Breaking the Binary: A Systematic Review of Gender-Ambiguous Voices in Human-Computer Interaction

Citation for the original published paper (version of record):
De Cet, M., Obaid, M., Torre, I. (2025). Breaking the Binary: A Systematic Review of Gender-Ambiguous Voices in Human-Computer
Interaction. Conference on Human Factors in Computing Systems - Proceedings.
http://dx.doi.org/10.1145/3706598.3713608

Breaking the Binary: A Systematic Review of Gender-Ambiguous Voices in Human-Computer Interaction

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

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

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

Abstract

Voice interfaces come in many forms in Human-Computer Interaction (HCI), such as voice assistants and robots. These are often gendered, i.e. they sound masculine or feminine. Recently, there has been a surge in creating gender-ambiguous voices, aiming to make voice interfaces more inclusive and less prone to stereotyping. In this paper, we present the first systematic review of research on gender-ambiguous voices in HCI literature, with an in-depth analysis of 36 articles. We report on the definition and availability of gender-ambiguous voices, creation methods, user perception and evaluation techniques. We conclude with several concrete action points: clarifying key terminology and definitions for terms such as gender-ambiguous, gender-neutral, and non-binary; conducting an initial acoustic analysis of gender-ambiguous voices; taking initial steps toward standardising evaluation metrics for these voices; establishing an open-source repository of gender-ambiguous voices; and developing a framework for their creation and use. These recommendations provide important insights for fostering the development and adoption of inclusive voice technologies.

CCS Concepts

• **Human-centered computing → Sound-based input / output**; *Auditory feedback*.

Keywords

gender-ambiguous, gender, ambiguous, gender-neutral, robot, agent, assistant, computer voice, Conversational User Interfaces

ACM Reference Format:
Martina De Cet, Mohammad Obaid, and Ilaria Torre. 2025. Breaking the Binary: A Systematic Review of Gender-Ambiguous Voices in Human-Computer Interaction. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3706598.3713608

1 Introduction

Since voice has been the main medium of communication between people for millennia, it is natural that new technologies are targeting voice-enabled devices as a promising avenue for efficient human-computer interactions, such as Voice Assistants (VAs) and robots. Nowadays, these interactions are increasingly prevalent in daily life, spanning personal, professional, and public domains. The widespread adoption of VAs is evident, with over 320 million in-home VAs installed globally and in the U.S. alone, around 35% of individuals own a VA [46]. Despite the technology's functionality, one consistent design choice has been the assignment of default female voices and names, such as "Siri" and "Alexa".

Historically, VAs have been designed to sound distinctly masculine or feminine. Indeed, assigning gender to agents is common in studies of Human-Computer Interaction (HCI) and has shown positive outcomes in terms of human attitudes, perceived trustworthiness, and user acceptance [14, 20]. One of the reasons for these positive outcomes is that it reinforces what is expected. Fridin and Belokopytov [24] investigated the acceptance of socially assistive robotics by teachers; their robot used a feminine voice, with the explanation that kindergarten staff are most often women. However, the practice of gendering voice assistants raises concerns. Torre et al. [80] suggest that assigning gender to voices may reinforce existing gender biases, propagate stereotypes, and potentially exclude gender-nonconforming people. Mahmood and Huang [45], reported that the association of commercial voice assistants with a female gender can perpetuate societal harm by reinforcing traditional stereotypes of women as submissive, and responsible for taking orders. This can contribute to a culture that normalises harassment, abuse, and discriminatory behaviour [26, 49]. Additionally, organisations like UNESCO have highlighted the risk that voice assistants might inadvertently amplify gender biases [82]. As societal awareness of gender diversity and inclusivity evolves, there is increasing interest in exploring more inclusive design alternatives.

In recent years, researchers have increasingly focused on developing new voice types that blend characteristics from both masculine and feminine voices, usually called "gender-neutral", "gender-ambiguous" or "non-binary". This phenomenon is not limited to research. Non-binary individuals –who do not fit neatly into the categories of "man" or "woman" or "male" or "female"– are increasingly seeking guidance from gender-affirming voice teachers to modulate their voices and achieve a more androgynous sound [41, 87]. For consistency, we will use the term "gender-ambiguous" throughout this paper to cover the range of terminology used in research when referring to such voices. Gender-ambiguous voices are designed to offer more inclusive options and to address potential biases related to gender representation. They can be used by virtual agents representing non-binary and transgender people. They can also be used as voices for embodied technologies such as robots to reduce

CHI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713608

robot gendering, which also risks propagating existing negative stereotypes [16, 80]. By introducing alternative voice options, researchers seek to expand user experiences and reflect a broader understanding of gender diversity in Human-Computer Interaction. Despite growing interest and advances in speech technology, little is known about gender-ambiguous voice perception. What do they sound like? Are they actually perceived as gender-ambiguous by listeners? And what does "gender-ambiguous" even mean?

This paper is structured as follows. First, we will delve deeper into the concept of gender neutrality (see Section 1.1). Subsequently, we will introduce our systematic review of the literature (see Section 1.2) and define key voice-technology terms to enhance the overall clarity and understanding of the paper (see Section 1.3). We will present the method followed in our systematic review in Section 2, the results in Section 3 and discuss implications and recommendations, as well as directions for future work, in Section 4.

1.1 Gender-Neutrality

Feminist theory, as exemplified by Gayle Rubin, distinguishes anatomical sex, defined by the biological characteristics of bodies, from gender, which encompasses the cultural constructs and norms associated with those characteristics. Rubin's work highlights that traits and roles traditionally linked to men and women are not innate but are shaped by societal and cultural influences [64]. Judith Butler, an American feminist philosopher, challenged traditional notions of gender and sexuality in their seminal work Gender Trouble [9]. Central to Butler's theory is that gender is not something one "has" but something one "does". Through their concept of "gender performativity", Butler argues that gender is constructed and maintained through the repetition of stylised acts, rather than being an inherent or fixed identity, suggesting that societal norms dictate and reinforce what is perceived as gendered behaviour.

This performative understanding of gender provides a critical lens for examining gender bias and stereotypes, which involve favouring one gender over another or arbitrarily assigning characteristics and roles based on gender. These biases are not only evident in human interactions but also manifest in how people engage with agents such as robots and chatbots, as research has shown [8, 73, 82]. Addressing these biases has led to growing interest in exploring alternative approaches to design agents that challenge stereotypes while maintaining credibility [86]. Furthermore, artificial gender-ambiguous voices in avatars or chatbots could be employed to perform non-binary gender identities, aligning with Butler's concept of gender as performative by deliberately disrupting binary expectations. Through such voices, nuanced and diverse expressions of gender could be enacted, including fluid, agender, or culturally specific non-binary identities, offering new ways to challenge and expand traditional understandings of gender in Human-Computer Interaction, or Computer-Mediated Communication.

The intersection of gender theory and linguistics offers further insights into how biases operate and evolve. Starting in the 1970s, linguistic studies were predominantly centred around the assumption that men's and women's languages were distinct [42]. This binary view reflected broader societal ideas of gender as opposites. However, feminist theory, especially Butler's work, prompted a shift in how these ideas are understood. By 1996, Bing and Bergvall [6] argued for moving beyond binary frameworks, pointing out that gender and biological sex exist on spectrums rather than as opposites. More recently, this line of inquiry has expanded to include non-binary and gender-neutral expressions, offering new linguistic options that challenge traditional gender constructs [30] in order to be more inclusive of gender-nonconforming people.

One such option, gender neutrality, has gained popularity as a method to disrupt stereotypes and avoid harmful associations in the field of HCI. The term "gender-neutral" refers to something not specifically linked to women or men and includes practices such as the removal of gender cues from language and design [10, 23]. Within HCI, the argument is that by eliminating all gender indicators, people may avoid assigning gender to agents, thereby mitigating the negative effects of gendering.

However, achieving gender neutrality does not come without challenges. Sutton [74] argues that artificial agents cannot be entirely gender-neutral, as humans tend to assign gender whenever human-like traits are present. Research supports this claim, showing that visual cues like an agent's hair length often elicit gendered attributions [20], while vocal characteristics also play a significant role in shaping user perceptions [21]. Additionally, the occupational context associated with an agent further influences how it is gendered [3]. It remains unclear how effective gender neutrality has been as a strategy for preventing gendering and mitigating stereotypical or negative reactions.

In this research, we use the term "gender-ambiguous" rather than "gender-neutral". We believe that completely removing gender from an agent is a complex challenge, and no matter how carefully researchers design their experiments to avoid gendering agents, there will always be influencing factors, such as participants' demographics (e.g., age, gender, education level, etc.), cultural background, and geographical context. The term "ambiguous" more accurately reflects the subjective nature of perception, where the same voice or agent might be interpreted as ambiguous by some, but perceived as more feminine or masculine by others. We argue that the solution is not to eliminate gender entirely but rather to reduce its prominence. We also believe that by incorporating more ambiguous options (such as in voice design, agent's appearance, etc.), we are taking a step forward in supporting gender-nonconforming people. This translates to offering more inclusive options in technology.

1.2 The Review

This review aims to examine the current state of research on gender-ambiguous voices in HCI, which, to our knowledge, has not been previously explored. Firstly, we examined how researchers describe ambiguous voices by asking:

RQ1 What defines an ambiguous voice in HCI literature?

Understanding how an ambiguous voice is defined is fundamental because the concept is relatively new and lacks a universal definition in the field. A clear and consistent definition is necessary for establishing a common understanding.

Afterwards, we examine the current state of available gender-ambiguous voices by asking:

RQ2 What are the currently available gender-ambiguous voices?
RQ2-a What is the availability of gender-ambiguous voices?
RQ2-b What characteristics do gender-ambiguous voices have?

RQ2-c What methods were utilised to create these gender-ambiguous voices?

Identifying the availability of gender-ambiguous voices is crucial for understanding their accessibility for research and development. Knowing their characteristics, such as acoustic features, helps researchers in creating effective ambiguous voices. Additionally, examining creation methods reveals advancements in technology and methodology, guiding future work in the field. Finally, we seek to understand user perceptions and evaluation methods by asking:

RQ3 What is the perception of gender-ambiguous voices?
RQ3-a What methods were used to evaluate the perception of gender-ambiguous voices?

Exploring how participants perceive these voices and how their perception is evaluated is fundamental for assessing whether ambiguous voices are useful and effective.

To answer these questions and identify missing research gaps, we systematically reviewed 36 studies investigating various aspects of gender-ambiguous voices. We begin by describing our methods for conducting this review, ensuring transparency and replicability in our approach. Afterwards, we report and discuss the results, highlighting key findings.

This systematic review aims to examine current research on ambiguous voices to enhance their visibility. By moving beyond the binary, we strive to foster more socially inclusive interactions between humans and technology. This work contributes: 1) A comprehensive review of the state of the art on gender-ambiguous voices in HCI and their availability status, 2) An analysis of the acoustic characteristics of gender-ambiguous voices, 3) A categorisation of the methods used to create ambiguous voices, and 4) A review of the perception of ambiguous voices and methods for evaluating them. Additionally, we aim to fill research gaps by proposing the following: 1) Clear definitions of terms used to describe non-gendered voices and a universal term for referring to non-gendered voices, 2) An explanation of why creating ambiguous voices is so complex, 3) Guidelines for evaluating ambiguous voices, 4) An open repository of ambiguous voice samples, 5) A reflection on the broader implications of ambiguous voices, and 6) Recommendations on key factors to consider when developing an ambiguous voice.

1.3 Defining Terms

In this paper, we will also delve into specific speech technology terminology. To make this paper accessible also by researchers who might not have a specific speech acoustics background, here we will define some terms that will be used in the following sections.

Throughout this paper, we are going to talk about **Text-To-Speech (TTS)** voices, which are computerised voices generated through an algorithmic process that transforms digital text into audio output resembling human speech [17].

