OPEN ACCESS
OPUS AUGSBURG

Universität Augsburg
Universitätsbibliothek

Evaluating gender ambiguity, novelty and anthropomorphism in humming and talking voices for robots

Johanna Magdalena Kuch, Jauwairia Nasir, Silvan Mertes, Ruben Schlagowski, Christian Becker-Asano, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Kuch, Johanna Magdalena, Jauwairia Nasir, Silvan Mertes, Ruben Schlagowski, Christian Becker-Asano, and Elisabeth André. 2024. "Evaluating gender ambiguity, novelty and anthropomorphism in humming and talking voices for robots." In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN), 26-30 August 2024, Pasadena, CA, USA*, 2219–25. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/ro-man60168.2024.10731423.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
Deutsches Urheberrecht
Weitere Informationen finden Sie unter: / For more information see:
https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/

2024 33rd IEEE International Conference on
Robot and Human Interactive Communication (RO-MAN)
August 26-30, 2024. Pasadena, California, USA.

Evaluating Gender Ambiguity, Novelty and Anthropomorphism in Humming and Talking Voices for Robots

Johanna Magdalena Kuch[1], Jauwairia Nasir[1], Silvan Mertes[1], Ruben Schlagowski[1], Christian Becker-Asano[2] and Elisabeth André[1]

Abstract— **This paper investigates the effects of gender neutralization on the perception of anthropomorphism, gender specificity, and novelty for human voices, comparing spoken and hummed voice modalities. We evaluated gender-neutralized and original voice samples in both spoken and hummed formats using an online survey. Our results confirm that gender-neutralizing filters effectively reduce perceived gender specificity in both modalities, supporting their use in creating gender-neutral voices for humanoid robots. Hummed voices were perceived as more anthropomorphic and less novel than spoken voices, suggesting that non-verbal sound modalities can enhance the human likeness of gender-neutral androids while maintaining gender ambiguity. The study contributes to HRI by highlighting the potential of humming to fulfill users' expectations of interaction with android robots.**

I. INTRODUCTION

The sound design of robots significantly impacts how humans perceive and interact with them [1], [2]. A suitable sound design can lead to a robot being perceived more positively, for example, in terms of acceptance [3] or empathy [4]. At the same time, a mismatch between the sound design of a robot and its visual appearance can reduce the quality of interaction [5] and make it appear uncanny [6]. These circumstances emphasize the need to develop individually suitable sound profiles for different robots, making this a relevant field of research in HRI [7].

This paper aims to investigate two main research questions: 1. Is gender-neutralization through manipulating pitch and formants effective for both spoken and hummed voice samples? 2. How does gender ambiguity in voice samples influence the perception of anthropomorphism and novelty?

Previous studies show that non-verbal sounds are essential in HRI, influencing emotion, interaction, or behavior [8], [9], [10]. However, existing work has mainly focused on non-human sounds [1], [11], [12], making current knowledge not applicable to anthropomorphic robots like ultra-realistic androids. With this work, we want to explore human-like non-verbal sound modalities that can be used in that case.

Humming is a language-independent [13], [14] and human-like non-verbal modality, which can allow more precise analysis of the influence of voice characteristics on perceptions of anthropomorphism and emotional expression [15], [16]. As previous work has indicated that musical utterances can evoke empathy and prosocial behavior [4], we investigate humming as a future means of interaction with ultra-realistic robots in this study. In ultra-realistic robots, sound design must be anthropomorphic to meet high user expectations. Since a robot's voice affects gender perception [17], it's vital to match the voice to the robot's gender, especially for gender-ambiguous androids. The use of standard gender-neutralization techniques [18] needs further exploration for humming.

To close this research gap, we recorded and then neutralized the gender characteristics of human humming and talking sounds. We then evaluated these in an online study to measure the effects of gender neutralization on the perception of anthropomorphism, gender specificity, and novelty. We hypothesize that gender neutralization leads to a comparable reduction in perceived gender specificity in both humming and speech. Previously it hast been shown that gender-neutral voices are perceived as comparatively less anthropomorphic than gender-specific voices and hypothesized that a cause might be the increased perceived novelty of such gender-ambiguous voices [19]. Thus, in this study, we investigate the perceived novelty of voices across all conditions to examine how anthropomorphism, gender specificity, and novelty correlate. Furthermore, we collected qualitative data to contextualize our findings on gender ambiguity in sound design for HRI.

II. RELATED WORK

Previous research has shed light on how the sounds and appearance of robots influence their perception. In the following, we provide an overview of relevant work focusing on non-verbal sound design and humming, the perception of anthropomorphism, and gender ambiguity in HRI.

