Human-Machine Communication
Volume 5, 2022
https://doi.org/10.30658/hmc.5.2

Gender Ambiguity in Voice-Based Assistants: Gender Perception and Influences of Context

Sandra Mooshammer[1] and Katrin Etzrodt[1]

1 TU Dresden, 01069 Dresden, Germany

Abstract

Recently emerging synthetic acoustically gender-ambiguous voices could contribute to dissolving the still prevailing genderism. Yet, are we indeed perceiving these voices as "unassignable"? Or are we trying to assimilate them into existing genders? To investigate the perceived ambiguity, we conducted an explorative 3 (male, female, ambiguous voice) × 3 (male, female, ambiguous topic) experiment (N = 343). We found that, although participants perceived the gender-ambiguous voice as ambiguous, they used a profoundly wide range of the scale, indicating tendencies toward a gender. We uncovered a mild dissolve of gender roles. Neither the listener's gender nor the personal gender stereotypes impacted the perception. However, the perceived topic gender indicated the perceived voice gender, and younger people tended to perceive a more male-like gender.

Keywords: gender neutrality, ambiguity, voice assistants, gender perception, context effects

Introduction

The use of voice-based assistants (VBAs) is rising and voice control is applied to an increasing number of devices (e.g., National Public Media, 2022). VBAs are dialogue systems that can understand human speech and use a synthesized, human-like voice to interact with users (Hoy, 2018; Knote et al., 2019) as personal smart, adaptive, and interactive artificial

CONTACT Sandra Mooshammer • TU Dresden, 01069 Dresden, Germany • sandra.mooshammer@tu-dresden.de

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

assistants (Knote et al., 2019). Alexa, Google Assistant, and Siri so far dominate the VBA market (Deloitte, 2018). The default voice gender[1] of these popular VBAs is mostly female (Abercrombie et al., 2021) and whereas around 70% of VBAs offer only a female voice, few offer only a male one or various gender options (Sey & Fesalbon, 2019). However, gendered voices in technology can elicit gender effects similar to the ones in human interaction (Lee et al., 2019; Nass et al., 1997). Thus, concerns have been raised that gender stereotypes may be reinforced by associating female voices with an assistant role (Abercrombie et al., 2021; West et al., 2019). However, as recent research indicates, female-voiced VBAs are not necessarily considered as female, but are also sometimes regarded as *neutral* or *different*, thus, a unique gender ontology for VBAs may occur (Fortunati et al., 2022).

Explicitly gender-ambiguous VBA voices such as the non-commercial pilot project *Q* are now being developed to address these concerns and reduce gender bias (genderlessvoice, 2022). In addition, Apple started providing an ambiguous voice option for English-speaking Siri, reaching for higher diversity and representation (Perez, 2022). However, recent research raises doubt about the *genderlessness* of such voices by claiming that people will assign a binary gender to them due to stereotypical design or context variables (e.g., Abercrombie et al., 2021; Sutton, 2020). Similar effects of stereotypical contexts have been found for other ambiguous stimuli like mixed-ethnical faces (Freeman et al., 2013). Hence, to investigate if and how gender-ambiguous voices alter gender biases and their effects, it is fundamental to know if they are actually perceived as ambiguous or not and what influences this perception.

To address the above-mentioned questions, we investigated to what extent acoustically gender-ambiguous voices are perceived as gendered or ambiguous and how this perception is altered by individual and contextual characteristics. Specifically, the study focused on the most critical influences on a voice's gender perception as of thematic context, prior personal stereotypes, and individual gender and age. To be able to meaningfully interpret the results, we studied the perception of the gender-ambiguous artificial voice compared to a female and a male gendered one. The following section begins with a review of human gender roles and stereotypes, ambiguity, and influences on human gender perception, which is then applied to the gender perception of artificial voices.

Theoretical Background

Gender Stereotypes

In a social context, the term *stereotype* refers to "qualities perceived to be associated with particular groups or categories of people" (Schneider, 2004, p. 24). These qualities are "*beliefs* about the characteristics, attributes, and behaviors of members of certain groups [. . .] [and]

1. We use the term *gender* instead of *sex* and differentiate between *acoustic gender* and *perceived gender* to describe the VBAs' properties in accordance with other scholars discussing ambiguous voices (e.g., Sutton, 2020) and for several reasons. First, whereas the human voice gender is determined by physical properties connected to biological sex, those factors induce a stereotypical assignment of the voice to a gender. Second, diverging from humans, VBAs do not actually possess a physical, biological sex but are only equipped with designed properties connected to socially determined genders. Third, *sex* would restrict our terminology to the two main biological sexes and hinder the view on identities outside of genderism—thus, on the central aspect of our paper.

theories about how and why certain attributes go together" (Hilton & von Hippel, 1996, p. 240). Whereas some literature suggests that many researchers conceptualize stereotypes to necessarily be inaccurate, rigid, negative, or that they have to be shared by many people (Ashmore & Del Boca, 1981; Schneider, 2004, p. 17), we agree with Schneider (2004, p. 24), who argues that these assumptions are limiting for the consideration of stereotypes as their essence lies in different aspects. At their core, stereotypes are initially mere generalizing assumptions about the association of attributes with certain groups. In contrast, prejudice refers to a (usually negative) attitude toward entire groups or individual group members *based on* stereotypes (Allport, 1954) which serves to create hierarchical status relationships between groups (Dovidio et al., 2010, p. 7). Similarly, discrimination refers to suppressing or excluding behavior *based on* stereotypes and prejudices (Allport, 1954). Hence, stereotypes can result in prejudice and discrimination.