In addition, we will focus on the following voice characteristics:

- **Pitch** refers to the fundamental frequency of the voice, perceived as the highness or lowness of a sound. The feminine pitch has a range from 145-275 Hz, while the masculine one is on average from 80-165 Hz [35].
- The **speech rate** is the speed at which speech is delivered, reflecting the temporal aspects of speech production. For conversational speech, the rate usually ranges from 2 to 11 syllables per second [39].
- **Jitter** is the cycle-to-cycle variation of the fundamental frequency. It represents the short-term (cycle-to-cycle) irregularity in the pitch period. Normal jitter values are typically less than 1.0% [78]. Acoustically, increased jitter can result in rough voice quality, making the voice sound less smooth and stable. When the jitter is low or within normal limits, the voice tends to sound smooth, steady, and controlled.
- **Harmonics-to-Noise Ratio (HNR)** is a measure of the amount of periodic signal (harmonics) to aperiodic signal (noise) in the voice. It quantifies the relative amount of noise in the voice signal and its mean value is around 20 dB [19]. Perceptually, a lower HNR indicates more noise in the voice, leading to a breathy or harsh quality. A higher HNR is associated with a clearer, more resonant, and pure voice quality.
- **H1-H2** refers to the amplitude difference between the first harmonic (H1) and the second harmonic (H2) of the voice spectrum. It is a measure of voice quality often associated with breathiness. The value for H1-H2 varies according to the pitch, language etc., with feminine values ranging between 0-5 dB, and masculine ones between -5 to 0 dB [57].

2 Method

In this section, we present the scheme employed to systematically review and categorise the literature on gender-ambiguous voices in HCI. By presenting our coding and selection methodology, we aim to offer transparency in how studies were chosen and analysed, facilitating a comprehensive meta-synthesis of the research landscape. The coding sheet for the final selection of papers is available in the supplementary material.

2.1 Coding Scheme

For each of the papers we were interested in collecting the information presented in Table 1.

2.2 Procedure

The systematic review was conducted following the PRISMA format [48] shown in Figure 1. This section outlines the methodology employed for conducting the systematic review and meta-synthesis of the literature.

2.2.1 Search Query Keywords. The databases employed in the search were ACM Digital Library and IEEE Xplore. The same query was adapted based on the database's requirements. The query keywords are shown in Figure 1. The search started on the 12th of December 2023 and ended on the 18th of January 2024. Inspiration for the query keywords was taken from the work done by Seaborn et al. [67] and then adapted to the topic of gender-ambiguous voices.

2.2.2 Eligibility Criteria. Inclusion and exclusion criteria were established according to the research question and by Kitchenham's [38] guidelines for selecting papers in the field of engineering. Inclusion criteria were: research that employed gender-ambiguous voices or focused on the design, generation, perception or evaluation of these voices. Regarding the exclusion criteria: inaccessible

Collected information	1. Type of paper (full paper, late-breaking report, extended abstract, etc.)
	2. Database (ACM Digital Library, IEEE Xplore)
	3. Name of the conference/journal
	4. Year of publication
	5. User study, if yes which user group was targeted
	6. Type of agent (voice assistant, virtual agent, robot, other). If a robot, specify which
	7. Voice technology (new TTS voice, commercial voice, modulated human voice, etc.)
	8. Name given to the voice (gender-neutral, gender-ambiguous, genderless, etc.)
	9. If there was a definition for the voice
	10. Whether the voice was the main investigated variable
	11. Whether the main focus was on voice generation
	12. Whether the voice is open source (yes, no, unknown)
	13. Whether the voice was evaluated
	14. Whether the voice agent was evaluated

Table 1: Information collected from each paper included in the review.

papers, papers not in English, survey papers, duplicate papers, review papers, and technical papers that are more towards electronics.

2.2.3 Selection Process. The process began with defining the keywords to include in the query. The search string was tested repeatedly by two authors to ensure no errors in the query and to verify the consistency of results across both databases. As databases, it was decided to use ACM Digital Library and IEEE Xplore given that most HCI research is published there. Once this phase was concluded and 778 papers were selected, the process continued with the elimination of duplicates found between ACM Digital Library and IEEE Xplore, reaching a total of 742 papers. Then, the screening phase began and after scanning all the abstracts, 62 papers were selected. The abstracts were assessed against the predefined inclusion and exclusion criteria listed above, using a structured form to ensure consistency. In the Eligibility phase, the papers were fully read and selected based on the inclusion/exclusion criteria. In this phase, 38 papers were selected. At the end of this process, 20% of the 38 selected papers (N = 8 papers), picked at random, were evaluated by two authors. The full-text assessment was conducted using a standardised form to ensure that both authors evaluated and annotated the papers according to the same criteria. The information collected during this process is presented in Table 1. Cohen's Kappa statistic was employed to evaluate the consistency between raters. Among the 8 papers selected, 100% agreement was reached in 13 out of 15 items to be evaluated (presented in Table 1), while in the remaining two papers, 20% agreement was reached for the "user group" category, and 75% for the "voice technology" category. Thus, the two authors discussed the discrepancies until 100% agreement was reached on those items as well. Therefore, this indicates strong agreement between the two raters according to Cohen [13]. After the inter-rater evaluation, 2 papers were excluded, leaving a total of 36 papers for the final review.

While we followed the majority of the PRISMA steps, it is important to note that certain aspects of the framework did not apply to this review. For example, since this is a systematic review and not a meta-analysis, statistical synthesis and calculations (such as heterogeneity analysis) were not performed. Additionally, we did not include a separate "data synthesis" step as we focused on synthesising findings through qualitative methods rather than quantitative ones. These limitations do not undermine the rigour of the review but highlight the specific scope of the current work.

3 Results

In this section, we present the data collected during our review, with all relevant papers listed in Table 2. Given that the topic of gender-ambiguous voices is relatively new and underexplored, we chose not to exclude studies based on their type of publication. Consequently, our review includes a diverse range of research: 31 full papers, 3 extended abstracts, and 2 late-breaking reports. Of the 36 papers reviewed, 22 were published in the ACM Digital Library, 9 in IEEE Xplore, and 5 were available in both. These papers were published between 2000 and 2024 (see Figure 2), with a notable concentration of 14 papers published between 2022 and 2023, highlighting the growing interest in this topic.

The selected papers span across 15 different venues, demonstrating the breadth of interest in this emerging field. The papers' venues are: ACM/IEEE International Conference on Human-Robot Interaction (HRI) [1, 2, 8, 12, 18, 25, 27, 31, 47, 88], ACM Conference on Human Factors in Computing Systems (CHI) [15, 53, 56, 68, 69, 76, 79], IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) [4, 40, 70, 80, 81], Conference on User Science and Engineering (i-USEr) [50, 51], Journal of Human-Robot Interaction [5, 32], International Journal of Child-Computer Interaction [34], International Journal of Human-Computer Studies [33], Journal of ACM on Human-Computer Interaction [89], ACM Conference on Conversational User Interfaces (CUI) [36], ACM Transactions on Interactive Intelligent Systems (TiiS) [55], Conference on Digital Avionics Systems (DASC) [22], ACM Designing Interactive Systems Conference (DIS) [61], International Conference (iConference) [44], Special Interest Group on Computer Graphics and Interactive Techniques Conference (SIGGRAPH) [58], and ACM International Conference on the Design of Communication (SIGDOC) [63].

Regarding user studies, only one paper was not a user study [88]. The user groups used are various: students (high school, college, university) and faculty members [1, 5, 12, 27, 40, 47, 51, 56, 61, 63, 70, 76], young adults [33], people with technical background [50], people with background in Human-Computer Interaction and User Experience and knowledge about VAs [53], native English

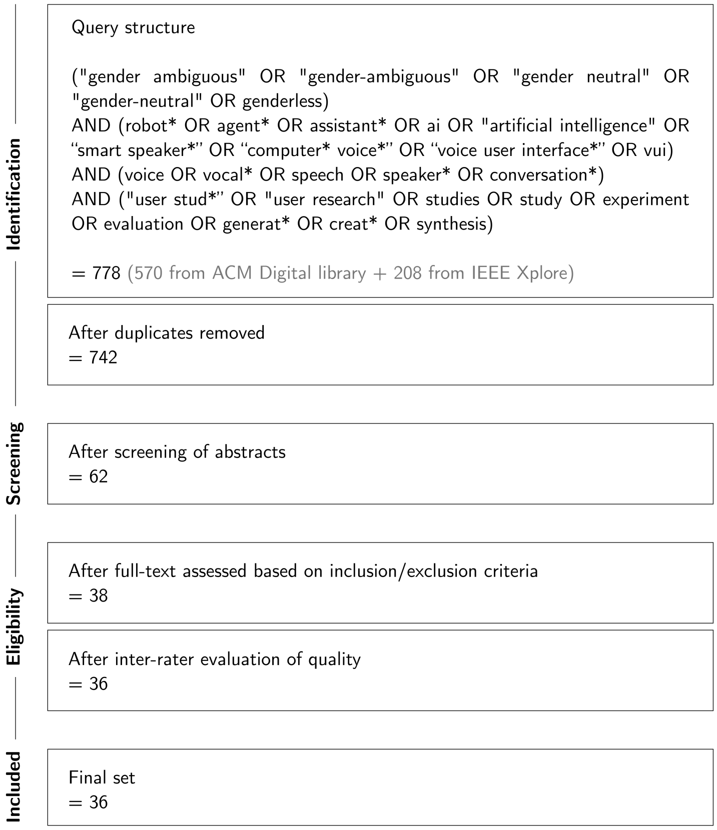

Identification
Query structure
("gender ambiguous" OR "gender-ambiguous" OR "gender neutral" OR "gender-neutral" OR genderless)
AND (robot* OR agent* OR assistant* OR ai OR "artificial intelligence" OR "smart speaker*" OR "computer* voice*" OR "voice user interface*" OR vui)
AND (voice OR vocal* OR speech OR speaker* OR conversation*)
AND ("user stud*" OR "user research" OR studies OR study OR experiment OR evaluation OR generat* OR creat* OR synthesis)
= 778 (570 from ACM Digital library + 208 from IEEE Xplore)
After duplicates removed
= 742
Screening
After screening of abstracts
= 62
Eligibility
After full-text assessed based on inclusion/exclusion criteria
= 38
After inter-rater evaluation of quality
= 36
Included
Final set
= 36

Figure 1: Flow diagram used for the systematic review process following the PRISMA format [48].

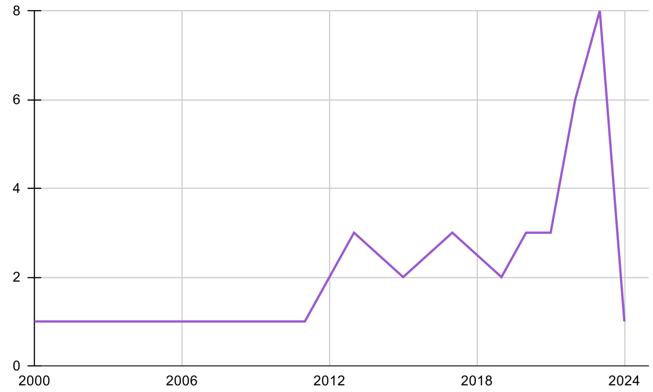


Figure 2: The year distribution of the analysed papers. Note that the data collection ended in January 2024, hence the low number of publications for this year. The figures in this paper use a colour palette inspired by the non-binary pride flag.

speakers [31, 32, 44], children [4, 34, 58, 81], social media users [89], aeroplane pilots [22], drivers [36], customers [18], lebanese and americans [2], participants recruited on online crowdsourcing [8, 25, 55, 68, 69, 79, 80], and non-binary and/or transgender people [15].