A. Non-verbal Sound Design in HRI

The design of sound in HRI has been established as an essential aspect of HRI. Previous studies have developed adaptive approaches to robot sound design based on non-expert human feedback [11], [20]. Studies on non-verbal communication in HRI emphasize the need for adaptive and context-aware sound design [21]. These works highlight the importance of tailoring robot voices to individual user preferences and emotional expressiveness [21]. These works highlight the importance of adapting robot voices to individual user preferences and emotional expressiveness.

B. Humming as a Non-Verbal Modality

Ishi et al. [16] and Suzuki et al. [15] have investigated the role of prosodic and voice quality characteristics in non-

[1]Chair for Human-Centered Artificial Intelligence, University of Augsburg, 86163 Augsburg, Germany
[2]Institute for Applied Artificial Intelligence, Stuttgart Media University, 70569 Stuttgart, Germany

verbal communication, showing that variations in prosodic features, such as speed and pitch, can significantly affect human perception and behavior. Jin et al. and Patil et al. provide insights into the unique characteristics of humming as a biometric and communicative tool, emphasizing the individual nature of non-verbal sound modalities [13], [14]. This research suggests that humming can be manipulated similarly to talking in terms of pitch and formants to achieve gender ambiguity. And shows potential of humming, as a language-independent and human-like non-verbal modality, to influence perceptions of anthropomorphism.

C. Anthropomorphism and Voice Perception in Robots

Anthropomorphism significantly influences the acceptance and interaction quality of robots by humans. Trovato et al. [22] and Kuhne et al. [23] show that cultural and human-like aspects can significantly influence the acceptance and perception of robots. They point out that a human-like voice, even if less appropriate, can make the robot seem more familiar and that human voices are generally preferred over synthetic voices. The research by Moore [24] and, building on this, Meah et al. [6] and Mitchell [25] show that inconsistencies in the human reality of robot features, such as voice and appearance, can lead to discomfort and rejection, known as the "Uncanny Valley" phenomenon. These studies emphasize the need for coordinated and consistent design of robot attributes. Roesler [26] and Mooshammer [27] complement these findings with their research on the influence of the application domain on the preferred degree of anthropomorphism and the perception of gender-neutral voices. They show that the context of robot use and the gender neutrality of the voice are critical factors for the acceptance and design of robots. Eyssel et al. [3] show that matching the robot's and user's gender, primarily when the robot uses a human-like voice, leads to a more positive perception and stronger psychological closeness. Steinhaeusser et al. [28] build on this by investigating the influence of a robot's voice on storytelling, highlighting the importance of anthropomorphic voice design in enhancing user relationships. Fink [29] and Schreibelmayr et al. [30] further emphasize the positive effects of anthropomorphism and human-like features on the acceptance of robots, particularly in social contexts.

D. Gender Ambiguity in Robot Voices

The complexity of gender perception in HRI is fundamental to our study. Seaborn et al. [31] and Sutton [32] highlight the importance of considering gender ambiguity to avoid projecting human gender models onto robots. Chang et al. [33] and Leung [34] provide insights into how gender-neutral or ambiguous auditory cues influence perceptions of gender in robotic interactions. Torre et al. [2] demonstrate that a gender-neutral voice can reduce the assignment of stereotypical gender attributes, supporting our approach to sound design. Mooshammer et al. [35] and Rizhinashvili et al. [18] present a technique for generating acoustically gender-neutral voices. Their methodology provides a basis for our study, which extends their work by investigating non-verbal sounds such as humming.

E. Novelty in Human-Robot Interaction

The perception of novelty in HRI is a critical factor that affects user acceptance and interaction quality. Smedegaard [36] argues that research on psychological novelty effects within Social Robotics and Human-Robot Interaction (SHRI) has been fragmented and heterogeneous. In this paper, we contribute to closing this gap by explicitly investigating novelty and its connections with anthropomorphism and gender ambiguity. Our investigation of gender ambiguity in HRI critically builds on the previous work [19], that investigated the effects of gender neutralization on the perception of anthropomorphism in speech signals. It was found that gender-neutral voices are perceived as less anthropomorphic than gender-specific voices, a phenomenon they possibly attributed to the novelty of such voices to the listener. We extend this line of inquiry by investigating whether the perceived novelty of non-verbal, human-like sound modalities, such as humming, similarly influences the perception of anthropomorphism and gender specificity. More concretely, building on previous findings on the novel perception of gender-neutral voices, we make an essential contribution by investigating how gender ambiguity influences the perception of human-like non-verbal sounds, particularly melodic humming. Specifically, using a mixed methods approach, we evaluate the effects of sound modulation on perceived anthropomorphism, gender specificity, and novelty to obtain actionable insights for the design of gender-ambiguous anthropomorphic voices.