Gender stereotypes result from the observation of (cisgender) men and women in society and ensuing conclusions about these groups' characteristics (Eagly & Wood, 2012; Ellemers, 2018). Thus, Eagly (1987) argued that gender stereotypes are not directly referring to the biological sexes but to their associated social *gender roles*. Gender roles are "shared expectations (about appropriate qualities and behaviors) that apply to individuals on the basis of their socially identified gender" (Eagly, 1987, p. 12). Thus, they are grounded in social norms (Eagly, 1987), displaying assumptions and expectations about how a certain gender ideally should (not) be and should (not) behave (Rudman & Glick, 2008).

Gender stereotypes can be systematized into the two dimensions, *warmth* and *competence*, depending on a group's level of status and degree of competition in intergroup relationships (Fiske et al., 2002; Fiske & Taylor, 2020). In this context, the *male* stereotype is associated with competence but not warmth, the *female* stereotype with warmth but not competence (Fiske, 2017). *Competence* refers to a person's ability to successfully accomplish tasks and is attributed to groups with a higher level of status (Fiske et al., 2002; Fiske & Taylor, 2020). It is associated with concepts of agency (Bakan, 1966; Eagly, 1987) and instrumentality (Sieverding & Alfermann, 1992), referring to task-relatedness, individuality, and the pursuit of competence. Thus, men are believed to be dominant, willing to take risks and performance-driven (Howansky et al., 2019; Williams & Best, 1990). *Warmth* refers to a person's intentions, and is primarily attributed to groups that are associated with a lower level of competition (Fiske et al., 2002; Fiske & Taylor, 2020). It is associated with concepts of communality (Bakan, 1966; Eagly, 1987) and expressivity (Sieverding & Alfermann, 1992), referring to the need for community, social-emotional support, or harmony (Altstötter-Gleich, 2004; Sieverding & Alfermann, 1992). Thus, women are believed to be helpful, emphatic, or friendly (Howansky et al., 2019; Williams & Best, 1990).

This stereotypization transfers to *occupations and hobbies* (Eagly & Wood, 2012) that are perceived as typically *male* or *female*, as various studies concerning different backgrounds, time periods, and samples show (Couch & Sigler, 2001; García-Mainar et al., 2018; Glick et al., 1995; A. J. Hancock et al., 2020; Janssen & Backes-Gellner, 2016; White & White, 2006). According to the dimension *competence*, these studies found that occupations associated with technology, but also power, responsibility, and prestige, such as builders or managers, are mostly perceived as male. In contrast, matching the dimension *warmth*, occupations that involve empathy, care, or knowledge about interpersonal relationships are typically perceived as female (e.g., marriage counselor or nurse).

Perception of Gender

Stereotypes can shape the category-based impression formation of new persons, based on the persons' stereotype-related "identifying features" (Schneider, 2004, p. 90) such as physical aspects like facial features, other optical features like the hairstyle (Mason et al., 2006; Rudman & Glick, 2008; Schneider, 2004) or behavioral aspects (Taylor, 1981). If these identifying features are ambiguous, people use other signs along available heuristics (Tversky & Kahneman, 1974) or priming effects (Graham & Lowery, 2004) for the categorization. However, impressions can be formed in any mode on a continuum from categorization to conscious, individual processing of new stimuli without relying on stereotypes (Fiske & Neuberg, 1990). Thus, the use of stereotyping is dependent on attention or personal motivation (Fiske & Neuberg, 1990) as well as emotional state (Smith & Mackie, 2010) and the strength of pre-existing stereotypes (Allen et al., 2009; Son Hing & Zanna, 2010).

Acoustical Voice Gender

Gender perception of voices is usually restricted to hearing; thus, the identifying features are the acoustical parameters that differ between sexes. Though features like creakiness or breathiness (Simpson, 2009) are also discussed, scholars agree that a convincing change in gender perception can be traced back to the *combined shift of* the *fundamental frequency (F0)* and *the formant frequencies (FF)* (Gelfer & Bennett, 2013; Hillenbrand & Clark, 2009; Skuk & Schweinberger, 2014; Whiteside, 1998).

Fundamental frequency (F0) refers to the height of the speaking voice, being on average lower for men (100–120 Hertz) and higher for women (200–220 Hertz) (Fitch, 1990; Gelfer & Bennett, 2013; Ma & Love, 2010; Simpson, 2009). However, there is evidence that older women often have a lower F0 than younger ones due to hormonal changes (D'haeseleer et al., 2011; Ma & Love, 2010), and that the German and English F0 of women has decreased in general (Berg et al., 2017). *Formants (FF)* are resonances that occur in the vocal tract when vowels are produced. As men's vocal tracts are longer on average, formants are deeper in male than in female voices. However, research has only identified a factor of 1.15–1.2 by which the formants for individual vowels of the sexes differ (Hillenbrand et al., 1995; Peterson & Barney, 1952; Wu & Childers, 1991). This is further complicated by the fact that findings for F0 and FF cannot easily be transferred into other languages (Simpson, 2009; Strange et al., 2004).

Perception of Gender Regarding Ambiguous Human Voices

Although male and female voices have typical ranges of the continuous scale of height (F0), there is a span of relative ambiguity in between these ranges. Here, *gender ambiguity* can occur for voices because they lack a distinct assignment to one of the ranges or even more extreme regions. This *ambiguity* ranges between 145–165 Hz (Gallena et al., 2018; Gelfer & Bennett, 2013). Although research lacks a distinct ambiguous range for formant frequencies, they too overlap in their ranges and standard deviations between men and women (Gelfer & Bennett, 2013; Pätzold & Simpson, 1997). Hence, similar to F0, it can be assumed that ambiguity occurs if formants are in between the identified distinct gender frequencies.