The studies predominantly focused on three agents: robots, voice assistants, and virtual agents. Voice assistants were employed in 6 [22, 36, 44, 53, 61, 79], virtual agents in 2 [50, 55] and robots were featured in 21 studies [1, 2, 4, 5, 8, 12, 18, 25, 27, 31–33, 47, 51, 56, 58, 63, 70, 76, 80, 81]. In 6 studies, no agent was involved since the focus was only on voices [15, 40, 68, 69, 88, 89], and, in one case, the study was about the evaluation of a game character [34]. Studies involving robots used the following robots: Lego Mindstorms [1, 2], KUBO [12], NAO [25, 27, 33], Pepper [8, 56, 81], inMoov [4], Furhat [47, 80], Robovie-X [51], Haksh-E [58], Robovie-R3 [18], Adept Pioneer 3 DX [5] and Wakamaru [32, 76].

3.1 Definition of Ambiguous Voices (RQ1)

Across the 36 papers reviewed, 32 used terms such as "gender-neutral", "gender-ambiguous", "genderless", "gender-free", "non-binary" or "androgynous" to describe the voice used in their studies. Specifically, 19 papers used the term "gender-neutral" [1, 2, 4, 5, 8, 12, 18, 22, 25, 27, 31, 33, 34, 40, 51, 53, 56, 58, 76], 5 used "gender-ambiguous" [32, 36, 44, 79, 80], 3 used "androgynous" [55, 70, 89], 2 used "genderless" [61, 81], 2 used "non-binary" [15, 47], and 1 chose "gender-free" [88]. Of the remaining 4 papers, 2 did not label the voice, and the other two described the voices as Kawaii, a style of speaking characterised by a high-pitched, soft, and cute tone, usually associated with Japanese pop culture [68, 69].

Citation	Year	User study	Agent				Voice technology					Voice							Voice definition	Voice as main variable	Focus on voice generation	Open source			Evaluation	
			Voice assistant	Virtual agent	Robot	Other	New TTS	Commercial voice	Customised artificial voice	Modulated human voice	Other	Gender-neutral	Gender-ambiguous	Genderless	Gender-free	Non-binary	Androgynous	Other				Yes	No	Unknown	Voice	Agent
Andrist et al. [1]	2013	•			•						•	•												•		•
Andrist et al. [2]	2015	•			•					•		•												•		•
Axelsoon et al. [4]	2019	•			•				•			•												•		•
Ball et al. [5]	2017	•			•						•	•												•		•
Bryant et al. [8]	2020	•			•			•				•												•		•
Christiansen et al. [12]	2022	•			•				•			•								•			•		•	•
Danielescu et al. [15]	2023	•				•	•									•				•	•	•			•	
Edirisinghe et al. [18]	2024	•			•			•				•											•			•
Faerber and Garloch [22]	2000	•	•					•				•												•		•
Green et al. [25]	2022	•			•			•				•											•			•
Hayes et al. [27]	2013	•			•				•			•											•		•	•
Huang and Mutlu [31]	2012	•			•					•		•												•		
Huang and Mutlu [32]	2013	•			•					•			•											•		•
Huang et al. [33]	2023	•			•			•				•											•			•
Hwang and Kang [34]	2023	•				•			•			•								•				•	•	
Jestin et al. [36]	2022	•	•						•				•						•	•				•	•	•
Kuch et al. [40]	2023	•				•					•	•								•		•			•	
Lopatovska et al. [44]	2022	•	•						•				•						•	•		•			•	•
Miranda et al. [47]	2023	•			•			•								•							•			•
Niculescu et al. [50]	2010	•		•					•									•						•		•
Nomura and Takagi [51]	2011	•			•				•			•												•		•
Parviainen and Søndergaard [53]	2020	•	•							•		•												•		•
Pejsa et al. [55]	2015	•		•						•							•							•		•
Peng et al. [56]	2019	•			•			•				•											•			•
Prabha et al. [58]	2022	•			•			•				•												•		•
Rinott et al. [61]	2021	•	•					•						•									•			•
Rose and Björling [63]	2017	•			•						•							•						•		•
Seaborn et al. [68]	2023	•				•		•										•		•				•	•	
Seaborn et al. [69]	2023	•				•					•							•		•				•	•	
Sembroski et al. [70]	2017	•			•			•									•							•		•
Szafir and Mutlu [76]	2012	•			•					•		•												•		•
Tolmeijer et al. [79]	2021	•	•						•				•						•	•			•		•	•
Torre et al. [80]	2023	•			•						•		•						•	•				•	•	•
Uluer et al. [81]	2020	•			•			•						•									•			•
Yu et al. [88]	2022					•	•								•				•	•	•			•	•	
Zhang et al. [89]	2021	•				•	•										•			•				•	•	
Count or range	**24 yrs**	**35**	**6**	**2**	**21**	**7**	**3**	**12**	**9**	**6**	**6**	**19**	**5**	**2**	**1**	**2**	**3**	**4**	**5**	**12**	**2**	**3**	**10**	**23**	**13**	**28**

Table 2: Overview of surveyed papers, listed alphabetically.

Only 5 papers out of 36 papers provided an explicit definition for the voice used [36, 44, 79, 80, 88], while the remaining 31 did not [1, 2, 4, 5, 8, 12, 15, 18, 22, 25, 27, 31–34, 40, 47, 50, 51, 53, 55, 56, 58, 61, 63, 68–70, 76, 81, 89]. The majority of the definitions were for "gender-ambiguous" voices. Indeed, Lopatovska et al. [44] wrote: *"Our definition of gender-ambiguous voice relied on the perception of this voice as neither clearly feminine nor masculine"*. Tolmeijer et al. [79] at first wrote: *"In accordance with Sutton [75], we use the term "gender-ambiguous" throughout this paper rather than calling a voice "genderless": many cues in the sound and content of VA speech can illicit gender ascription, even when the pitch is gender-neutral"*. Afterwards, they wrote: *"Gender-ambiguous refers to a voice that falls into both spectrums, meaning that different people would assign different genders to it based on prior mental models"*. Torre et al. [80] defined both gender-ambiguous and agender voices: *"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"*. Another definition of gender-ambiguous comes from Jestin et al. [36], we decided to include it even though they cite Johnson [37] at the end of their definition: *"Gender ambiguous voices can be pulled into being seen as either male or female despite being carefully designed to sound non-binary"*. Lastly, Yu et al. [88] defined gender-free: *"...gender-free speech means that humans cannot recognise the gender when hearing the speech audios."*.

Of 36 reviewed papers, 32 used terms like "gender-neutral", "gender-ambiguous", or "non-binary" to describe voices, with "gender-neutral" being the most common (19). Only five defined these terms, mostly focusing on "gender-ambiguous" voices, with varying definitions—some describing voices that lacked clear gender alignment and others as sounding both masculine and feminine.

3.2 Gender-Ambiguous Voices in HCI (RQ2) and Their Availability Status (RQ2-a)

In our analysis, we categorised the voice technology used in each experiment as "New TTS voice", "Commercial voice", "Customised artificial voice", "Modulated human voice", or "Other". The category "New TTS voice" includes studies that created a Text-To-Speech voice from scratch. Three studies ticked this box [15, 88, 89]. The label "Commercial voice" was applied to studies that used pre-existing voices from the market or literature, such as Amazon Polly, robot voices, or other TTS systems. Twelve studies fell into this category [8, 18, 22, 25, 33, 47, 56, 58, 61, 68, 70, 81]. "Customised artificial voice" includes studies where researchers modified some acoustic characteristics of pre-existing voices (pitch, speech rate, etc.) to achieve a neutral sound. Nine papers used this approach [4, 12, 27, 34, 36, 44, 50, 51, 79]. With "Modulated human voice", we meant all the human voices that have been modulated in pitch, for example, to achieve a specific frequency that can be perceived as ambiguous or neutral. Six studies used this method [2, 31, 33, 53, 55, 76]. Finally, 6 studies were labelled "other" because they either used a mix of different voice technologies [40, 80], or did not specify the technology behind the voice [1, 5, 63, 69].

In terms of availability, in 23 studies, researchers did not report whether the voice used was open source [1, 2, 4, 5, 8, 22, 31, 32, 34, 36, 50, 51, 53, 55, 58, 63, 68–70, 76, 80, 88, 89]. Ten studies did not use an open source voice, such as Amazon Polly or Siri [12, 18, 25, 27, 33, 47, 56, 61, 79, 81] and only three papers stated that the voice used was open source [15, 40, 44].

Notably, even though all these studies employed voices in their experiments, only 12 explicitly investigated voice as a main variable of interest [12, 15, 34, 36, 40, 44, 68, 69, 79, 80, 88, 89]. The rest used voice as part of their experiments without focusing on it [1, 2, 4, 5, 8, 18, 22, 25, 27, 31–33, 47, 50, 51, 53, 55, 56, 58, 61, 63, 70, 76, 81]. Out of the 36 papers reviewed, only two [15, 88] dedicated their entire study to the generation of an ambiguous voice. The remaining studies combined voice generation with other experimental factors.

In short, most studies used commercial or customised voices, with only a few creating new ones. Most studies did not specify if the voices used were open-source or used open-source voices.

3.3 Characteristics of the Voices (RQ2-b)

To understand what characteristics gender-ambiguous voices have, we reached out to the developers of the voice systems described in the 36 papers to request audio files. We were able to obtain samples from 4 different gender-ambiguous TTS voices: 1) The Sam voice (SAM) used by Danielescu et al. [15], 2-3) The customised Amazon Polly voice (CAN_1) and the ad-hoc experiment TTS voice (CAN_2) from Torre et al. [80], and 4) The voice of the robot Haksh-E (MINI) by Prabha et al. [58]. We conducted acoustic analyses on the same sentence uttered by all 4 voices. The sentence was *"Hello! I'm a robot and I work as a tour guide. My work consists of guiding people around museums and galleries!"*. We chose this sentence because we plan to use the samples for a separate experiment on gender-ambiguous voices which is beyond the scope of the current review.

Regarding the characteristics, we measured pitch (Hz), speech rate (segment/s), jitter (%), HNR (dB), and H1-H2 (dB) as typical measures that are used to describe voice quality in the phonetics literature [57]. Voice features were extracted using Praat [7], a free package for phonetic analyses, in Python. In Table 3, we present the acoustic characteristics of the four ambiguous voices.

It is important to note that these analyses are exploratory and based on a limited sample of four voices, and as such, the findings should not be considered representative or generalisable to all gender-ambiguous voices.

3.3.1 *Ambiguous Voices Evaluation.*

The four voices listed above were also included in a separate investigation, submitted elsewhere, that focused on the perception of these voices in conjunction with robot bodies [16]. For completeness, here we also report the results of a pilot study aimed at assessing the gender perception of these sample voices. However, it should be noted that this investigation is limited by the sample size (we were able to obtain only 4 out of 36 surveyed voices) and thus it is not representative of the entire breadth of gender-ambiguous voices recently generated by the TTS community. For this pilot study, we recruited 180 online participants from the UK, aged 18–72 (μ = 39.6, σ = 13.8), comprising 109 women, 67 men, and 4 non-binary individuals, who evaluated six gender-ambiguous voices, including the four discussed here.