III. Methodology

This section describes in detail the procedure for creating different voice samples and their subsequent evaluation in a mixed-methods study.

A. Creation of the voice samples

To generate the spoken voice samples, recordings of four adults (two men and two women) were made in a studio setting to ensure consistent audio quality. Each person first read out the "Harvard sentences" [37], a standardized text set used in phonetic research and speech quality assessment. The first seven sentences of each speaker were used, with each piece having a length of around 25 to 30 seconds. For the humming samples, they then hummed the melody of the children's song "Little Hans". Again, each sample had a length of around 25 to 30 seconds. For the gender-neutralized samples, the initial recordings were processed using a filter presented by [18] to remove gender-specific characteristics while maintaining other aspects of the voice. This filter works by adjusting the pitch and formant frequencies of the voice to lie within a range that is perceived as neither distinctly male nor female. Specifically, the pitch is shifted to a median frequency between typical male and female pitches, and the formant frequencies are adjusted to neutralize resonances that are characteristic of male or female voices. As a result, we

obtained eight spoken voice samples, four of which were categorized as gender-specific (i.e., either male or female) and four as gender ambiguous. Analoguosly, we obtained eight hummed samples, where again four were categorized as gender-specific and four as gender-ambiguous[3].

During a pilot study (online survey, n=9), we identified that filtering produced artifacts (cracking sounds) in the voice samples that made unfiltered and filtered samples less comparable. To ensure comparability and control for the artifacts introduced by the filtering process, all voice samples, including the unfiltered ones, were subjected to the same level of artificial distortion. Therefore we applied a gain of 10 dB to the audio signal, overlaid the audio with white noise, adjusted to a noise level of 25 dB, reduced the sample rate of the audio to 6 kHz, applied a bandpass filter with a low cutoff frequency of 500 Hz and a high cutoff frequency of 5000 Hz and finally muted random frequency bands by applying silence to 100 bands, each with a silence duration of up to 200 ms.

B. Structure of the study

The study was conducted in the form of an online survey via Amazon Mechanical Turk. The created voice samples were rated in terms of their perceived gender specificity, anthropomorphism and novelty. Participants were also asked to characterize each voice in three words by imagining a character with that voice ("Imagine a character with this voice and describe it in three words."). The evaluation covered several aspects:

Gender specificity: Participants rated separately on a scale from 1 to 5 how well the voice corresponded to the genders female, male, and neutral. First, the neutrality scale was used to calculate a *specificity factor*, cf. Equation 1, which ranges from 0 (least specific) to 1 (most specific). It is then used to scale the absolute difference of the *male* and *female* scales (for the same sample) to calculate each sample's gender specificity from 0 to 4, cf. Equation 2.

Anthropomorphism: An average over the items of the Godspeed questionnaire [38] is calculated to determine an overall degree of anthropomorphism for each sample. As the item "moving rigidly vs. moving elegantly" cannot be applied to a sound source, the terms at the end of the Likert scale were replaced with "rigid" and "elegant" respectively. Instead of rating the perception of a robot we asked: "Please rate your impression of the voice on these scales".

Novelty: Two Likert scales from 1 to 5, representing the opposites familiar vs. unfamiliar and usual vs. unusual, were used with the overall mean value as a final measure of a sample's perceived novelty.

$$specificity\_factor = 1 - \frac{neutral - 1}{4} \quad (1)$$

$$gender\_specificity = (|m - f|) \times specificity\_factor \quad (2)$$

To minimize fatigue effects for each participant, each participant only evaluated 8 of the 16 voice samples in random order. Four from an originally female, and four from an originally male voice. In addition, attention checks were conducted at three points in the survey to test participants' attention and ensure their engagement. On average, participants took half an hour to complete the survey.

[3]All audio files used in the online study can be found here: https://git.rz.uni-augsburg.de/kuchjoha/ro-man24˙files.git

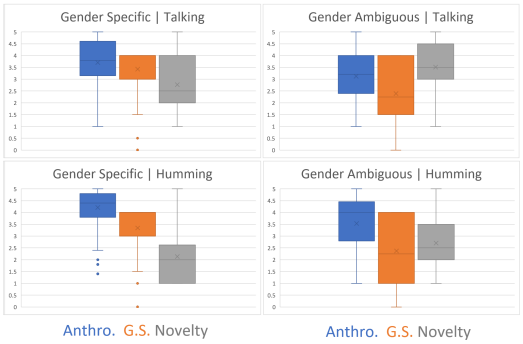

Fig. 1. Boxplots of Anthropomorphism (values from 1-5), Gender specificity (values from 0-4) and Novelty (values from 1-5) for each Condition

C. Study participants

The participants were recruited via Amazon Mechanical Turk and consisted exclusively of native English speakers from North America and the UK. After careful screening and elimination of invalid results, valid responses from 49 individuals remained for analysis.