Thus, it is plausible that a voice which lies in between typical male and female ranges for both frequencies is acoustically gender-ambiguous. However, research has ambivalent outcomes if and how gender ambiguity is *ascribed* to an acoustically gender-ambiguous voice. Here, the type of measurement seems to be a major factor. Whereas the use of the categories *male*, *female*, and *other* resulted in the assignment of a distinct gender to acoustically ambiguous voices (Mullennix et al., 1995), gradual measurements revealed that perceived ambiguity exists (Bralley et al., 1978; Gallena et al., 2018; A. B. Hancock et al., 2014; Mullennix et al., 1995; van Borsel et al., 2009).

Perception of Gender Regarding Artificial Voices

Although gender perception of human voices can be applied to human-like artificial ones, the perception of the latter has peculiarities and still often depicts contradictory findings. For example, *Q*, which was designed to sound gender-ambiguous, seems to be ambiguous overall. However, two of the producers stated that only 50% of 4,500 participants rated it as ambiguous on a 5-point scale, whereas the other half perceived a gender, equally divided between male and female (MacLellan, 2019).

In addition, findings on ambiguous voices of social robots are ambivalent. For example, three out of six participants in a study by Behrens et al. (2018) rated a gendered robot voice (male or female, but synthesized and deliberately kept mechanical) explicitly as gender-ambiguous. This is underlined by the perception of Amazon's Alexa in a study by Fortunati et al. (2022), where, despite the female name, female personality narration, and female default voice, 20% of participants explicitly labeled Alexa as *neutral* or *different* (from male, female, or neutral). More pronounced even, half of them talked about Alexa without using any gender-specific language, with another 15% using gender-neutral language and pronouns such as *they*. Vice versa, the social robot *Pepper* was supposed to be androgynous in voice and appearance (SoftBank Robotics, 2022), but was more likely to be associated with a female voice based on its appearance in an Irish study (McGinn & Torre, 2019) or to be perceived as a boy in Japanese culture (Sugiyama, 2021). Whereas the differing associations might be due to cultural differences, both indicate that Pepper in its original state is neither related to the typical voice of grown-up men nor is entirely gender-ambiguous. In a further study, meanwhile, 30% of 50 participants judged Pepper, presented with a gender-ambiguous voice designed by the researchers, to be *neither male nor female*. However, the majority (64%) perceived it as male (Bryant et al., 2020). Since the authors were able to rule out other influencing variables such as the gender of the subjects, it is plausible that the voice was not entirely ambiguous. The robot's voice was synthetically generated using different F0 values (unfortunately not explicitly described) while formants and other acoustic parameters were ignored.

Independent if artificial ambiguous voices are accompanied by visual cues, some people actually assign ambiguity to them, whereas others still assign a distinct gender. The extent varies greatly between the different technologies, but fundamental and formant frequencies are seldom reported adequately. Thus, studies can barely be compared. In addition, the finding that even gendered artificial voices may be perceived as ambiguous indicates that the artificiality of the voice as such may cause this perception. To address this issue, the perception of the artificial gender-ambiguous voice needs to be compared with artificial gendered voices. Therefore, we asked:

> ***RQ1*****:** How is the gender of a VBA's ambiguous voice perceived compared to a VBA's distinct male or female voice?

Besides the explicit gender assignment, stereotypes associated with the voice can give further insight into the impression that an ambiguous voice may create in the listeners. Especially in comparison with explicitly gendered voices and gender stereotypes (warmth for women, competence for men), investigating the (non-)ascription of gender stereotypes to the ambiguous voice and, thus, its placement in the SCM, is helpful for understanding its perception and the traits associated with it. This can, in turn, give insights for practitioners on further voice design and gender use for specific tasks. However, there is a dearth of research on the ascription of stereotypes to such voices. Thus, an open research question will be formulated to address these aspects:

> ***RQ2*****:** To what extent are gender-stereotypical traits ascribed to the VBA's ambiguous voice?

Social and Psychological Influences on Voice Gender Perception

Gender assignment to acoustically gender-ambiguous voices can be explained by social and psychological factors. Additional information can be considered when categorizing ambiguous stimuli.

A voice will always be perceived as talking about something and will appear in some form of environment. Thus, an omnipresent, potentially influential variable on gender perception is the *context*. Sutton (2020, p. 6) reduces this to the core points *Activity or Topic* in an essay targeted directly at gender-ambiguous VBAs, stating that the topic is at the same time the VBA's activity. Based on social role theory, she argues that if an ambiguous voice speaks about a gender-stereotyped topic, the voice may be assigned that respective gender. Indeed, female and male voices are further perceived as less feminine and masculine, respectively, when talking about products stereotypically assigned to the other gender (Nass & Brave, 2005). However, there is a dearth of research on this specific hypothesis.

Meanwhile, this focus on the topic might need to be broadened as voices also appear in an environment, which can be stereotyped. Children used higher voices when they are asked to speak like a beautician or nurse, and lower voices when they are asked to speak like a builder or mechanic (Cartei et al., 2020). This voice change included changes in both F0 and formants. The same study also showed significant differences when children should speak according to a neutral occupation (e.g., a student), depending on the age group and gender of the children. The frequencies used in these cases were between those used for stereotypically *male* and *female* occupations. Notably, the children were only asked to speak according to the specific occupational group, but were not directed to speak about a matching topic.

