" ratings ($b = 0.00$, 95% CI $[-0.11, 0.11]$, $t(109) = 0.00$, $p = .999$). Thus, we find partial support for H3.

Finally, to answer RQ4, we conducted a series of Kruskal-Wallis rank sum tests to see whether having gender-conforming or gender-non-conforming robots influenced ratings of likeability, trustworthiness, uncanniness, creepiness, scariness, and eeriness of the robot. We do not find any significant effect ($\chi^2(1) = 0.06, p = .809$; $\chi^2(1) = 0.09, p = .767$; $\chi^2(1) = 0.32, p = .572$; $\chi^2(1) = 0.31, p = .581$; $\chi^2(1) = 0.39, p = .534$, $\chi^2(1) = 0.20, p = .656$, respectively). Thus, we do not find support for H4.

V. Discussion

A gender-ambiguous voice can help reduce gendering towards a robot that is endowed with stereotypically feminine or masculine attributes (**RQ1**). People rated the robot as more "feminine" when there was no voice compared to when

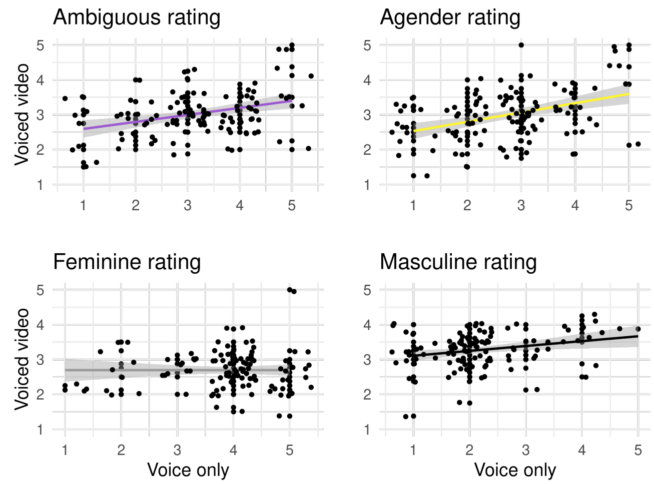

Fig. 4: Correlations between gender ratings of the voice alone and of the voiced videos. Jitter added to aid visualisation.

the robot was combined with a gender-ambiguous voice. As there tend to be more negative stereotypes and attitudes towards feminine robots than masculine ones [34], [35], this is an important result for the HRI community, suggesting that these stereotypes could potentially be reduced with the use of a gender-ambiguous voice.

The results from the other gender scales give additional insights. Adding a gender-ambiguous voice to the robot did not change participants' ratings of it being "ambiguous" or "agender". If people perceived the robot's appearance as "ambiguous" or "agender", this perception happened whether an ambiguous voice was present or not. Thus, a gender-ambiguous voice does not seem to drastically change the perception of a robot that is already seen as such.

Still, the gender-ambiguous voice increased the "masculine" ratings of the robot. This is surprising given that the voice by itself was perceived as slightly "feminine." Given the prevalence of a "male robot" bias [36], [37], robots perceived as "masculine" may have been more strongly categorised as such, with no room for ambiguity or fluidity. This is consistent with categorical perception theories, where the perceived difference between a stimuli and the prototype of its category will be smaller than the actual difference [38]. In our case, the voice may not have been disruptive enough to force the stimuli to the border between perceived genders, in which case a stronger Uncanny Valley effect would have been expected [39]. If so, then the addition of an ambiguous voice simply reinforced the existing conceptual masculine paradigm, providing multi-modal information that lead to even stronger categorisation than uni-modal information (the video without voice) [40]. This interpretation is supported by the fact that we did not find any evidence of an Uncanny Valley or related phenomena (**RQ4**), suggesting that people never perceived the stimuli as incongruent or "uncomfortable." This may be due to the fact that the robot occupation manipulation was not successful. Or, in the presence of mismatching cues (a robot perceived as "masculine" and an ambiguous or feminine-leaning voice), people integrated these two conflicting sources of information and perceived the stimulus as something entirely different—a more "masculine" robot—akin to the "McGurk effect" [41].

We also did not find evidence that robots conforming with human gender stereotypes were gendered more than "mismatched" robots (**RQ2**). Again, this may be due to the robot occupation manipulation not being successful. Still, Rea et al. [42] also did not find strong evidence for occupation-related gender stereotypes in robots. Perhaps the robot simply stating its occupation was not enough for people to form a gender-occupation connection. On the one hand, this is a positive result, suggesting that occupations until recently perceived as gendered [4], [5] are not so anymore, at least for the sample of participants we selected. Still, as gender stereotypes are still prevalent in society [20], [43], we must extend this line of work and conduct additional pilot testing to find occupations, or other embodiment factors, that are currently stereotyped and have room to be disrupted.