D. Data analysis

First, we wanted to evaluate how the filter affects the perception of gender specificity, anthropomorphism, and novelty both humming and talking samples. As such, we tested for differences between perception of anthropomorphism of original vs. filtered *talking*, gender specificity of original vs. filtered *talking*, novelty of original vs. filtered *talking*, anthropomorphism of original vs. filtered *humming*, gender specificity of original vs. filtered *humming* and novelty of original vs. filtered *humming* samples.

Further, to test if there are differences between talking and humming samples, we tested for differences between perception of anthropomorphism of humming vs. talking, gender specificity of humming vs. talking and novelty of humming vs. talking samples.

For the qualitative answers we conducted a qualitative content analysis [39] to determine which specific terms were mentioned more than five times and in which contexts they appeared particularly often in order to gain deeper insights into the participants' perceptions and interpretations.

IV. Results

Boxplots of the participants' answers are given in Figure 1. The quantitative and qualitative results of the online survey are presented below.

A. Quantitative results

The quantitative results are based on the aggregated ratings of the participants, which were carried out using the scales provided to assess the gender specificity, anthropomorphism and novelty of the voices.

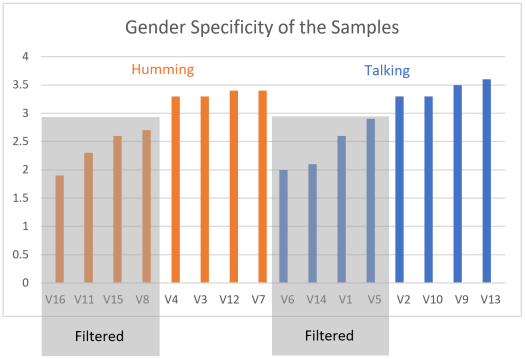

Fig. 2. Rating of the Gender Specificity for Talking and Humming

1) Filtered vs. original voice samples: To determine how gender specific the voice samples were perceived, we calculated the mean of the gender specificity score from all responses for each voice. These values were used to analyze and compare the gender specificity of the individual voices. The results of this analysis are shown in Figure 2. As can be seen, the gender specificity of humming and talking samples is in a similar range. The filtered voice samples were consistently perceived as less gender-specific compared to the original recordings. Due to a lack of normality of our data, Wilcoxon tests were applied for all significance tests. A Bonferroni correction was applied to all nine Wilcoxon tests setting the level of significance to 0.0056.

Variables Original vs. Filtered	p	Effect Size
Talking Anthropomorphism	$< .0001$	0.752
Talking Gender Specificity	$< .0001$	0.865
Talking Novelty	$< .0001$	-0.745
Humming Anthropomorphism	$< .0001$	0.743
Humming Gender Specificity	$< .0001$	0.750
Humming Novelty	0.0005	-0.644

TABLE I
WILCOXON TEST (ORIGINAL AND FILTERED VOICES)

The results of the first six tests (i.e., the tests comparing original vs. filtered samples) are shown in Table I. In all tests, we found a significant difference between original and filtered samples, suggesting that the filter is effective in both talking and humming samples.

2) Humming vs. talking: The results of the three tests comparing humming and talking are shown in Table II.

Variable Talking vs. Humming	p	Effect Size
Anthropomorphism	$< .001$	-0.425
Gender Specificity	0.700	-
Novelty	$< .001$	0.606

TABLE II
WILCOXON TEST (TALKING AND HUMMING)

The analysis of the anthropomorphism values between the talking and humming conditions using the Wilcoxon test showed a significant difference ($p < .001$), with an effect size of -0.425, indicating a medium to strong negative effect. This means that the participants perceived the humming voices as significantly more anthropomorphic than the talking voices. For gender specificity, however, we did not find a significant difference between the talking and humming conditions. This suggests that the type of vocalization - whether spoken or hummed - has no substantial influence on the perception of gender specificity, which also explains why the filter works well in both conditions. Finally, the examination of novelty showed clear differences between the two conditions, with a significant Wilcoxon test result ($p < .001$). The effect size was 0.606, indicating a strong positive effect. This means that the humming voices were perceived as significantly less novel than the talking voices.