Whereas research about the contextual influences on the perception and categorization of ambiguous voices is scarce, findings for visual ambiguous stimuli indicate a strong influence of the context. Higgins et al. (1985) showed a priming effect of the mentioning of positive or negative attributes on the perception of neutrally described stimuli (animals and humans). Freeman et al. (2013) demonstrated that ambiguous faces (mixed Asian and White) were more likely to be interpreted as Asian against a typical Chinese background image, and as White against a typical American background. Moreover, the finding that certain brain areas are more strongly activated with visual context congruence proves a measurable effect of context effects on physical perception mechanisms (Freeman et al., 2015). Context also has an influence on other factors to be assessed; faces in front of threatening backgrounds, for example, are rated as correspondingly less trustworthy (Brambilla et al., 2018).

By referring to gender stereotype theory, these findings for visual stimuli can be transferred to voice gender perception: Occupations with instrumental traits are associated with men, whereas communal occupations are associated with women. These stereotypical associations could be a sufficient cue to assign a gender to an ambiguous voice when it is talking about a gender-stereotyped topic or appears in a gender-stereotyped environment. Thus, we formulate the third research question:

> ***RQ3*:** How does the topic affect the gender perception of the VBA's ambiguous voice?

In addition to the topic, a person's attitudes, prejudices, and heuristics could partially alter the influence of contextual factors. Research on the stereotyping of faces indicates that a person's *strong implicit prejudice* may impact the ascription of ethnicity to an either happy- or angry-looking mixed-ethnical face (Hugenberg & Bodenhausen, 2004; Hutchings & Haddock, 2008). Similarly, *political and personal attitude* may influence the strength of stereotyping: Conservative U.S. Americans seem more likely to classify a mixed-ethnic face as Black than liberal ones, mediated by personal attitudes toward equal treatment (Krosch et al., 2013). In addition, a person's *social context and experience* may affect availability heuristics as persons living in a multicultural environment are significantly more likely than inhabitants of a primarily White environment to judge mixed-ethnic people as mixed-ethnic (Pauker et al., 2018). Transferred to gender perception, these results indicate that participants' gender stereotypes and social contexts may impact how pronounced their stereotyping—and thus, their gender ascription to the voice—is. Furthermore, *age* might be influential as gender roles are changing over time and older persons might have been socialized with different gender stereotypes than younger ones. Finally, participants' *gender* itself could be influential by the use of a similarity heuristic or the over-exclusion of ambiguous persons from the own ingroup as Bodenhausen and Peery (2009) suggest, meaning that a person identifying as male might lean toward perceiving an ambiguous voice as female and vice versa. Thus, a fourth research question will be included to address possible impacts on the individual level:

> ***RQ4*:** How do individual factors like age, gender, and personal gender stereotypes influence the gender perception of the VBA's ambiguous voice?

Apart from contextual and individual factors, other aspects can also play a role in gender perception. These will be specified briefly as they are wide-ranging and have to be kept as neutral as possible in the study design in order to prevent confounding effects on the perception of the ambiguous voice. First, *details of spoken language* may alter gender perception. These include, for example, pronunciation of words or word endings (Hillenbrand et al., 1995; Trudgill, 1972), phonetic patterns (A. B. Hancock et al., 2014), and also details of word choice and sentence formation (Holtgraves & Leaper, 2014; Newman et al., 2008; Singh, 2001). It is conceivable that such differences—although effects are mostly small and overlap between genders (Holtgraves & Leaper, 2014)—may contribute to gender identification once the frequencies of a voice cannot be categorized: Style and content already allow for the identification of an author's gender via machine learning (Baker, 2014; Cheng et al., 2011).

Second, gender perception of VBAs may be affected by *design elements* such as embodiment (e.g., the visuals of a smart speaker) and names—or *Object and Brand*, as Sutton (2020, p. 4) called it. *Visual cues* can be traced back to the categorization of men and women via optical identifying features (see above) and include hairstyles (Eyssel & Hegel, 2012) and the ratio of hip, waist, and chest (Trovato et al., 2018) in robots, but also subtle cues like color (Cunningham & Macrae, 2011; Hess & Melnyk, 2016), or round versus edged shapes (Lieven et al., 2015; Tilburg et al., 2015) in objects and designs. Through such cues, gender and stereotypical traits are assigned to an object (Hess & Melnyk, 2016), which in turn can influence behavior toward the object itself (Cunningham & Macrae, 2011). Furthermore, specific *names* (Pilcher, 2017) or a *name's sound* (Guevremont & Grohmann, 2015; Slepian & Galinsky, 2016) are associated with a gender, possibly creating the expectation of a likewise-gendered persona for an acoustically ambiguous VBA. Huart et al. (2005) found this association for ambiguous faces.

Method

Design

An online experiment was conducted with German-speaking voices and participants. Participants were randomly assigned to the groups which differed in terms of the topic, but also of the voice gender, which was necessary for meaningful comparisons in terms of gender perception and stereotype assignment as argued above. Thus, although the research questions center mostly around the ambiguous voice, male and female variants were included, resulting in a 3 (male, ambiguous, female voice) × 3 (male, ambiguous, female topic) between-subject design and a total of nine groups.

The voices were designed by the authors using the Google WaveNet technology (Google, 2022) for text-to-speech generation of audio files from the texts and fine-tuned in F0 and FF (the central variables in voice gender perception, see section "Acoustical Voice Gender") with the program "Praat" (Praat, 2020). Linguistic gender markers like typical speech patterns were controlled by using one single voice as a basis for the production of a wide range of variants, subsequently pre-testing those variants for their perceived gender and choosing the final variants based on these results. This ensured that even if possibly influencing gender markers occurred, their effect was already included in the voices' final gender perception, and avoided confounding effect of different speech patterns caused by different speakers. The process of designing and pre-testing is described in detail by Mooshammer and Etzrodt (2022). The final voices' acoustic parameters are specified in A1_ Voice parameters (see our OSF: https://osf.io/39pts/).