Participants listened to each voice in a randomised order and evaluated them based on gender. All voices articulated the same sentence: *"Hello! I'm a robot, and I work as a tour guide. My work consists of guiding people around museums and galleries!"* The set included six gender-ambiguous voices (four of which are analysed

Voice name	Pitch (Hz)	Speech rate (segment/s)	Jitter (%)	HNR (dB)	H1-H2 (dB)
Sam [15]	$\mu = 142.635$, $\sigma = 14.086$	$\mu = 1.469$, $\sigma = 0.903$	$\mu = 1.948$, $\sigma = 0.744$	$\mu = 15.861$, $\sigma = 1.692$	$\mu = 6.545$, $\sigma = 2.213$
CAN_1 [80]	$\mu = 138.173$, $\sigma = 12.024$	$\mu = 3.401$, $\sigma = 2.780$	$\mu = 2.107$, $\sigma = 0.504$	$\mu = 12.333$, $\sigma = 4.983$	$\mu = 7.697$, $\sigma = 4.429$
CAN_2 [80]	$\mu = 157.652$, $\sigma = 35.365$	$\mu = 2.557$, $\sigma = 1.722$	$\mu = 2.489$, $\sigma = 0.847$	$\mu = 12.022$, $\sigma = 3.410$	$\mu = 2.656$, $\sigma = 3.908$
MINI [58]	$\mu = 248.677$, $\sigma = 22.445$	$\mu = 1.633$, $\sigma = 0.963$	$\mu = 1.542$, $\sigma = 0.534$	$\mu = 17.659$, $\sigma = 1.955$	$\mu = 3.734$, $\sigma = 3.082$

Table 3: Acoustic characteristics of the 4 collected ambiguous voices (pitch, speech rate, jitter, HNR (Harmonics-to-Noise Ratio), H1-H2 (Amplitude difference between the first harmonic, H1, and the second harmonic, H2)).

in this study), along with a male and a female voice (Cristopher and Sara from speechgen.io [71]), used as attention checks. Evaluations were conducted using a 5-point Likert scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*), with participants rating the following statements: *"The voice sounds feminine"*, *"The voice sounds masculine"*, and *"The voice sounds ambiguous"*. A definition of *ambiguous* was provided at the start: *"By 'ambiguous', we mean that the voice sounds neutral or androgynous"*. Demographic information was collected at the end. In Figure 3, we present the distribution of ratings for the statement *"The voice sounds ambiguous"*, with CAN_1, CAN_2, and SAM emerging as the most ambiguous.

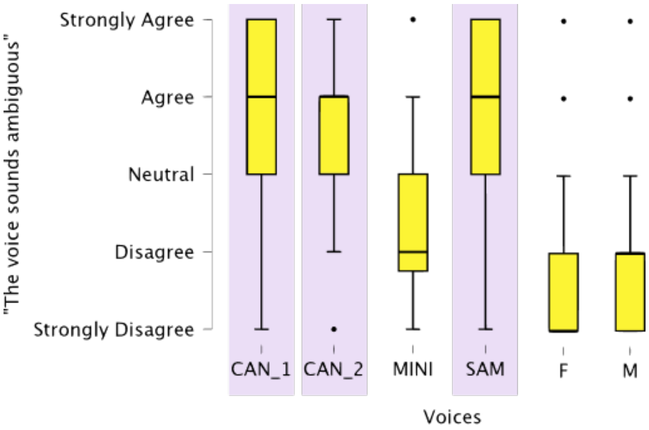


Figure 3: The distribution of ratings for the ambiguous statement is presented for the four gender-ambiguous voices and the two gendered voices (F = female voice, M = male voice). The shaded boxes highlight voices that were perceived as more ambiguous than the other voices but were not perceived to be different among each other.

Initially, Levene's test for homogeneity of variance was performed to assess whether variances in the ratings of ambiguous voices were consistent between samples. The test indicated a statistically significant variance ($p < 0.001$). Consequently, a non-parametric ANOVA (Kruskal-Wallis test) was performed, revealing a statistically significant difference in how participants rated the voices ($H(7) = 560.122, p < 0.001$). Post hoc analyses using Dunn's test showed significant differences between most pairs of voices, with the exceptions of CAN_1 vs. CAN_2 ($p = 0.124$), CAN_1 vs. SAM ($p = 0.724$), and CAN_2 vs. SAM ($p = 0.058$), suggesting that these 3 voices were perceived as equally ambiguous.

3.4 Methods Used to Create the Voices (RQ2-c)

As explained in the previous section, some studies used a commercial voice, some decided to customise it, some started with a human voice and then modulated it and a few created a new TTS voice. We will now present the voices used by the researchers and explain, if possible, how they created or modified them.

3.4.1 Agent's Built-in Voice. Huang et al. [33] used the agent's built-in voice since it was perceived as gender-neutral. Edirisinghe et al. [18] described the robot's voice as child-like and gender-neutral. Similarly, Uluer et al. [81] said that the robot's voice was both childlike and genderless, characterised by a high pitch and slow articulation speed.

Some researchers used the agent's built-in voice to generate speech but did not provide details on their methodology. For instance, Peng et al. [56] and Hayes et al. [27] used the built-in software of the robot to produce gender-neutral speech. Axelsoon et al. [4] convert a human-like, feminine voice into a gender-neutral one by adjusting its pitch.

3.4.2 Modulated Human Voice. Some studies focused on creating a gender-ambiguous voice by modifying a human voice. Some did not provide any specific detail on the method used [31, 33]; some reported starting with a feminine voice [2, 76], while one study [53] specifically mentioned changing the pitch of a feminine voice from 168 Hz to 153 Hz to create a gender-neutral one.

3.4.3 Amazon Voices. A few studies used Amazon voices, particularly the Amazon Polly voice named "Ivy", without any changes [25, 61]. Two studies used the VoiceFlow platform and Amazon voices. Lopatovska et al. [44] used the Amazon Alexa developer console and the VoiceFlow platform. They programmed all experimental utterances and responses into the voice-activated Alexa skill app using Node.js and JSON syntax. The ambiguous voice in their study took inspiration from the Q genderless voice [11], which was created by increasing the frequency of a masculine voice to 160 Hz. Jestin et al. [36] also used Voiceflow to convert text prompts entered into the speak block into voice output. They created an ambiguous voice by manipulating a feminine voice from Amazon Alexa using Speech Synthesis Markup Language – an XML-based markup language used to control various aspects of speech synthesis – to adjust prosody elements such as pitch and rate.

3.4.4 *Text-To-Speech, Computer and AI-generated voices.*

The majority of the studies used TTS, computer, and AI-generated voices. A few did not provide any additional detail about the creation process [8, 68]. Some reported that they used a pre-existing TTS feminine voice with low pitch [50], and others said they used a TTS program in a slightly high-pitched, androgynous voice [70].

In the following paragraph, we will provide a more detailed explanation of how these voices were created. By necessity, we will use TTS-specific terminology.

Seven of the papers used specific voice-generation techniques. Yu et al. [88] developed a gender-free TTS system consisting of three sub-models: a speech gender encoder, a TTS synthesiser, and a vocoder network. The speech gender encoder creates gender-specific embeddings from masculine and feminine speech samples, which are then combined by a rule-based model to form a gender-free speech style embedding. This embedding is converted into a Mel spectrogram by the Tacotron 2-based TTS synthesiser, and the final gender-free speech audio is produced using a WaveNet-based vocoder. Nomura and Takagi [51] synthesised voices using tools including Easy Speech, Text-To-Speech Engine Japanese version, Sound Engine Free, Microsoft SPAI 4.0, and L&H TTS 3000. Prabha et al. [58] built a conversational AI system using RASA, an open-source machine learning framework, in combination with Google's Automatic Speech Recognition and TTS engines. Tolmeijer et al. [79] created an ambiguous voice by shifting a Google WaveNet feminine voice down by three semitones. Hwang and Kang [34] used CLOVA AI for generating Text-To-Speech AI voice files, which were then modified with the pitch shifter function in Adobe Premiere Pro to achieve a gender-neutral sound. Zhang et al. [89] generated voices using IBM Watson and Natural Reader to produce both gendered and androgynous voices. Danielescu et al. [15] created a non-binary TTS voice by first recording a voice actor with a range of masculine and feminine speech characteristics. They processed these recordings to develop initial synthetic voices, which were evaluated through a survey conducted with the non-binary and transgender community and conducted additional evaluations.

3.4.5 *Mix of Voices.*

Two studies employed more than one voice for their experiments. Torre et al. [80] initially modified the "Kendra" Amazon Polly voice to make it sound ambiguous by lowering its pitch to an average fundamental frequency of approximately 135 Hz. In addition, they created three new TTS voices, with a process described in detail in [77]. These voices were trained using a multi-speaker version of the Tacotron 2 neural TTS engine, utilising large corpora of masculine and feminine voices. To achieve a gender-ambiguous timbre, they averaged speaker embeddings and used Speech Gender Recognition to ensure the voices were not easily classified as masculine or feminine. Kuch et al. [40] used both natural and synthetic voices, which were then adjusted for gender neutrality. They followed the method proposed by Rizhinashvili et al. [62], which involves manipulating the mean fundamental frequencies of the voices to fall within a gender-neutral range, situated between the average frequencies of masculine and feminine voices. For synthetic voices, sentences were generated using Microsoft Azure with two feminine and two masculine voices, which were then processed with the gender-neutralisation filter.

3.4.6 *Unspecified Voice Technology.*

In four studies, the specific technology used was not presented. Andrist et al. [1] described using two modulated voices intended to be gender-neutral. Ball et al. [5] used a pre-recorded spoken message that was modified to sound both synthetic and gender-neutral. Rose and Björling [63] reported designing a voice characterised by a mid-range pitch, childlike, and robotic, without explicitly conveying any particular gender. Lastly, Christiansen et al. [12] generated verbal utterances using a voice generator, but the exact procedure is not clear.

All in all, the studies examined used various techniques to create gender-ambiguous voices. These techniques included mainly modifying built-in agent voices, modulating human voices, and using commercial TTS systems like Amazon Polly. Some studies employed custom techniques, such as pitch shifting or creating new TTS voices using machine learning models like Tacotron 2 and WaveNet. Several studies involved both synthetic and natural voices, with modifications to reach gender neutrality.

3.5 Perception of the Voices (RQ3)

A few studies examined how participants perceive ambiguous voices, but the findings are inconclusive. In some cases, these voices were negatively perceived. For instance, Christiansen et al. [12] found that participants described the robot's gender-ambiguous voice as annoying and preferred a different voice. Jestin et al. [36] found that participants rated the ambiguous voice as more artificial and less desirable than a male voice. Danielescu et al. [15] observed that the ambiguous voice in their study was generally perceived less favourably than masculine and feminine voices. However, non-binary individuals responded more positively to the ambiguous voice, particularly in terms of trustworthiness and intelligence, especially when the voice had more feminine vocal characteristics (such as intonation). This feminine influence was also found by Torre et al. [80], who noted that participants generally perceived their ambiguous voice as more "feminine", even though the average ratings remained close to "neutral" on the Likert scales.

Tolmeijer et al. [79] found that perceptions of gendered and gender-ambiguous voices varied by participant gender. For example, women rated the ambiguous voice as significantly more friendly and polite than men did. Both the feminine high-pitched voice and the gender-ambiguous voice were not associated with stereotypically feminine traits like being delicate, family-oriented, or sensitive. In terms of trust, female participants reported trusting the gender-ambiguous voice more than male participants did.

In Lopatovska et al. [44], participants described the gender-ambiguous voice as: deep, low, monotone, slow, weird, not clear, and with a good timbre. When it came to gender identification, many participants struggled to determine the voice's gender. Also, 26 out of 65 participants preferred the ambiguous voice to a gendered one.

Finally, Kuch et al. [40], found that natural gender-specific voices were rated highest for anthropomorphism, followed by natural gender-ambiguous voices, synthetic gender-specific voices, and synthetic gender-ambiguous voices.

In summary, the analysed studies revealed mixed results. Some people found them annoying or less desirable compared to gendered voices, while non-binary people perceived them more positively. Perceptions also varied by participant gender, with women often rating ambiguous voices more favourably. In the comparison between natural and synthetic ambiguous voices, the natural ambiguous voices were preferred to synthetic ones.