3) Explorative analysis of correlations between the Variables: In an explorative manner, we analyzed the correlations between anthropomorphism, novelty, and gender specificity to identify potential relationships and to distinguish differences between the talking and humming conditions. Based on preliminary observations, we hypothesized a positive correlation between gender specificity and anthropomorphism, and a negative correlation of both variables with novelty. Given the non-parametric distribution of our data, we employed Spearman's Rho for this analysis. This approach allowed us to uncover underlying trends and distinctions in how voices are perceived across different vocal expressions.

Talking		
	Anthropomorphism	**Gender specificity**
Gender specificity		
Spearman's Rho	0.240	—
p-value	$< .001$	
Novelty		
Spearman's Rho	-0.535	-0.215
p-value	$< .001$	0.002
Humming		
	Anthropomorphism	**Gender specificity**
Gender specificity		
Spearman's Rho	0.275	—
p-value	$< .001$	
Novelty		
Spearman's Rho	-0.654	-0.364
p-value	$< .001$	$< .001$

TABLE III
CORRELATIONS FOR TALKING AND HUMMING SAMPLES

In our study, we explored significant correlations between anthropomorphism, gender specificity, and novelty in both conditions, talking and humming.

Talking Conditions: We found a weak positive correlation between anthropomorphism, and gender specificity (Spearman's rho = 0.240, $p < .001$), indicating that voices recognized as more human-like tend to have slightly higher gender-specific attributes. There was a strong negative correlation between anthropomorphism and novelty (Spearman's rho = -0.535, $p < .001$), suggesting that the more human-like the voices are perceived, the less novel they seem.

Humming Conditions: A moderate positive correlation was noted between anthropomorphism, and gender specificity (Spearman's rho = 0.275, $p < .001$). A very strong negative correlation emerged between anthropomorphism and novelty (Spearman's rho = -0.654, $p < .001$), indicating

a significant decrease in perceived novelty as voices are perceived to be more human-like.

The data suggest stronger correlations between anthropomorphism and novelty in humming conditions compared to talking. The patterns between gender specificity and novelty are consistently negative in both conditions.

B. Qualitative results

The results of this study suggest that certain attributes show universal and gender-specific tendencies in the perception of voice samples (TableIV). In particular, the attribute ”happy” stands out with a total of 34 mentions, which is predominantly the case in hummed voice samples. This indicates that a cheerful mood is seen as universal and is less linked to a specific gender. Interestingly, ”male” is mainly mentioned in hummed, gender-ambiguous voices, indicating a certain association of humming sounds with male identity, even without clear gender-specific characteristics. The even distribution of the terms ”calm” and ”kind” across all conditions - in both humming and speaking and regardless of gender - suggests their universal characteristics in voice perception. These results emphasize the general view of these characteristics as fundamental to human interaction, regardless of the voice pitch or gender of the speaker.

Word	frequency	ambiguous		specific	
		humming	talking	humming	talking
happy	34	16	1	17	0
male	32	11	11	7	3
calm	27	8	6	6	7
kind	24	5	5	7	7
friendly	22	8	3	8	3
old	17	0	10	1	6
musical	16	8	0	8	0
foreign	16	0	9	0	7
accent	15	0	7	0	8
educate	14	0	5	0	9
serious	14	0	5	0	9
bore	14	6	3	4	1
feminine	13	0	0	5	8
female	13	0	1	5	7
young	13	3	0	8	2
care	13	4	5	2	2
intelligent	12	0	5	1	6
helpful	12	0	6	1	5
thoughtful	12	4	1	3	4
relax	11	5	0	5	1
masculine	11	1	2	5	3
smart	9	0	3	0	6

TABLE IV
MOST FREQUENT WORDS USED TO DESCRIBE THE VOICES

In contrast, the attribution of education and seriousness, represented by the terms ”educated” and ”serious”, is observed primarily in spoken voices. This could indicate that these attributes are more strongly related to the content and nature of the spoken language than to the pure tone of voice.

Examining the frequency of common descriptions in the voice samples also shows patterns: in the humming samples, attributes such as ”male”, ”happy”, ”friendly”, ”musical” and ”relax” were each mentioned at least five times by five different people, resulting in a total of 93 common descriptions. In the talking samples, other attributes such as ”old”, ”foreign”, ”accent”, ”educated”, ”serious”, ”intelligent” and ”helpful” were mentioned more frequently, with a total of 97 common descriptions. These results highlight the different associations with hummed versus spoken voices and emphasize the role of language and tone in shaping perceptions.

V. DISCUSSION

A. Interpretation of main findings

The results of our study provide important insights into the perception of anthropomorphic voices and the effects of a filter that manipulates pitch and formants for gender neutralization on gender specificity. Our hypothesis was confirmed that the filter works equally well in both humming and talking. This result shows that the filter is suitable for significantly reducing gender specificity in both cases, not just talking as previously shown by Kuch et al. [19].