The gendered topic was chosen in accordance to the stereotypes associating gender with certain occupations described above. As a stereotypically male topic, *airplanes* were chosen to represent instrumentality and technical work. *Love* was chosen as female topic in association with communality and typical occupations like social work or marriage counselor. The neutral topic *penguins* was chosen according to the occupation *biologist*, as this, along with jobs like accountant or journalist, seems to be perceived mostly neutral (Couch & Sigler, 2001; Teig & Susskind, 2008; White & White, 2006). For each topic, a short text containing an interesting fact was prepared (the texts can be accessed in the study's OSF repository: A2_ Stimulus texts). By choosing a neutral presentation of a fact, gender-typical language details like discourse markers should be kept as scarce as possible.

All voices were presented as the VBA *Kim*. This name does not imply a definite gender due to its use for both sexes in German. Kim was framed as an example for a new VBA system usable on the device on which the participants were currently completing the survey and which the participants should be testing. Kim was presented as disembodied voice with as few visual characteristics as possible to avoid optical influences on gender perception. The setting consisted of a mocked *dialog* with Kim, including a greeting by Kim, a topic-dependent pre-set written question (standing for the participants) and the respective audio files with Kim's answers consisting of the spoken text about airplanes, love, or penguins, which could be activated by the participants. The audio files can be found in our OSF.

Subsequently, participants rated the *perceived gender of voice and topic* on 5-point scales (1 = male, 5 = female, 3 = neither/I cannot judge). *Perceived gender stereotypes* were measured on a scale by Altstötter-Gleich (2004), consisting of 16 randomly rotated items rated on a 5-point scale indicating the factors *instrumentality* and *expressivity*. For *personal gender stereotypes*, participants rated eight tasks which might occur in heterosexual relationships (e.g., *decorating the house* or *proposing marriage*) on a 5-point scale according to whether male or female partners should predominantly perform that task in a relationship (Mills et al., 2012). Since this study was part of a larger study, further questions followed before the sociodemographics including age and gender (offering the options male, female, or diverse) were collected, and participants were debriefed (see also A3_ Survey in the OSF).

Factor Analyses

Principal factor analyses were conducted with R to identify the dimensions of perceived and personal gender stereotypes, using parallel analysis (indicating the intersection of simulated with real data) as an extraction criterium. For perceived stereotypes, the factors *instrumentality* and *expressivity* could be confirmed, using promax rotation, and after excluding the items *willingness to take risks* and *professionality* (KMO = 0.9, $\chi 2(91) = 2753.36$, $p < .001$) due to insufficient communality.

In addition, *willingness to take risks* caused parallel analysis to suggest three factors instead of two. The resulting factors explained 57% of the total variance, had eigenvalues of $\lambda_{Instrumentality}$ = 3.08 and $\lambda_{Expressivity}$ = 4.93 and possessed a high internal consistency ($\alpha_{Instrumentality}$ = 0.85, $\alpha_{Expressivity}$ = 0.93).

For personal gender stereotypes, varimax rotation was used as theory states that having male and female stereotypes is not necessarily interdependent. After excluding two items (*mow the lawn* and *shovel snow*, which loaded onto a third factor), a two-factor solution of *male* and *female* stereotypes was found (KMO = 0.71, $\chi2$ (15) = 392.17, $p < .001$), explaining 61% of total variance. The resulting factors were interpretable as male (λ = 2.41, α = 0.72) and female stereotypes (λ = 1.23, α = 0.58), with *decorating the house* loading inversely on the male factor. Details regarding the factor analyses can be found in the OSF (A4_ Factor analyses). The factor for female stereotypes was used despite the lower α value as it is theoretically grounded, necessary as a counterpart for the male stereotypes, and only consists of three items which can affect α values (Field et al., 2012).

Sample

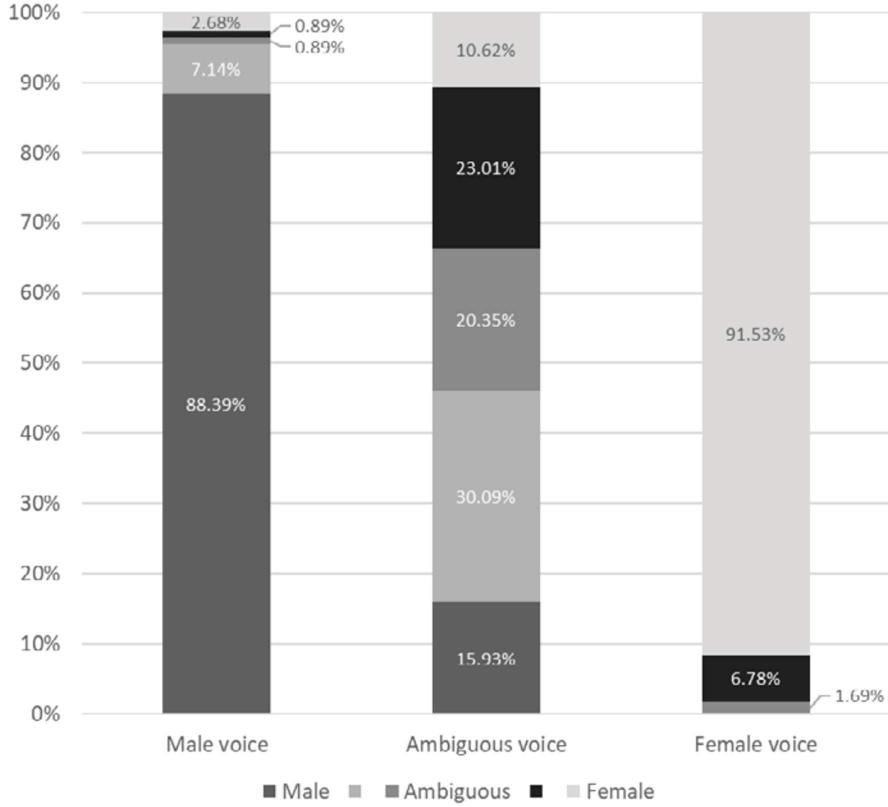
Participants were recruited via online survey platforms and survey exchange groups on social media during May and June 2020. Although 380 persons completed the experiment, 37 had to be excluded due to missing answers (less than 40% of centrally relevant dependent variables), or having spent less than 20 seconds hearing the stimulus, resulting in a final sample size of N = 343, who took an average of 10 minutes to complete the survey. The sample consisted of 58% persons identifying as female and 42% identifying as male (the option *diverse* was offered but not chosen by any of the participants who remained in the final sample), was relatively young (M = 30.04; SD = 11.84) and had achieved a high level of education (36% had a high school degree, 52% a university degree).