Finally, we found that voice gendering alone influenced gendering of the combined robot and voice stimuli for all gender scales apart from the "feminine" one **RQ3**. We must also consider that the voice alone was rated as slightly "feminine." This could be a ceiling effect. However, this suggests that we are on the right track: a voice that is more consistently judged to be "agender" or "ambiguous" can alter gender perceptions of robots that are otherwise, perhaps uncontrollably, gendered by participants. We commend the work that the HRI and TTS communities are performing towards designing and developing such voices. There is still work to be done, as people tend to categorise voices as either "feminine" or "masculine" and it is difficult to create a truly "neutral" voice [44], [45]. In particular, we plan to involve the genderqueer community in the design process.

There are a few limitations that should be mentioned. Most notably, despite our pilot study suggesting otherwise, participants in the main study still found our chosen voice to be more "feminine" than "agender" or "ambiguous". This might be due to the context in which they heard the voice (reading the encyclopaedia entry) which might have been perceived as more "feminine". Nonetheless, we consider our results to be informative and promising, and we plan to continue exploring this topic. Furthermore, we could not examine the interplay between the voice, the robot and participants' individual differences—most notably, their own gender, attitudes towards robots, and propensity to gender-stereotype—due to space limitations. But we plan to conduct such analyses for future work. Also, we plan to conduct follow-up studies in person. The larger research question that we want to address is whether gender-ambiguous voices on robots can reduce gender stereotypes towards robots, but also towards people.

Acknowledgements

This work was partially funded by the Digital Futures research centre. All figures were generated with a colour palette based on the non-binary pride flag, created by Joel Le Forestier (`https://joelleforestier.com/#pridepalettes`).

References

[1] R. A. Søraa, "Mechanical genders: how do humans gender robots?" *Gender, Technology and Development*, vol. 21, no. 1-2, pp. 99–115, 2017.

[2] S. Brahnam and A. De Angeli, "Gender affordances of conversational agents," *Interacting with Computers*, vol. 24, no. 3, pp. 139–153, 2012.

[3] C. I. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1994, pp. 72–78.

[4] F. Eyssel and F. Hegel, "(s) he's got the look: Gender stereotyping of robots," *Journal of Applied Social Psychology*, vol. 42, no. 9, pp. 2213–2230, 2012.

[5] B. Tay, Y. Jung, and T. Park, "When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction," *Computers in Human Behavior*, vol. 38, pp. 75–84, 2014.

[6] M. Paetzel, C. Peters, I. Nyström, and G. Castellano, "Congruency matters-how ambiguous gender cues increase a robot's uncanniness," in *International conference on social robotics*. Springer, 2016, pp. 402–412.

[7] J. Bernotat, F. Eyssel, and J. Sachse, "The (fe) male robot: how robot body shape impacts first impressions and trust towards robots," *International Journal of Social Robotics*, vol. 13, no. 3, pp. 477–489, 2021.

[8] C. McGinn and I. Torre, "Can you tell the robot by the voice? an exploratory study on the role of voice in the perception of robots," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 211–221.

[9] B. F. Malle, K. Fischer, J. E. Young, A. Moon, and E. Collins, "Trust and the discrepancy between expectations and actual capabilities," in *Human—Robot Interaction: Control, Analysis, and Design*, D. Zhang and B. Wei, Eds. Cambridge Scholars Publishing, 2020, ch. 1, pp. 1–23.

[10] J. Goetz, S. Kiesler, and A. Powers, "Matching robot appearance and behavior to tasks to improve human-robot cooperation," in *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003.* IEEE, 2003, pp. 55–60.

[11] D. Bryant, J. Borenstein, and A. Howard, "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*, 2020, pp. 13–21.

[12] K. Seaborn, N. P. Miyake, P. Pennefather, and M. Otake-Matsuura, "Voice in human–agent interaction: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–43, 2021. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3386867

[13] K. Seaborn and P. Pennefather, "Neither "hear" nor "their": Interrogating gender neutrality in robots," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2022, pp. 1030–1034. [Online]. Available: https://doi.org/10.1109/HRI53351.2022.9889350

[14] R. K. Moore, "Appropriate voices for artefacts: some key insights," in *1st International workshop on vocal interactivity in-and-between humans, animals and robots*, 2017.

[15] J. Cambre and C. Kulkarni, "One voice fits all? social implications and research challenges of designing voices for smart devices," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–19, 2019.

[16] J. Butler, *Gender trouble*. Routledge, 2002.

[17] C. Tannenbaum, R. P. Ellis, F. Eyssel, J. Zou, and L. Schiebinger, "Sex and gender analysis improves science and engineering," *Nature*, vol. 575, no. 7781, pp. 137–146, 2019. [Online]. Available: https://www.nature.com/articles/s41586-019-1657-6

[18] S. J. Sutton, "Gender ambiguous, not genderless: Designing gender in voice user interfaces (vuis) with sensitivity," in *Proceedings of the 2nd conference on conversational user interfaces*, 2020, pp. 1–8. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3405755.3406123

[19] G. Aşkın, İ. Saltık, T. E. Boz, and B. A. Urgen, "Gendered actions with a genderless robot: Gender attribution to humanoid robots in action," *International Journal of Social Robotics*, pp. 1–17, 2023.