In terms of anthropomorphism, gender specificity, and novelty, we found that humming is perceived as significantly more anthropomorphic and less novel. These results are particularly promising as they indicate that humming can achieve an anthropomorphic yet gender-ambiguous voice, which is relevant for ultra-realistic, gender-ambiguous androids. However, it remains to be tested whether the applied voice has the desired effects on the robot head. It is also relevant that we found no significant difference in gender specificity between humming and talking. This finding allows us to focus on the gender perception of the robot head without worrying about the modality of the voice.

We found the expected correlations in terms of novelty in both conditions. Gender-ambiguous voices were perceived as less anthropomorphic than gender-specific voices, which is most likely due to their novelty. Interestingly, the correlations, particularly between novelty and anthropomorphism and novelty and gender specificity, were slightly stronger for humming. This difference suggests that it may be worthwhile to use occasional humming to precisely manipulate certain aspects, such as anthropomorphism or gender specificity, even when otherwise working with talking.

In our qualitative analysis, we found that humming tended to be described in terms related to emotional state, whereas talking tended to be associated more with cognitive characteristics or gender of the voice. This finding suggests that, depending on what is to be emphasized, humming or talking can be used specifically in the design of robot sounds.

B. Practical implications and application areas

Our research results suggest that gender-ambiguous voices, mainly through the tested gender-neutralization filter, significantly impact the perception of humanoid robots. When applied to humming and speaking voices, the filter’s identified equivalence underlines its universal applicability and emphasizes the need to extend design considerations beyond traditional spoken voice output.

The higher anthropomorphic perception of hums compared to spoken voices implies a precise strategic application of this sound modality in robotics to increase user acceptance in sensitive areas such as care or customer care. Integrating

humming sounds could be particularly important in the development of assistive robots for therapeutic purposes or stress reduction by providing a more subtle, emotionally soothing form of communication. Furthermore, our findings on reservations about novel voices highlight the critical balance between innovation and user familiarity. This finding about novelty points to a key design consideration: harmonizing experimental soundscapes with established user expectations is necessary to minimize dissonance in perception and acceptance. The qualitative analysis also differentiates between emotional and cognitive associations about humming versus speaking voices, which is crucial for the targeted development of robot audio outputs in different application contexts. This differentiated perception should be incorporated into speech design to adapt robots more effectively to their respective fields of application and the associated user expectations.

C. Limitations of the study

This study has several limitations that should be noted. First, the participant base was limited to English speakers from North America and the UK, which may limit the generalizability of the results across different linguistic and cultural backgrounds. The voices used in our research came from individuals who are not native English speakers, some with noticeable accents, which could have affected perceptions of the spoken voices. It would be beneficial for future studies to include a more diverse range of speakers with different backgrounds to better understand the impact of accents on voice perception. In addition, gender-neutralizing the voice samples resulted in audio artifacts that were carried over to all samples to keep them comparable, which may have influenced participants' perceptions. A further limitation arises from the online setting of the study, which only displayed the audio files for evaluation. This means there was no direct physical interaction between the participants and robots, neglecting the interplay between auditory, visual, and tactile elements; hence, potentially limiting the transferability of the results to real-life interactions. Further, we only used the melody "Little Hans", known for its positive connotations, which might have influenced how people perceived the hummed voices. Future research should explore a variety of melodies in different musical keys for a more balanced view. Another limitation relates to the use of the Godspeed questionnaire to measure anthropomorphism. Although widely used in HRI research, the Godspeed questionnaire has known pitfalls, such as its limited ability to capture the full complexity of anthropomorphism and its potential biases in self-reported measures. These limitations should be acknowledged when interpreting the results, and future studies should consider complementing the Godspeed questionnaire with other validated measures or observational data to gain a more comprehensive understanding of anthropomorphism.

D. Comparison with related work

By exploring the gender-neutral design of speech and hum sounds, we contribute to the dialog initiated by Eyssel et al. and Trovato et al. on user perception and adaptation of robotic voices [3], [22]. Our research shows that hums modified for gender neutrality are perceived as less gender-specific and expand into non-verbal sound domains with conversation, making our approach relevant for inclusive sound profiling, especially for gender-ambiguous robots. Building on discussions of anthropomorphism, our study extends the concept to non-verbal sounds. In contrast to previous work focusing mainly on linguistic features [3], [30], [26], our results show that non-verbal sounds, such as humming, can also enhance perceived human likeness. This suggests that sound design in robotics can go beyond spoken words and encompass a broader range of human-like sounds. Furthermore, our approach meets the demands of Ishi et al. [16] and Suzuki et al. [15] for a deeper understanding of the effects of melodic humming on the perception of humanlike voices. Lastly, we extend Kuch et al.'s [19] findings to the non-verbal domain demonstrating that gender neutralization techniques applied to speech [18] can also be effectively applied to non-verbal sounds such as humming.