3.6 Methods Used to Evaluate the Voices (RQ3-a)

About one-third of the studies evaluated voices, mainly with questionnaire scales. Tolmeijer et al. [79] asked participants to classify the voice into one of three categories: female, male, or unsure. Similarly, Hayes et al. [27] asked participants to rate the audio samplings using three scales: male, female, or gender-neutral. Some used a 5-point Likert scale, such as Torre et al. [80], who used items ranging from 1 (strongly disagree) to 5 (strongly agree) to assess whether the voice sounded "Feminine", "Masculine", "Agender", or "Ambiguous". Also, Jestin et al. [36] used a five-point Likert scale (1 = strongly disagree, 5 = strongly agree) to measure participants' overall gender preference for the voice assistant. Seaborn et al. [68, 69] assessed gender perception with a nominal scale that included categories for feminine, masculine, both, and neither (with the latter two indicating gender ambiguity). Participants could also provide their descriptions or select an alternate choice. Kuch et al. [40] also used a 5-point Likert scale with options for "Neutral", "Female", and "Male". They also explored the anthropomorphism of each voice using four scales from the Ho and MacDorman questionnaire [29], each rated on a 5-point Likert scale from 1 (less anthropomorphic) to 5 (more anthropomorphic). Hwang and Kang [34] used a 7-point one, where 1 indicated a male voice, 7 indicated a female voice, and scores between 3 and 5 were considered gender-neutral. Christiansen et al. [12] used a Likert scale to rate the helpfulness of the robot's gender-neutral voice, where 1 represented "Distracting" and 7 represented "Helpful". Similarly, Zhang et al. [89] employed a 5-point Likert scale to determine participants' likelihood of selecting the voice for their social media profiles.

Some researchers employed qualitative methods, such as interviewing participants about the voice quality and other factors in the experiment [50]. Danielescu et al. [15] used both quantitative and qualitative approaches to evaluate their non-binary TTS voice. They asked non-binary participants to listen to each voice sample individually and rate it on a 5-point Likert scale based on whether it sounded non-binary. Afterwards, they asked how comfortable they felt with it representing non-binary individuals, and how natural it sounded. Participants also answered open-ended questions about additional considerations for creating a non-binary TTS voice and provided further comments. Following feedback, the voice was refined, and another survey was conducted with different participants to reassess the same questions.

Yu et al. [88] did not evaluate their voice but used Principal Component Analysis (PCA) to visualise and verify the effectiveness of gender-free embedding extraction. By reducing high-dimensional gender style embeddings into a 3D space, they confirmed that gendered voices cluster in distinct regions while gender-free ones fall in between, validating the effectiveness of their approach.

In short, around one-third of the studies evaluated the voices. Most of them used quantitative methods – typically a 5-point Likert scale – to measure gender perception, although a few used qualitative or mixed methods. One study assessed the helpfulness of gender-neutral sounds rather than gender perception.

3.6.1 Demographics of participants. We also looked at who were the participants of the perception and evaluation studies reported above (papers listed in Sections 3.5 and 3.6), specifically regarding their gender, which, together with age, is the most typically collected participant demographics [66, 84]. These are reported, quoting the wording from the authors of the papers, in Table 4.

4 Discussion

This systematic review examines current research trends on gender-ambiguous voices within the Human-Computer Interaction field. It explores essential aspects such as definitions, availability, characteristics, creation methods, user perceptions and evaluation of these voices. By analyzing 36 studies, the review aimed to clarify existing practices and identify research gaps. In this section, we will highlight key findings from Section 3, present related open challenges, and propose potential solutions.

4.1 Missing Definitions

In the literature we analysed, several terms were used to describe voices that do not conform strictly to traditional masculine or feminine categories. Initially, we encountered "gender-ambiguous", "gender-neutral", and "genderless". As our review continued, we also found terms like "non-binary", "gender-free", and "androgynous". In Section 3.1, we outlined the five definitions researchers use to describe non-gendered voices. Although these terms are often used interchangeably, they do not always have the same meanings.

The challenges: 1) Agree on a standardised term to describe a voice that does not conform to binary gender categorisations "female" or "male", 2) Clarify the definitions of terms for beyond-binary voices, as the lack of commonly agreed-upon definitions in HCI can lead to confusion about their similarities and differences.

Addressing the challenges: 1) We propose that "gender-ambiguous" is the most appropriate term for these voices. Terms like "gender-neutral" and "genderless" suggest a complete removal of gendered traits, which is not achievable with voices [74]. Indeed, people tend to categorise voices in binary terms, linking certain traits to "male" or "female". Additionally, many cues in both the sound and content of speech can influence gender perception [75, 79]. "Gender-ambiguous" acknowledges that these voices are not intended to eliminate gender but rather resist clear categorisation. This term more accurately describes voices that blend masculine and feminine traits while emphasising their ambiguity and fluidity. We recommend "gender-ambiguous" over "androgynous" because it allows for greater flexibility in how these voices are perceived, without having the connotations of physical appearance one may think of when hearing "androgynous". 2) We provide aggregate definitions in Table 5, grouping them in terms of similarity.

4.2 Complexity Behind Ambiguous Voices

As discussed in Section 3.4, many studies in our review used pre-existing voices rather than creating new ones. This preference

Paper	Total N	Gender of participants (authors' wording)
Christiansen et al. [12]	13	10 "male", 3 "female"
Danielescu et al. study 1 [15]	26	26 "non-binary"
Danielescu et al. study 2 [15]	25	25 "non-binary"
Danielescu et al. study 3 [15]	1010	"45% female, 45% male, 10% not exclusively male-or female-identifying"
Hayes et al. [27]	8	7 "male", 1 "female"
Hwang and Kang [34]	11	not reported
Jestin et al. [36]	18	9 "male", 9 "female"
Kuch et al. [40]	20	not reported
Lopatovska et al. [44]	108	71 "female", 34 "male", 2 "non-binary", 1 N/A
Niculescu et al. [50]	48	not reported
Seaborn et al. [68]	94	53 "female", 37 "male", 4 "another gender or N/A"
Seaborn et al. [69]	157	76 "woman", 73 "man", 0 "another gender", 8 "preferred not to say"
Tolmeijer et al. [79]	234	138 "female", 96 "male"
Torre et al. [80]	62	31 "male", 30 "female", 0 "non-binary", 1 N/A
Zhang et al. [89]	15	9 "female", 1 "woman", 1 "genderqueer", 4 "male"

Table 4: Gender distribution of participants in user studies about perception or evaluation of gender-ambiguous voices. "N" refers to the participant count after any exclusion criteria have been applied.

Term	Definition
Gender-ambiguous, androgynous	Terms to describe a voice that does not clearly fit into traditional male or female categories by blending both masculine and feminine characteristics, making it difficult to identify the speaker's gender based solely on vocal characteristics.
Gender-neutral, Genderless, Gender-free	Terms to describe a voice that aims to avoid masculine or feminine characteristics, with no alignment to male or female vocal traits, aiming for a balance that does not lean towards either gender.
Non-binary	A term to describe a voice that incorporates elements from either feminine and/or masculine voices or neither, and is more linked to someone's identity rather than their voice.

Table 5: Definitions describing non-gendered voices derived from the ones presented in Section 3.1.

likely stems from the complexity of creating gender-ambiguous voices, combined with the lack of guidelines or frameworks. Since this is a relatively new research area, there is limited literature on the technical aspects, such as fundamental frequency, speech rate, voice quality, etc., that make a voice sound ambiguous. This gap in guidance makes the process experimental and uncertain for researchers and developers. Additionally, creating such voices requires specialised knowledge of technology, including Text-To-Speech engines, machine learning algorithms, and audio processing techniques. The required technical expertise further complicates efforts to develop ambiguous voices. Generative AI tools are a new addition to the possibilities of generating artificial gender-ambiguous voices. However, these also raise new issues, such as lack of transparency on how the voices are generated, what data the models were trained on, and whether any human voice sources are appropriately acknowledged or compensated [59].

The challenge: Establish a clear understanding of which vocal characteristics contribute to the perception of a voice as gender-ambiguous rather than gendered.

Addressing the challenge: Our analysis of the voice samples (see Table 3) shows great variability among the acoustic characteristics. We suggest merging this (limited) data with the expertise of gender-affirming voice teachers – who help individuals modify their voices to better align with their gender identity. In particular, we are focusing on their knowledge about how to achieve an androgynous voice. According to gender-affirming voice teachers, key characteristics for achieving a more androgynous voice include pitch, resonance, weight, and prosody [54, 87]. Regarding **pitch**, for androgynous voices, it is typically recommended to find a midpoint between masculine and feminine pitches. For instance, feminine pitches have an average range between 145-275 Hz, and masculine pitches range between 80-165 Hz [35]. A potential target might be around 125-185 Hz, taking a bigger overlap between the two ranges. A brighter **resonance** is generally associated with a more feminine sound, while a darker resonance is often perceived as more masculine. Achieving an androgynous voice involves finding a balance between these two extremes. Voice **weight** refers to how heavy or light the voice sounds. Heavier voices are typically associated with masculinity, while lighter voices are associated with femininity. The goal of an androgynous voice is to balance these qualities. Finally, **prosody** includes elements such as pitch, duration, intensity, and speech rate. In English, certain prosodic features are commonly perceived as masculine or feminine, making prosody an important aspect of achieving an androgynous voice [41].

Combining these insights with our data, we observe that SAM, CAN_1, and CAN_2 fall within the ambiguous pitch range. H1-H2 values provide insight into resonance: a smaller H1-H2 difference generally indicates stronger resonance and a more "full-bodied" voice, typically associated with masculine voices. In comparison, a larger H1-H2 difference is often found in feminine voices. SAM and CAN_1 have higher H1-H2 values compared to CAN_2 and MINI. For voice weight, which can be derived from HNR values, a higher HNR correlates with a "heavier" voice due to stronger harmonics and less breathy noise, usually associated with masculine voices. Lower HNR values are related to "lighter" voices, often associated with feminine ones. In our data, SAM and MINI exhibit higher HNR values, while CAN_1 and CAN_2 have lower HNR values. Lastly, prosody involves various speech elements, making it complex to compare our data with the recommendations of speech therapists.

In Table 6 we present a simplified overview of the gender categorisation (feminine, masculine or ambiguous) of the characteristics (pitch, resonance, weight) of the four analysed voices. Although our analysis is based on only four voice samples, combined with insights from gender-affirming voice teachers, it provides a starting point for future research. We aim to assist researchers by offering initial insights into pitch, resonance, voice weight, and prosody in the context of ambiguous voices. As discussed in Section 3.4, some researchers aiming to create gender-ambiguous voices focused solely on adjusting the pitch. However, based on the samples we collected and insights from gender-affirming voice experts, pitch is just one of several factors influencing the perception of gender ambiguity. Therefore, those working on designing ambiguous voices should consider a range of features beyond pitch alone. We encourage future studies to build on our findings by expanding the dataset and refining the analysis. This will help develop more comprehensive guidelines for creating ambiguous voices.

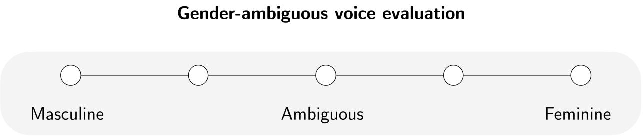
4.3 Guidelines for Voice Evaluation

As detailed in Section 3.6, the methods used to evaluate the gender perception of ambiguous voices vary widely, highlighting the need for standardised guidelines. About one-third of the studies reviewed employed evaluation techniques such as Likert scales. However, inconsistencies in labels for gender-ambiguous voices pose a significant issue. Some scales included options like *male*, *female*, *unsure*, *both*, or *neither*, but lacked specific categories for gender ambiguity. This lack of clarity can lead to inaccurate classifications and unreliable data, as participants may struggle to categorise voices correctly. Furthermore, the absence of detailed categories for ambiguity can obscure important nuances in how these voices are perceived, leading to an incomplete understanding of their impact.