[20] A. Galatolo, G. I. Melsión, I. Leite, and K. Winkle, "The right (wo) man for the job? exploring the role of gender when challenging gender stereotypes with a social robot," *International Journal of Social Robotics*, pp. 1–15, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s12369-022-00938-2

[21] K. Seaborn and A. Frank, "What pronouns for pepper? a critical review of gender/ing in research," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–15. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3491102.3501996

[22] C. Yu, C. Fu, R. Chen, and A. Tapus, "First attempt of gender-free speech style transfer for genderless robot," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2022, pp. 1110–1113.

[23] S. Mooshammer and K. Etzrodt, "Social research with gender-neutral voices in chatbots–the generation and evaluation of artificial gender-neutral voices with praat and google wavenet," in *Chatbot Research and Design: 5th International Workshop, CONVERSATIONS 2021, Virtual Event, November 23–24, 2021, Revised Selected Papers*. Springer, 2022, pp. 176–191.

[24] J. Carpenter, "Why project q is more than the world's first nonbinary voice for technology," *Interactions*, vol. 26, no. 6, pp. 56–59, 2019.

[25] J. Gay, J. C. Pepusch, and W. Nicholson, "Polly," 2023. [Online]. Available: https://docs.aws.amazon.com/polly/latest/dg/refphoneme-tables-shell.html

[26] J. Shen, R. Pang, R. J. Weiss, M. Schuster, N. Jaitly, Z. Yang, Z. Chen, Y. Zhang, Y. Wang, R. Skerrv-Ryan, *et al.*, "Natural TTS synthesis by conditioning wavenet on mel spectrogram predictions," in *Proc. ICASSP*, 2018, pp. 4779–4783.

[27] K. Ito and L. Johnson, "The LJ speech dataset," https://keithito.com/LJ-Speech-Dataset/, 2017.

[28] R. Zandie, M. H. Mahoor, J. Madsen, and E. S. Emamian, "Ryanspeech: A corpus for conversational text-to-speech synthesis," in *Proc. Interspeech*, 2021.

[29] D. Rizhinashvili, A. H. Sham, and G. Anbarjafari, "Gender neutralisation for unbiased speech synthesising," *Electronics*, vol. 11, no. 10, p. 1594, 2022.

[30] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "Darpa timit acoustic-phonetic continous speech corpus cd-rom. nist speech disc 1-1.1," *NASA STI/Recon technical report n*, vol. 93, p. 27403, 1993.

[31] C. Busso, S. Lee, and S. S. Narayanan, "Using neutral speech models for emotional speech analysis." in *Proc. Interspeech*, 2007, pp. 2225–2228.

[32] E. H. Jung, T. F. Waddell, and S. S. Sundar, "Feminizing robots: User responses to gender cues on robot body and screen," in *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, 2016, pp. 3107–3113.

[33] B. Kim, E. de Visser, and E. Phillips, "Two uncanny valleys: Re-evaluating the uncanny valley across the full spectrum of real-world human-like robots," *Computers in Human Behavior*, vol. 135, p. 107340, 2022.

[34] M. K. Strait, C. Aguillon, V. Contreras, and N. Garcia, "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, 2017, pp. 1418–1423.

[35] M. West, R. Kraut, and H. Ei Chew, "I'd blush if I could: closing gender divides in digital skills through education," Tech. Rep., 2019. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1

[36] E. Roesler, L. Naendrup-Poell, D. Manzey, and L. Onnasch, "Why context matters: The influence of application domain on preferred degree of anthropomorphism and gender attribution in human–robot interaction," *International Journal of Social Robotics*, pp. 1–12, 2022.

[37] E. Roesler, M. Heuring, and L. Onnasch, "(hu) man-like robots: The impact of anthropomorphism and language on perceived robot gender," *International Journal of Social Robotics*, pp. 1–12, 2023.

[38] S. Harnad, *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, 1987.

[39] R. K. Moore, "A bayesian explanation of the 'uncanny valley'effect and related psychological phenomena," *Scientific reports*, vol. 2, no. 1, pp. 1–5, 2012.

[40] B. De Gelder and P. Bertelson, "Multisensory integration, perception and ecological validity," *Trends in cognitive sciences*, vol. 7, no. 10, pp. 460–467, 2003.

[41] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, no. 5588, pp. 746–748, 1976.

[42] D. J. Rea, Y. Wang, and J. E. Young, "Check your stereotypes at the door: an analysis of gender typecasts in social human-robot interaction," in *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7*. Springer, 2015, pp. 554–563.

[43] A. Danielescu, "Eschewing gender stereotypes in voice assistants to promote inclusion," in *Proc. CUI*, 2020, pp. 1–3.

[44] M. Hope and J. Lilley, "Cues for perception of gender in synthetic voices and the role of identity." in *Proc. Interspeech*, 2020, pp. 4143–4147.

[45] S. Mooshammer and K. Etzrodt, "Gender ambiguity in voice-based assistants: Gender perception and influences of context," *Human-Machine Communication*, vol. 5, no. 1, p. 2, 2022.