Results

Perception of Gender

The analysis of voice gender perception (*RQ1*) showed that although the *gender-ambiguous voice* was rated as approximately ambiguous on the 5-point scale (M = 2.82, SD = 1.26), the standard deviation indicates that the participants used the entire scale. As depicted in Figure 1, only 20% of the participants perceived the voice as fully ambiguous (expressed by choosing the scale midpoint 3), whereas 80% had a tendency toward a gender (male: 46%, female: 34%). However, more than 50% used the *rather male/rather female* categories, whereas only one quarter used the distinct gender poles. In contrast, the *gendered voices* were clearly perceived as male (M = 1.22, SD = 0.69) respectively female (M = 4.90, SD = 0.34), with the majority of about 90% using the scale's poles.

FIGURE 1 Gender Perception of the Three Voices



Ascription of Stereotypical Traits

In terms of ascribed stereotypes (*RQ2*), the acoustically ambiguous voice was associated with an average *instrumentality* (M = 2.97, SD = 0.94), whereas distinctly gendered voices were assigned with a higher instrumentality (male: M = 3.21, SD = 0.82; female: M = 3.25, SD = 0.79). Analyses of variance (ANOVAs) confirmed that the voices differed significantly ($F(2, 340) = 3.54$, $\eta^2 = 0.02$, $p = .030$), with the ambiguous voice perceived as less instrumental than the female voice ($p = .040$) whereas ambiguous and male ($p = .120$) as well as male and female voice did not differ ($p = 1.00$), according to a post-hoc test with Bonferroni correction. Due to the non-normal distribution of instrumentality, robust tests[2] were used in addition, confirming the ANOVA ($F_{rob}(2, 180.4) = 2.91$, $\xi = 0.02$, $p_{rob} = .057$) as well as the post-hoc test's finding for the ambiguous and female voices' difference on a 10% level

2. Robust tests (tests using trimming and estimates to control for non-normal distributions) were conducted using the WRS2-package in R, with t1way for robust ANOVA and lincon for post hoc analysis (Mair & Wilcox, 2020). They were conducted in addition to each standard analysis to validate the results because normal distribution (a requirement of many standard analyses) was not given for many dependent variables.

(p_{rob} = .078). In addition, the difference between ambiguous and male voice was significant on a 10% level (p_{rob} = .078).

Expressivity was also averagely pronounced for the ambiguous voice (M = 2.28, SD = 0.98); however, slightly less strong than instrumentality. This time male voices were perceived as less expressive (M = 2.05, SD = 0.86), female voices as more expressive (M = 2.47, SD = 0.97) than the ambiguous one. These differences were partially confirmed by the ANOVA (F(2, 338) = 6.01, η^2 = 0.03, p = .003) and post-hoc analyses, with male and female voices differing significantly (p = .002). However, the ambiguous voice did neither differ significantly from male (p = .175) nor female voices (p = .372). The robust ANOVA validated the results (F_{rob}(2, 180.98) = 6.55, ξ = 0.23, p_{rob} = .002). Robust post-hocs confirmed differences for male and female (p_{rob} = .001) and indicated a relevant difference between ambiguous and male (p_{rob} = .081) as well as between ambiguous and female voices (p_{rob} = .081) on the 10% level. A closer look on the results for the single stereotypical attributes in the scale revealed that for expressivity, the male voice scored lowest and the female voice highest for every single item, which further strengthens this impression of gender stereotyping for typically female characteristics. In contrast, the scale items for instrumentality are distributed less clear for the three voices.

Regarding the question what gender-stereotypical traits are ascribed to the ambiguous voice (*RQ2*), we thus conclude that it was perceived as more instrumental than expressive. A comparison with the gendered voices showed that in terms of female stereotypes, the ambiguous voice was situated between male and female voices; thus, in fact, representing gender ambiguity. Male stereotypes, however, were rated lowest compared to the gendered voices. An overview of the ascription of instrumentality and expressivity as well as the single stereotypical traits can be found in Figure 2.