Additionally, it is important to consider who are the people who evaluate the voices. Recent works within the Human-Computer Interaction [43, 52] and Human-Robot Interaction [84] research communities have highlighted that there exists a systematic over-representation of certain participant demographics in technological user research, which is skewed specifically toward Western, young, white, male, and technologically literate individuals. This is problematic, as it potentially reinforces and reiterates existing power structures in society [85], i.e. the "male-gaze" [28]. When the research in question deals with creating technology that breaks existing binary gender barriers (such as gender-ambiguous voices), this technology must be evaluated by a diverse (and especially gender-diverse) sample of people. Although all the articles collected in our systematic review dealt with "non-binary" artificial voices, from the demographics of the participants reported in Table 4, we can infer that only half of the studies (7 out of 14) allowed participants to select options other than a binary "female, male" option in the collected demographics. For the remaining 7 papers, we do not know if any non-binary individuals were actually recruited since they were (seemingly) not given the option to state so. This is problematic, as people who do not fall within the binary "male/female" categories are effectively made invisible [52]. Apart from studies that targeted non-binary individuals [15], we can also observe that participants' gender distribution was often unbalanced, notably with a tendency to over-represent women (e.g. [68, 79]).

The challenge: There is a need for standardised guidelines to evaluate gender-ambiguous voices. Such guidelines would ensure consistency across studies, facilitating comparison and integration of results. Standardised scales with well-defined categories for gender ambiguity would lead to more precise evaluations and better insights into voice characteristics and user perceptions, ultimately enhancing the quality of research in voice technology.

Addressing the challenge: Based on our review, we propose a preliminary evaluation scale for gender-ambiguous voices. We acknowledge that this is not a standardised measurement, but rather a scale that still needs to be validated. Given that most studies used Likert scales, we suggest a Likert-scale-based starting method. Psychometric comparisons between 5-point and 7-point scales show no clear advantages for either; however, one study suggests that a 7-point scale may be preferred by respondents with higher cognitive abilities, while a 5-point scale might be more suitable for the general public [83]. Another study [60] advocates for the 5-point scale, claiming it produces higher-quality data than 7- or 11-point scales, which aligns with the recommendations in [83]. Therefore, we suggest using a 5-point Likert scale for evaluating gender-ambiguous voices (see Figure 4). The Likert scale includes five points ranging from "masculine" to "feminine" with the midpoint labelled as "ambiguous". To ensure participants understand the term "ambiguous" we define a reference for those unfamiliar with the word.



Ambiguous: A term to describe a voice that does not clearly fit into traditional male or female categories and seems to blend both masculine and feminine characteristics.

Figure 4: Our proposal for a 5-point Likert scale evaluation for gender-ambiguous voices.

Furthermore, regarding participant recruitment and reporting, the HCI community has suggested guidelines on how to ask for, and report, participant gender [72], when it is needed for research purposes [65], and we urge developers of TTS voices to follow these guidelines. In terms of recruiting participants, diversity is crucial when user testing new technology, especially when the technology aims to reduce inequalities, break binary barriers, and promote

Voice name	Pitch	Resonance	Weight
SAM [15]	Ambiguous	Feminine	Masculine
CAN_1 [80]	Ambiguous	Feminine	Feminine
CAN_2 [80]	Ambiguous	Masculine	Feminine
MINI [58]	Feminine	Masculine	Masculine

Table 6: Simplified overview of the categorisation of the characteristics of the four analysed voices.

inclusivity. Even though diversity encompasses various dimensions such as age, race, nationality, language, ability, socio-economic status, and gender, among others, most empirical works in HCI tend to report only a few of these demographics [52, 66]. This might be due to existing norms in e.g. Psychology and Social Sciences [84]. We encourage developers of gender-ambiguous voices to consider who is evaluating their voices, and strive to achieve a balanced gender distribution among their participants. This does not mean necessarily recruiting 1/3 male, 1/3 female, and 1/3 non-binary individuals, but rather consider who the target user group will be, and recruit testers accordingly. For example, in the first two studies reported by Danielescu et al. [15], the focus was specifically on how a gender-ambiguous voice would be perceived in terms of representativeness by nonbinary transgender persons, and thus only people from these communities were recruited; in their third study, aimed at a larger scale evaluation of the voice, representative of a wider population, the sample of participants was recruited following a 45%, 45%, 10% split.

4.4 Importance of Open-Source Voices

As we explained in Section 3.2, a gap emerged regarding the availability of gender-ambiguous voices. Indeed, out of 36 papers, 23 did not disclose whether the voice was freely available, 10 used non-open-source voices, and 3 made their voices open-source.

Several factors might have contributed to the limited availability of open-source voices. Firstly, non-open-source voices are often easier and quicker for researchers to use. For many studies, the voice itself was not the main investigated variable, and thus generating and sharing a gender-ambiguous voice might have been an unnecessary effort. Secondly, existing commercial voices most often have restrictions on redistribution. Integrating commercial voices into research frameworks is typically more straightforward due to their established APIs and support, in comparison to the complexities of open-source platforms.

Despite these challenges, having a database of open-source gender-ambiguous voices would be beneficial for the HCI community, for several reasons: 1) They are cost-effective, as many researchers cannot afford commercial voices due to licensing fees, 2) They enhance transparency and reproducibility, allowing other researchers to access and replicate experiments, thus leading to more reliable and comparable results, 3) They promote inclusivity and innovation by enabling researchers from diverse backgrounds to contribute to and benefit from advancements in voice technology, encouraging more extensive experimentation and modification, 4) Knowledge of how the voice was generated allows researchers to critically address whether the voice adheres to their standards for inclusivity. For example, some researchers might want to use voices that have been generated with corpora of recordings of non-binary speakers, while others might prefer using voices that have been generated with corpora of masculine and feminine voice recordings.

The challenge: Institute a database of all available open-source gender-ambiguous voices. This database would allow researchers to easily find, compare, analyse, and select voices for their studies.

Addressing the challenge: To address this challenge, we have created an open repository* of the open-source voices we have encountered so far. We hope this resource will serve as a starting point to encourage the development and sharing of new open-source, gender-ambiguous voices within the HCI research community.

4.5 Implication of Ambiguous Voices

Integrating ambiguous voices into voice assistants or robots may introduce challenges, such as potential confusion or discomfort, given that users are accustomed to gendered voices. Not all users might feel comfortable interacting with voices that do not fit traditional gender categories. Additionally, these voices might face biases or prejudices from users with strong gender preconceptions, which could affect acceptance and usage of the technology. However, by developing and using ambiguous voices, we do not mean that technology should remove gendered voices, but rather offer the user the possibility of choosing between gendered and ambiguous voices.

We believe that incorporating non-binary voices is a natural evolution in response to a changing world, driven by several important factors: 1) Inclusivity and Representation: gender-ambiguous voices can make technology more inclusive for gender-nonconforming individuals, fostering systems that do not enforce binary gender norms, 2) Reducing Stereotypes: by moving away from traditional gender markers, such as assigning soft, kind voices to women and authoritative ones to men, systems can prioritise functionality over conforming to gender expectations, 3) Creative Voice Design: expanding beyond binary voices allows designers to experiment with a broader range of voice parameters, such as frequency and speed, leading to a richer diversity of vocal expressions that are not constrained by gender norms.

4.6 Recommendations

The flowchart in Figure 5 outlines a step-by-step guide for creating or selecting a gender-ambiguous voice, helping practitioners make informed decisions at each stage of the process:

(1) **Initial Decision**: The process begins by asking whether to create a new ambiguous voice or use an existing one. If an existing voice meets the project needs, Section 4.4 provides open-source gender-ambiguous voices.

*Open Repository of Gender-Ambiguous Voices: https://github.com/martirator/Gender-Ambiguous-Voices

(2) **Voice Characteristics**: For those opting to create a new voice, four key voice characteristics are identified as critical elements to focus on: pitch, resonance, weight, and prosody.

(3) **Evaluation of Gender-Ambiguity**: Voices are then evaluated on a scale ranging from masculine to feminine, with an "ambiguous" option in the middle. The scale includes an "Ambiguous" definition, describing voices that blend characteristics from both ends of the spectrum, thus resisting clear classification as masculine or feminine. In addition, as discussed in Section 4.3, it is crucial also to consider who is conducting the evaluation. When the research involves developing technology that challenges binary gender norms—such as gender-ambiguous voices—it is essential to evaluate this technology with a gender-diverse sample of individuals.

(4) **Open-Source Sharing**: After the voice is created and evaluated, the final step encourages practitioners to make the voice open-source. This promotes collaboration and allows others to build on the work to foster inclusive voice design.

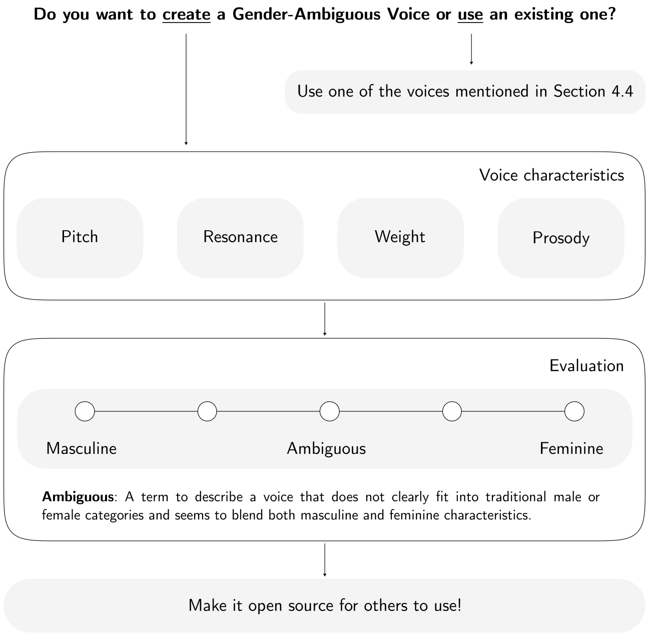

Do you want to create a Gender-Ambiguous Voice or use an existing one?
Use one of the voices mentioned in Section 4.4
Voice characteristics
Pitch
Resonance
Weight
Prosody
Evaluation
Masculine
Ambiguous
Feminine
Ambiguous: A term to describe a voice that does not clearly fit into traditional male or female categories and seems to blend both masculine and feminine characteristics.
Make it open source for others to use!

Figure 5: Flowchart of our suggestion for creating or using a gender-ambiguous voice.

4.7 Future Works and Limitations

We believe that the current paper proposes novel insights into a new and fascinating area of research and potential innovation. However, some limitations should be mentioned. Firstly, the scope of the existing literature is limited, as only 36 papers were found and analysed. However, as interest in inclusive voice technology grows, we are optimistic that more studies will emerge, offering broader perspectives and deeper investigations into this topic.

Furthermore, the review may not capture all relevant literature, especially studies published in languages other than English or in venues outside the ACM Digital Library and IEEE Xplore. Consequently, the generalisability of the findings might be limited. Moreover, the acoustic characteristics analysed were derived from only four voice samples, all of which were in English. This small and linguistically uniform dataset may not sufficiently capture the acoustic features of ambiguous voices across diverse languages and cultural contexts, which limits the generalisability of our findings. For example, certain acoustic markers of gender may vary significantly across tonal versus non-tonal languages, or languages with differing phonological structures. Future research should aim to address these gaps by incorporating a more diverse and representative dataset, including voice samples from various languages and cultural backgrounds. Broadening the scope of literature reviews to include studies from less widely indexed databases or those published in languages other than English would further enrich the understanding of this field. Similarly, the pilot study reported in Section 3.3.1 only compared 4 gender-ambiguous TTS voices. As more and more voices are developed and become available, a more comprehensive comparison should be performed. Additionally, this pilot study focused specifically on the gender perception of these voices, but did not delve into the perception of other traits that would be important for usability, such as e.g. friendliness, trustworthiness, representativeness, or likeability. Future studies should assess user preferences of gender-ambiguous voices beyond gender itself.