VI. Conclusion

In this paper, we explored the impact of gender-neutralizing voice filters on the perception of anthropomorphism, gender specificity, and novelty in gender-specific and gender-ambiguous voices, focusing on both spoken and hummed voices. Our study confirmed that a gender-neutralizing filter, manipulating pitch and formants, significantly reduces perceived gender specificity across both vocal modalities (i.e., humming and talking), supporting the filter's applicability for creating gender-ambiguous voices for humanoid robots. Hummed voices were perceived as more anthropomorphic and less novel than spoken voices, suggesting that non-verbal sound modalities like humming can enhance the human likeness of ultra-realistic, gender-ambiguous androids while maintaining gender ambiguity. Future work has to investigate the applicability of the developed voices in scenarios involving real android robots. Our results highlight the need to consider non-verbal sounds in robot design to meet user expectations for human-like interaction.

ACKNOWLEDGMENT

This work was funded by Deutsche Forschungsgemeinschaft (DFG) under project number 442607480, PANORAMA.

References

[1] B. J. Zhang and N. T. Fitter, "Nonverbal sound in human-robot interaction: A systematic review," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 4, pp. 1–46, 2023.

[2] I. Torre, E. Lagerstedt, N. Dennler, K. Seaborn, I. Leite, and É. Székely, "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)*, pp. 106–112, IEEE, 2023.

[3] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. de Ruiter, and F. Hegel, "'if you sound like me, you must be more human'," in *Proceedings of the seventh annual ACMIEEE international conference on Human-Robot Interaction* (H. Yanco, ed.), ACM Conferences, (New York, NY), p. 125, ACM, 2012.

[4] T. Höfker, "Musical utterances to evoke empathy and prosocial behavior toward a hospital robot," 2023.

[5] B. Sarigul, I. Saltik, B. Hokelek, and B. A. Urgen, "Does the appearance of an agent affect how we perceive his/her voice?," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (T. Belpaeme, J. Young, H. Gunes, and L. Riek, eds.), (New York, NY, USA), pp. 430–432, ACM, 2020.

[6] L. F. S. Meah and R. K. Moore, "The uncanny valley: A focus on misaligned cues," in *Social Robotics* (M. Beetz, B. Johnston, and M.-A. Williams, eds.), vol. 8755 of *Lecture Notes in Computer Science*, pp. 256–265, Cham: Springer Nature, 2014.

[7] P. van Rijn, S. Mertes, K. Janowski, K. Weitz, N. Jacoby, and E. André, "Giving robots a voice: Human-in-the-loop voice creation and open-ended labeling," 2024.

[8] E. Frid and R. Bresin, "Perceptual evaluation of blended sonification of mechanical robot sounds produced by emotionally expressive gestures: Augmenting consequential sounds to improve non-verbal robot communication," *International Journal of Social Robotics*, vol. 14, no. 2, pp. 357–372, 2022.

[9] C. Bethel, ed., *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ACM Digital Library, (New York,NY,United States), Association for Computing Machinery, 2021.

[10] S. Saunderson and G. Nejat, "How robots influence humans: A survey of nonverbal communication in social human–robot interaction," *International Journal of Social Robotics*, vol. 11, no. 4, pp. 575–608, 2019.

[11] H. Ritschel, I. Aslan, S. Mertes, A. Seiderer, and E. Andre, "Personalized synthesis of intentional and emotional non-verbal sounds for social robots," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, (Piscataway, NJ), pp. 1–7, IEEE, 2019.

[12] Fernández De Gorostiza luengo, Javier, F. Alonso Martín, Á. Castro-González, and M. Á. Salichs, "Sound synthesis for communicating nonverbal expressive cues," *IEEE Access*, vol. 5, pp. 1941–1957, 2017.

[13] M. Jin, J. Kim, and C. D. Yoo, "Humming-based human verification and identification," in *2009 IEEE International Conference on Acoustics, Speech, and Signal Processing*, (Piscataway, NJ), pp. 1453–1456, IEEE, 2009.

[14] H. A. Patil, M. C. Madhavi, and N. H. Chhayani, "Person recognition using humming, singing and speech," in *2012 International Conference on Asian Language Processing (IALP 2012)* (D. Xiong, ed.), (Piscataway, NJ), pp. 149–152, IEEE, 2012.

[15] N. Suzuki, K. Kakehi, Y. Takeuchi, and M. Okada, "Social effects of the speed of hummed sounds on human–computer interaction," *International Journal of Human-Computer Studies*, vol. 60, no. 4, pp. 455–468, 2004.