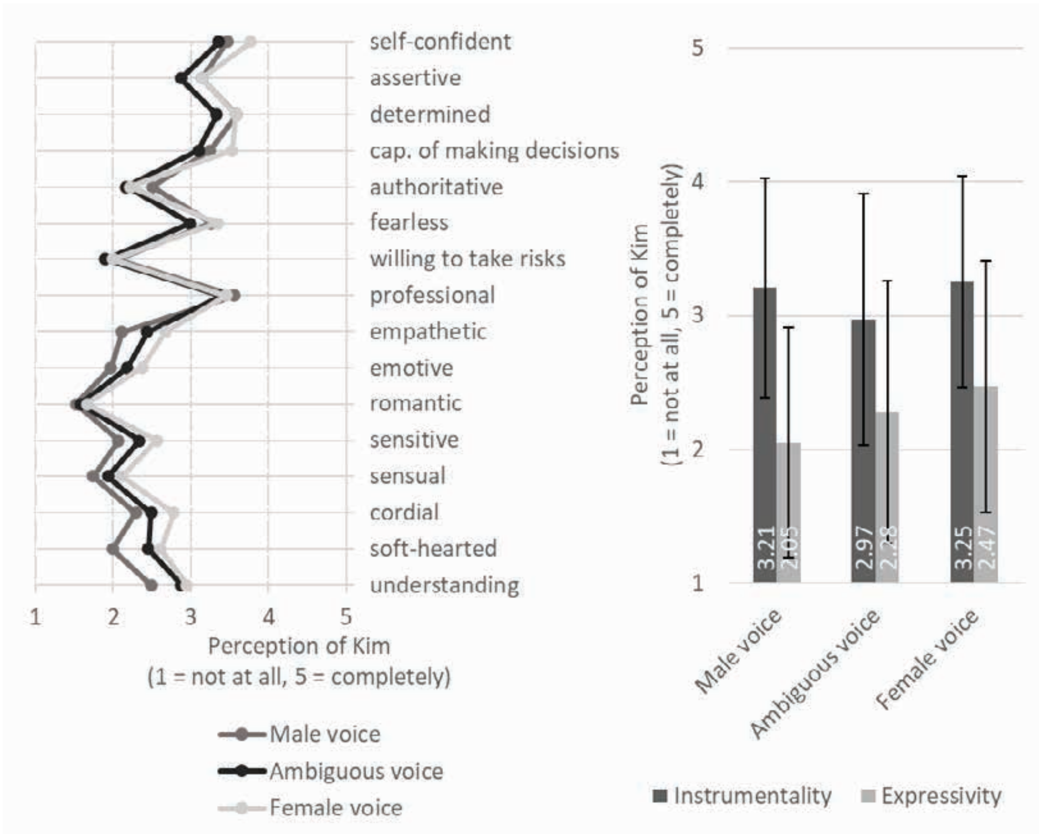
Influences on Gender Perception

To evaluate the possible factors affecting gender perception for acoustically ambiguous voices (*RQ3* & *RQ4*), an ANCOVA[3] was conducted on those participants who had assessed the ambiguous voice stimulus (n = 113). Included predictors were *topic*, *personal gender stereotypes*, age, and gender.

The analysis showed no significance for the influence of the topic (F(2, 106) = 1.03, p = .361, η^2 = 0.02), but revealed a small and significant influence of participants' age (F(1, 106) = 3.99, p = .048, η^2 = 0.04, see Table 1): Older participants perceived the voice to be more female.

3. There was a non-normal distribution in the groups for both predictors, *topic* and *perceived topic gender*. In addition, group sizes were uneven for *perceived topic gender*. This would require robust or nonparametric methods. Currently available methods for robust ANCOVA in R allow only one covariate and a predictor with two groups (Mair & Wilcox, 2020). For this reason, robust regression with dummy variables and the neutral topic as baseline comparison was conducted. However, as the included interaction effects are barely interpretable when applied to every dummy variable separately and the results of dummy variables are not fully comparable to the results of one factorial predictor in an ANCOVA, these were only used for background validation and will not be reported here in detail.

FIGURE 2 Stereotype Ascription to the Voices



Notes: The left figure shows the ascription of single stereotypical items to the voices, the right figure shows the ascription of the male respectively female stereotype dimensions *instrumentality* and *expressivity*, which were calculated from the items contained in the scale depicted on the left as described in the article. As mentioned, instrumentality does not contain the variables *willing to take risks* and *professional* due to the results of the factor analysis.

TABLE 1 ANCOVA Model 1 With Predictor "Topic"

Predictor and Covariates	F(df)	η^2	p
Intercept	$F(1, 106) = 3.80$		.054^{+}
Topic	$F(2, 106) = 1.03$	.019	.361
Gender	$F(1, 106) = 0.02$	.000	.898
Age	$F(1, 106) = 3.99$	.036	**.048***
Male stereotypes	$F(1, 106) = 0.45$	.004	.505
Female stereotypes	$F(1, 106) = 0.00$	.000	.999

It is possible that context effects are induced not by the topic itself, but the subjectively perceived stereotypicality of the topic. Thus, the analyses were repeated, including perceived topic gender as a predictor instead of the actual topic. The ANCOVA uncovered an effect of perceived topic gender on the perceived voice gender, $F(4, 104) = 3.69$, $p = .008$, with an effect size of $\eta^2 = 0.12$. The high significance was confirmed by the robust regression. Age remained influential, $F(1, 104) = 4.15$, $p = .044$, $\eta^2 = 0.04$. The other covariates did not affect the perceived gender (see Table 2). A post-hoc Tukey analysis of the model (predictor *perceived topic gender*, without interactions) uncovered differences in gender perception if the topic was perceived as female, compared to a perception as male (difference = 2.06, $t = 3.51$, $p = .006$) or rather male (difference = 1.41, $t = 2.92$, $p = .033$) (Figure 3).