The studies included in this review used diverse methodologies for creating and evaluating gender-ambiguous voices. The lack of standardised methods poses a challenge for future works for drawing uniform guidelines, as different studies may yield results that are not directly comparable. We recommend that future research align with consistent evaluation methods, potentially adopting and validating the evaluation scale we proposed (see Figure 4).

As discussed in Section 4.5, gender-ambiguous voices hold great potential for inclusivity and representation of gender-non-conforming individuals. However, our review found that only one study [15] involved non-binary and transgender individuals in the voice creation process. Future research should include these users to assess whether they find gender-ambiguous voices a useful resource and if they feel represented. This will help establish more inclusive guidelines and ensure voice technologies effectively meet the needs of diverse user groups.

In conclusion, it is important to note that the study of gender-ambiguous voices is still evolving within HCI, with concepts, terminologies, and practices continuously developing. Thus, our findings and guidelines may need to be updated as new research emerges.

5 Conclusion

In this study, we conducted a systematic review of the current research on gender-ambiguous voices within the Human-Computer Interaction field, focusing on three key areas. First, we examined how gender-ambiguous voices are defined in the literature. Second, we evaluated the availability, characteristics, and creation methods of existing gender-ambiguous voices. Finally, we explored user perceptions and evaluation methods to assess the effectiveness and usability of these voices. This comprehensive approach aimed to clarify the state of the field and guide future research in creating and utilising gender-ambiguous voices.

Exploring the creation, use, and evaluation of voices that go beyond the binary has revealed several research gaps and opportunities. These include inconsistencies in terminology, and a lack of standardised evaluation methods, guidelines for voice creation, and a clear understanding of how gender-ambiguous voices should sound. Specifically, terms such as "gender-neutral", "gender-ambiguous", "genderless", "gender-free", "non-binary", and "androgynous" are used interchangeably, leading to confusion; thus, we proposed definitions to clarify these terms. Evaluation methods also vary widely across studies, highlighting the need for standardised ones to enhance comparability and accuracy. Lastly, there are no guidelines on how a gender-ambiguous voice should sound. In response to this, we offer ranges for certain voice characteristics that aid in making a voice sound ambiguous.

In conclusion, this study finds that research on gender-ambiguous voices is an emerging field, significantly shaped by recent advancements in speech technology. The analysis of 36 studies reveals both progress and limitations in this area, highlighting the need for clear definitions, standardised evaluation methods, and the development of open-source voices. As the field continues to evolve, it is essential to critically assess current methodologies and establish robust guidelines to guide future research. This foundational work provides a starting point for more nuanced exploration and innovation, setting the stage for further advancements in creating inclusive and effective gender-ambiguous voices.

References

[1] Sean Andrist, Erin Spannan, and Bilge Mutlu. 2013. Rhetorical robots: Making robots more effective speakers using linguistic cues of expertise. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE Press, New York, NY, USA, 341–348. https://doi.org/10.1109/HRI.2013.6483608

[2] 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

[3] Gaye Aşkın, İmge Saltık, Tuğçe Elver Boz, and Burcu A Urgen. 2023. Gendered actions with a genderless robot: Gender attribution to humanoid robots in action. *International Journal of Social Robotics* 15, 11 (2023), 1915–1931.

[4] Minja Axelsson, Mattia Racca, Daryl Weir, and Ville Kyrki. 2019. A Participatory Design Process of a Robotic Tutor of Assistive Sign Language for Children with Autism. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, New York, NY, USA, 1–8. https://doi.org/10.1109/RO-MAN46459.2019.8956309

[5] Adrian Keith Ball, David C. Rye, David Silvera-Tawil, and Mari Velonaki. 2017. How should a robot approach two people? *J. Hum.-Robot Interact.* 6, 3 (dec 2017), 71–91. https://doi.org/10.5898/JHRI.6.3.Ball

[6] Victoria L. Bergvall, Janet Mueller Bing, and Alice F. Freed. 1996. *Rethinking Language and Gender Research: Theory and Practice.* Longman, London. https://digitalcommons.odu.edu/english_books/7 English Faculty Bookshelf, Book 7.

[7] Paul Boersma and David Weenink. 2001. Praat, a system for doing phonetics by computer. http://www.praat.org/.

[8] 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 *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE Press, New York, NY, USA, 13–21.

[9] Judith Butler. 1990. *Gender Trouble: Feminism and the Subversion of Identity.* Routledge, New YOrk.

[10] Julia Cambre and Chinmay Kulkarni. 2019. One Voice Fits All? Social Implications and Research Challenges of Designing Voices for Smart Devices. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 223 (Nov. 2019), 19 pages. https://doi.org/10.1145/3359325

[11] J. Carpenter. 2019. Why Project Q Is More Than the World's First Nonbinary Voice for Technology. *Interactions* 26, 6 (2019), 56–59. https://doi.org/10.1145/3358912

[12] Caroline Gjerlund Christiansen, Sidsel Hardt, Stine Falgren Jensen, Kerstin Fischer, and Oskar Palinko. 2022. Speech Impact in a Usability Test - A Case Study of the KUBO Robot. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE Press, New York, NY, USA, 723–726. https://doi.org/10.1109/HRI53351.2022.9889621

[13] Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20, 1 (1960), 37–46. https://doi.org/10.1177/001316446002000104

[14] Charles Crowell, Matthias Scheutz, Paul Schermerhorn, and Michael Villano. 2009. Gendered voice and robot entities: perceptions and reactions of male and female subjects. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE Press, New York, NY, USA, 3735–3741.

[15] 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* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 390, 17 pages. https://doi.org/10.1145/3544548.3581281

[16] Martina De Cet, Miriam Sturdee, Mohammad Obaid, and Ilaria Torre. 2025. Sketching Robots: Exploring the Influence of Gender-Ambiguous Voices on Robot Perception. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction.*

[17] Dictionary.com. 2023. Text-to-Speech. https://www.dictionary.com/browse/text-to-speech Accessed: 2024-06-13.

[18] Sachi Edirisinghe, Satoru Satake, Drazen Brscic, Yuyi Liu, and Takayuki Kanda. 2024. Field Trial of an Autonomous Shopworker Robot that Aims to Provide Friendly Encouragement and Exert Social Pressure. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 194–202. https://doi.org/10.1145/3610977.3635007

[19] Ana Paula Engelbert. 2014. Cross-Linguistic Effects on Voice Quality: A Study on Brazilians' Production of Portuguese and English. *Concordia Working Papers in Applied Linguistics* 5 (01 2014), 157.

[20] Friederike Eyssel and Frank Hegel. 2012. (S)he's Got the Look: Gender Stereotyping of Robots 1. *Journal of Applied Social Psychology* 42 (07 2012). https://doi.org/10.1111/j.1559-1816.2012.00937.x

[21] 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. (03 2012). https://doi.org/10.1145/2157689.2157717

[22] R.A. Faerber and J.L. Garloch. 2000. Usability evaluation of speech synthesis and recognition for improving the human interface to next generation data link communication systems. In *19th DASC. 19th Digital Avionics Systems Conference. Proceedings (Cat. No.00CH37126)*, Vol. 2. IEEE Press, New York, NY, USA, 5A4/1–5A4/7 vol.2. https://doi.org/10.1109/DASC.2000.884871

[23] European Institute for Gender Equality. 2024. Gender-neutral. https://eige.europa.eu/publications-resources/thesaurus/terms/1321?language_content_entity=en Accessed: 2024-12-04.

[24] Marina Fridin and Mark Belokopytov. 2014. Acceptance of socially assistive humanoid robot by preschool and elementary school teachers. *Comput. Hum. Behav.* 33 (apr 2014), 23–31. https://doi.org/10.1016/j.chb.2013.12.016

[25] Haley N. Green, Md Mofijul Islam, Shahira Ali, and Tariq Iqbal. 2022. Who's Laughing NAO? Examining Perceptions of Failure in a Humorous Robot Partner. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE Press, New York, NY, USA, 313–322. https://doi.org/10.1109/HRI53351.2022.9889353

[26] Andrea Guzman. 2017. *Making AI Safe for Humans: A Conversation With Siri.* Routledge, London, UK, 69–85.

[27] Cory J. Hayes, Charles R. Crowell, and Laurel D. Riek. 2013. Automatic processing of irrelevant co-speech gestures with human but not robot actors. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction* (Tokyo, Japan) *(HRI '13)*. IEEE Press, New York, NY, USA, 333–340.

[28] Inês Hipólito, Katie Winkle, and Merete Lie. 2023. Enactive artificial intelligence: subverting gender norms in human-robot interaction. *Frontiers in Neurorobotics* 17 (2023), 1149303.

[29] Chin-Chang Ho and Karl F. MacDorman. 2010. Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* 26, 6 (nov 2010), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

[30] Sharone Amalia Horowit-Hendler. 2020. *Navigating the Binary: Gender Presentation of Non-Binary Individuals.* PhD Dissertation. University at Albany, State University of New York. https://scholarsarchive.library.albany.edu/legacy-etd/2485

[31] Chien-Ming Huang and Bilge Mutlu. 2012. Robot behavior toolkit: Generating effective social behaviors for robots. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE Press, New York, NY, USA, 25–32.

[32] Chien-Ming Huang and Bilge Mutlu. 2013. The repertoire of robot behavior: enabling robots to achieve interaction goals through social behavior. *J. Hum.-Robot Interact.* 2, 2 (jun 2013), 80–102. https://doi.org/10.5898/JHRI.2.2.Huang

[33] Ivy S. Huang, Yoyo W.Y. Cheung, and Johan F. Hoorn. 2023. Loving-kindness and walking meditation with a robot: Countering negative mood by stimulating

creativity. *International Journal of Human-Computer Studies* 179 (2023), 103107. https://doi.org/10.1016/j.ijhcs.2023.103107

[34] Daeun Hwang and Younah Kang. 2023. How Does Constructive Feedback in an Educational Game Sound to Children? *International Journal of Child-Computer Interaction* 36 (2023), 100581. https://doi.org/10.1016/j.ijcci.2023.100581

[35] J. Jackson and R. D. A. Taylor. 2019. Vocal pitch and intonation characteristics of those who are gender non-binary. In *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019)*. International Phonetic Association, London, UK. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_2734.pdf

[36] Iris Jestin, Joel Fischer, Maria Jose Galvez Trigo, David Large, and Gary Burnett. 2022. Effects of Wording and Gendered Voices on Acceptability of Voice Assistants in Future Autonomous Vehicles. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) *(CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 24, 11 pages. https://doi.org/10.1145/3543829.3543836

[37] Keith Johnson. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34 (10 2006), 485–499. https://doi.org/10.1016/j.wocn.2005.08.004

[38] Barbara Kitchenham. 2004. Procedures for Performing Systematic Reviews. *Keele, UK, Keele Univ.* 33 (08 2004).

[39] Xaver Koch and Esther Janse. 2016. Speech rate effects on the processing of conversational speech across the adult life spana). *The Journal of the Acoustical Society of America* 139, 4 (04 2016), 1618–1636. https://doi.org/10.1121/1.4944032 arXiv:https://pubs.aip.org/asa/jasa/article-pdf/139/4/1618/15316310/1618_1_online.pdf

[40] Johanna Magdalena Kuch, Frank Melchior, and Christian Becker-Asano. 2023. Effects of gender neutralization on the anthropomorphism of natural and synthetic voices. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, New York, NY, USA, 2080–2085. https://doi.org/10.1109/RO-MAN57019.2023.10309479

[41] Seattle Voice Lab. 2024. Androgynous Voice. https://www.seattlevoicelab.com/services/androgynous/#:~:text=What%20Is%20Considered%20the%20Androgynous,your%20range%20in%20either%20direction. Accessed: 2024-09-07.