[16] C. T. Ishi, H. Ishiguro, and N. Hagita, "Evaluation of prosodic and voice quality features on automatic extraction of paralinguistic information," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 374–379, 2006.

[17] 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 *HRI'19*, (Piscataway, NJ), pp. 211–221, IEEE, 2019.

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

[19] J. M. Kuch, F. Melchior, and C. Becker-Asano, "Effects of gender neutralization on the anthropomorphism of natural and synthetic voices," in *Proc. of 32nd IEEE Int. Conf. on Robot and Human Interactive Communication (RO-MAN)*, pp. 2080–2085, 2023.

[20] M. Schwenk and K. O. Arras, "R2-d2 reloaded: A flexible sound synthesis system for sonic human-robot interaction design," in *The 23rd IEEE international symposium on robot and human interactive communication*, pp. 161–167, 2014.

[21] R. Read and T. Belpaeme, "People interpret robotic non-linguistic utterances categorically," *International Journal of Social Robotics*, vol. 8, no. 1, pp. 31–50, 2016.

[22] G. Trovato, J. G. Ramos, H. Azevedo, A. Moroni, S. Magossi, H. Ishii, R. Simmons, and A. Takanishi, "Designing a receptionist robot: Effect of voice and appearance on anthropomorphism," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2015)*, (Piscataway, NJ), pp. 235–240, IEEE, 2015.

[23] K. Kühne, M. H. Fischer, and Y. Zhou, "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*, vol. 14, p. 593732, 2020.

[24] R. K. Moore, "A bayesian explanation of the 'uncanny valley'effect and related psychological phenomena," *Scientific reports*, vol. 2, no. 1, p. 864, 2012.

[25] W. J. Mitchell, K. A. Szerszen, A. S. Lu, P. W. Schermerhorn, M. Scheutz, and K. F. Macdorman, "A mismatch in the human realism of face and voice produces an uncanny valley," *i-Perception*, vol. 2, no. 1, pp. 10–12, 2011.

[26] 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*, vol. 14, no. 5, pp. 1155–1166, 2022.

[27] S. Mooshammer and K. Etzrodt, "Gender ambiguity in voice-based assistants: Gender perception and influences of context," *Human-Machine Communication*, vol. 5, pp. 49–74, 2022.

[28] S. C. Steinhaeusser, P. Schaper, O. Bediako Akuffo, P. Friedrich, J. Ön, and B. Lugrin, "Anthropomorphize me!," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (C. Bethel, ed.), ACM Digital Library, (New York,NY,United States), pp. 529–534, Association for Computing Machinery, 2021.

[29] J. Fink, "Anthropomorphism and human likeness in the design of robots and human-robot interaction," in *Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012. Proceedings 4*, pp. 199–208, 2012.

[30] S. Schreibelmayr and M. Mara, "Robot voices in daily life: Vocal human-likeness and application context as determinants of user acceptance," *Frontiers in psychology*, vol. 13, p. 787499, 2022.

[31] K. Seaborn and P. Pennefather, "Neither hear nor their: Interrogating gender neutrality in robots," *Proceedings of the*.

[32] S. J. Sutton, "Gender ambiguous, not genderless," in *Proceedings of the 2nd Conference on Conversational User Interfaces* (M. I. Torres, ed.), ACM Digital Library, (New York,NY,United States), pp. 1–8, Association for Computing Machinery, 2020.

[33] Y.-C. Chang, D. J. Rea, and T. Kanda, "Investigating the impact of gender stereotypes in authority on avatar robots," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 106–115, 2024.

[34] Y. Leung, J. Oates, and S. P. Chan, "Voice, articulation, and prosody contribute to listener perceptions of speaker gender: A systematic review and meta-analysis," *Journal of Speech, Language, and Hearing Research*, vol. 61, no. 2, pp. 266–297, 2018.

[35] 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* (A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, E. Luger, M. Goodwin, and P. B. Brandtzaeg, eds.), vol. 13171 of *Lecture Notes in Computer Science*, pp. 176–191, Cham: Springer International Publishing, 2022.

[36] C. V. Smedegaard, "Novelty knows no boundaries: Why a proper investigation of novelty effects within shri should begin by addressing the scientific plurality of the field," *Frontiers in robotics and AI*, vol. 9, p. 741478, 2022.

[37] "Ieee recommended practice for speech quality measurements," *IEEE Transactions on Audio and Electroacoustics*, vol. 17, no. 3, pp. 225–246, 1969.

[38] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[39] V. Braun and V. Clarke, "Thematic analysis," 2012.