[42] Robin Lakoff. 1973. Language and Woman's Place. *Language in Society* 2, 1 (1973), 45–80. http://www.jstor.org/stable/4166707

[43] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How weird is CHI?. In *Proceedings of the 2021 chi conference on human factors in computing systems*. 1–14.

[44] Irene Lopatovska, Diedre Brown, and Elena Korshakova. 2022. Contextual Perceptions of Feminine-, Masculine- and Gender-Ambiguous-Sounding Conversational Agents. In *Information for a Better World: Shaping the Global Future*, Malte Smits (Ed.). Springer International Publishing, Cham, 459–480.

[45] Amama Mahmood and Chien-Ming Huang. 2024. Gender Biases in Error Mitigation by Voice Assistants. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 60 (apr 2024), 27 pages. https://doi.org/10.1145/3637337

[46] Alex Mari, Andreina Mandelli, and René Algesheimer. 2024. Empathic voice assistants: Enhancing consumer responses in voice commerce. *Journal of Business Research* 175 (2024), 114566. https://doi.org/10.1016/j.jbusres.2024.114566

[47] Lux Miranda, Ginevra Castellano, and Katie Winkle. 2024. A Case for Diverse Social Robot Identity Performance in Education. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 28–35. https://doi.org/10.1145/3610978.3640768

[48] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and The PRISMA Group. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine* 6, 7 (07 2009), 1–6. https://doi.org/10.1371/journal.pmed.1000097

[49] Nora Ni Loideain and Rachel Adams. 2019. From Alexa to Siri and the GDPR: The gendering of Virtual Personal Assistants and the role of Data Protection Impact Assessments. *Computer Law and Security Review* 36 (12 2019), 105366. https://doi.org/10.1016/j.clsr.2019.105366

[50] Andreea Niculescu, Dennis Hofs, Betsy van Dijk, and Anton Nijholt. 2010. How the agent's gender influence users' evaluation of a QA system. In *2010 International Conference on User Science and Engineering (i-USEr)*. IEEE Press, New York, NY, USA, 16–20. https://doi.org/10.1109/IUSER.2010.5716715

[51] Tatsuya Nomura and Satoru Takagi. 2011. Exploring effects of educational backgrounds and gender in human-robot interaction. In *2011 International Conference on User Science and Engineering (i-USEr)*. IEEE Press, New York, NY, USA, 24–29. https://doi.org/10.1109/iUSEr.2011.6150530

[52] Anna Offenwanger, Alan John Milligan, Minsuk Chang, Julia Bullard, and Dongwook Yoon. 2021. Diagnosing bias in the gender representation of HCI research participants: how it happens and where we are. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[53] Emmi Parviainen and Marie Louise Juul Søndergaard. 2020. Experiential Qualities of Whispering with Voice Assistants. 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, 1–13. https://doi.org/10.1145/3313831.3376187

[54] Connected Speech Pathology. 2024. Crafting an Androgynous Voice for Gender-Neutral Expression. https://connectedspeechpathology.com/blog/crafting-an-androgynous-voice-for-gender-neutral-expression#:~:text=An%20androgynous%20voice%20type%20avoids,masculine%22%20voice%20sounds%20and%20styles. Accessed: 2024-09-07.

[55] Tomislav Pejsa, Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2015. Gaze and Attention Management for Embodied Conversational Agents. *ACM Trans. Interact. Intell. Syst.* 5, 1, Article 3 (mar 2015), 34 pages. https://doi.org/10.1145/2724731

[56] Zhenhui Peng, Yunhwan Kwon, Jiaan Lu, Ziming Wu, and Xiaojuan Ma. 2019. Design and Evaluation of Service Robot's Proactivity in Decision-Making Support Process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300328

[57] Erwan Pépiot. 2014. Voice, speech and gender: Male-female acoustic differences and cross-language variation in English and French speakers. *Corela* 12, 1 (2014). https://doi.org/10.4000/corela.3783

[58] Pranav Prabha, Sreejith Sasidharan, Devasena Pasupuleti, Anand Das, Gayathri Manikutty, and Rajesh Sharma. 2022. A Minimalist Social Robot Platform for Promoting Positive Behavior Change Among Children. In *ACM SIGGRAPH 2022 Educator's Forum* (Vancouver, BC, Canada) *(SIGGRAPH '22)*. Association for Computing Machinery, New York, NY, USA, Article 3, 2 pages. https://doi.org/10.1145/3532724.3535597

[59] Ido Ramati. 2024. Algorithmic Ventriloquism: The Contested State of Voice in AI Speech Generators. *Social Media+ Society* 10, 1 (2024), 20563051231224401.

[60] Melanie A. Revilla, Willem E. Saris, and Jon A. Krosnick. 2014. Choosing the Number of Categories in Agree–Disagree Scales. *Sociological Methods & Research* 43, 1 (2014), 73–97. https://doi.org/10.1177/0049124113509605 arXiv:https://doi.org/10.1177/0049124113509605

[61] Michal Rinott, Shachar Geiger, Neil Nenner, Ori Topaz, Ayelet Karmon, and Kiersten Blake. 2021. Designing an embodied conversational agent for a learning space. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1324–1335. https://doi.org/10.1145/3461778.3462108

[62] Davit Rizhinashvili, Abdallah Hussein Sham, and Gholamreza Anbarjafari. 2022. Gender Neutralisation for Unbiased Speech Synthesising. *Electronics* 11, 10 (2022). https://doi.org/10.3390/electronics11101594

[63] Emma J. Rose and Elin A. Björling. 2017. Designing for engagement: using participatory design to develop a social robot to measure teen stress. In *Proceedings of the 35th ACM International Conference on the Design of Communication* (Halifax, Nova Scotia, Canada) *(SIGDOC '17)*. Association for Computing Machinery, New York, NY, USA, Article 7, 10 pages. https://doi.org/10.1145/3121113.3121212

[64] Gayle Rubin. 1975. The Traffic in Women: Notes on the 'Political Economy' of Sex. In *Toward an Anthropology of Women*, Rayna R. Reiter (Ed.). Monthly Review Press, New York, 157–210.

[65] Morgan Klaus Scheuerman, Katta Spiel, Oliver L Haimson, Foad Hamidi, and Stacy M Branham. 2020. HCI guidelines for gender equity and inclusivity. *UMBC Faculty Collection* (2020).

[66] Ari Schlesinger, W Keith Edwards, and Rebecca E Grinter. 2017. Intersectional HCI: Engaging identity through gender, race, and class. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 5412–5427.

[67] Katie Seaborn, Norihisa P. Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. 2021. Voice in Human–Agent Interaction: A Survey. *ACM Comput. Surv.* 54, 4, Article 81 (may 2021), 43 pages. https://doi.org/10.1145/3386867

[68] Katie Seaborn, Somang Nam, Julia Keckeis, and Tatsuya Itagaki. 2023. Can Voice Assistants Sound Cute? Towards a Model of Kawaii Vocalics. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 63, 7 pages. https://doi.org/10.1145/3544549.3585656

[69] Katie Seaborn, Katja Rogers, Somang Nam, and Miu Kojima. 2023. Kawaii Game Vocalics: A Preliminary Model. In *Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Stratford, ON, Canada) *(CHI PLAY Companion '23)*. Association for Computing Machinery, New York, NY, USA, 202–208. https://doi.org/10.1145/3573382.3616099

[70] Catherine E. Sembroski, Marlena R Fraune, and Selma Šabanović. 2017. He said, she said, it said: Effects of robot group membership and human authority on people's willingness to follow their instructions. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, New York, NY, USA, 56–61. https://doi.org/10.1109/ROMAN.2017.8172280

[71] Speechgen.io. 2024. Speechgen.io. https://speechgen.io Accessed: 2024-05-18.

[72] Katta Spiel, Oliver L Haimson, and Danielle Lottridge. 2019. How to do better with gender on surveys: a guide for HCI researchers. *Interactions* 26, 4 (2019), 62–65.

[73] Yolande Strengers and Jenny Kennedy. 2020. *The Smart Wife: Why Siri, Alexa, and Other Smart Home Devices Need a Feminist Reboot.* https://doi.org/10.7551/mitpress/12482.001.0001

[74] Selina Jeanne Sutton. 2020. Gender Ambiguous, not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 11, 8 pages. https://doi.org/10.1145/3405755.3406123

[75] Selina Jeanne Sutton, Paul Foulkes, David Kirk, and Shaun Lawson. 2019. Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300833

[76] Daniel Szafir and Bilge Mutlu. 2012. Pay attention! designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/2207676.2207679

[77] Eva Székely, Joakim Gustafson, and Ilaria Torre. 2023. Prosody-controllable gender-ambiguous speech synthesis: a tool for investigating implicit bias in speech perception. In *ISCA*. International Speech Communication Association.

[78] João Teixeira and André Gonçalves. 2014. Accuracy of Jitter and Shimmer Measurements. *Procedia Technology* 16 (12 2014), 1190–1199. https://doi.org/10.1016/j.protcy.2014.10.134

[79] Suzanne Tolmeijer, Naim Zierau, Andreas Janson, Jalil Wahdatehagh, Jan Marco Leimeister, and Abraham Bernstein. 2021. Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3411763.3451623

[80] 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 Press, New York, NY, USA, 106–112. https://doi.org/10.1109/RO-MAN57019.2023.10309500

[81] Pinar Uluer, Hatice Kose, Bulent Koray Oz, Turgut Can Aydinalev, and Duygun Erol Barkana. 2020. Towards An Affective Robot Companion for Audiology Rehabilitation: How Does Pepper Feel Today?. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, New York, NY, USA, 567–572. https://doi.org/10.1109/RO-MAN47096.2020.9223534

[82] UNESCO. 2019. I'd Blush If I Could: Closing Gender Divides in Digital Skills through Education. https://unesdoc.unesco.org/ark:/48223/pf0000367416.locale=en/. Accessed: 2024-08-22.

[83] Bert Weijters, Elke Cabooter, and Niels Schillewaert. 2010. The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing* 27, 3 (2010), 236–247. https://doi.org/10.1016/j.ijresmar.2010.02.004

[84] 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.

[85] Katie Winkle, Donald McMillan, Maria Arnelid, Katherine Harrison, Madeline Balaam, Ericka Johnson, and Iolanda Leite. 2023. Feminist human-robot interaction: Disentangling power, principles and practice for better, more ethical HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 72–82.

[86] Katie Winkle, Gaspar Melsión, Donald McMillan, and Iolanda Leite. 2021. Boosting Robot Credibility and Challenging Gender Norms in Responding to Abusive Behaviour: A Case for Feminist Robots. 29–37. https://doi.org/10.1145/3434074.3446910

[87] Renee Yoxon. 2024. Beyond Binary: Crafting Your Own Androgynous Sound. https://www.reneeyoxon.com/blog/beyond-binary-crafting-your-own-androgynous-sound Accessed: 2024-09-07.

[88] Chuang Yu, Changzeng Fu, Rui Chen, and Adriana Tapus. 2022. First Attempt of Gender-free Speech Style Transfer for Genderless Robot. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE Press, New York, NY, USA, 1110–1113. https://doi.org/10.1109/HRI53351.2022.9889533

[89] Lotus Zhang, Lucy Jiang, Nicole Washington, Augustina Ao Liu, Jingyao Shao, Adam Fourney, Meredith Ringel Morris, and Leah Findlater. 2021. Social Media through Voice: Synthesized Voice Qualities and Self-presentation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 161 (apr 2021), 21 pages. https://doi.org/10.1145/3449235