TABLE 2 ANCOVA Model 2 With Predictor "Perceived Topic Gender"

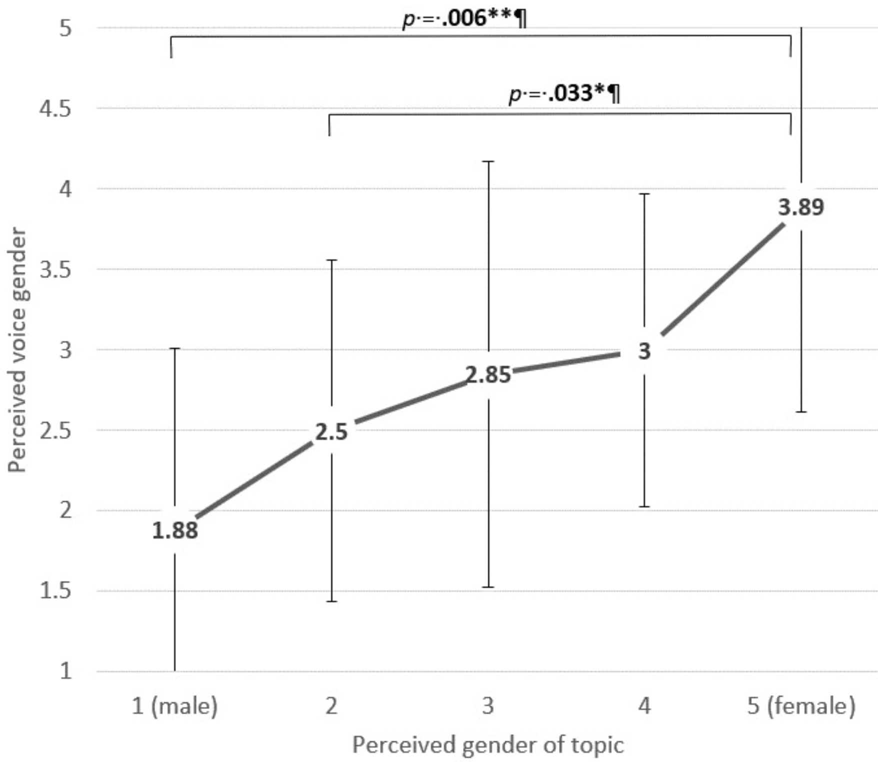
Predictor and Covariates	*F*(df)	η^2	*p*
Intercept	*F*(1, 104) = 2.82		.096^{+}
Perceived topic gender	*F*(4, 104) = 3.69	.124	**.008****
Gender	*F*(1, 104) = 0.01	.000	.914
Age	*F*(1, 104) = 4.15	.038	**.044***
Male stereotypes	*F*(1, 104) = 1.26	.012	.264
Female stereotypes	*F*(1, 104) = 0.00	.000	.956

As the topics were indeed perceived as gendered (airplanes: $M = 2.38$, $SD = 1.04$; penguins: $M = 3.05$, $SD = 0.76$; love: $M = 3.58$, $SD = 1.06$), we conducted another ANCOVA including topic, age, gender, and stereotypes to investigate the effect of topic and control variables on the topic gender perception. It could be demonstrated that topic gender was highly significant ($F(2, 106) = 17.62$, $p < .001$), whereas the covariates showed no effect. Again, these results were controlled by a robust regression due to missing normal distribution of the dependent variable. The robust regression confirmed the influence of topic gender, but additionally showed a significant effect of female stereotypes ($p = .041$, estimate $= 0.46$).

Post-hoc tests of the standard ANCOVA model revealed significant differences between all three groups: The corrected means of the male and ambiguous topic differed by 0.68 ($t = 3.36$, $p = .003$), male and female by 1.21 ($t = 5.92$, $p < .001$), and ambiguous and female topic by 0.53 ($t = 2.61$, $p = .028$).

For the possible influences on gender perception, we found that gendered topics (*RQ3*) were not significant, but showed an indirect effect on voice gender perception via the perceived topic gender. Individual characteristics of the participants (*RQ4*) were also not significant with the exception of age.

FIGURE 3 Effects of Perceived Topic Gender on Perceived Voice Gender of the Ambiguous Voice



Discussion

We explicated that acoustic gender ambiguity is not simply perceived as gender ambiguity. In contrast, in accordance with research on other ambiguous stimuli, the results uncovered that acoustically gender-ambiguous voices are perceived very differently, depending on people's age and the perceived topic gender. Whereas less than a quarter used the scale's center to represent their ambiguous perception, most used the full range of the scale, hesitating to ascribe explicit ambiguity and primarily tending toward a slight gender assignment by preferring the *rather male/rather female* options. Although this can be partially explained by the impact of the perceived gender of the topics, even the people who perceived the topic as neutral used the full scale to categorize the ambiguous voice. In contrast, they were clearly more confident in assessing the gender-distinct voices, using the poles to a greater extent. This indicates that some kind of ambiguity was perceived nonetheless—especially compared to the gender perception for the distinctly gendered voices. In accordance to Piaget (1997), it could be interpreted as an evoked equilibration process due to the uncertainty in gender ascription: Therefore—similar to other ambiguous objects (Etzrodt & Engesser, 2021)—when confronted with ambiguity, people most of the time use the less exhausting strategy of accommodating the voice by modifying an existing category stemming from genderism, instead of building a hybrid or entirely new classification. This finding also has implications for further research as it demonstrates the necessity of using gradual scales for measuring gender perception of ambiguous stimuli. Categories (male, female, ambiguous) could not have captured the slight tendencies that most people displayed.

If a topic was perceived as gendered or gender-neutral, the voice's gender perception was altered, which aligns with prior research. Hence, in accordance with Sutton's (2020) hypothesis based on stereotype theory, it can be assumed that stereotypical context information is used for the categorization of ambiguous stimuli. The impact of the *perceived* stereotypicality of the topics emphasizes the importance of user's perception on context's effects in future research on voice effects, since it is not sufficient to rely on theoretically predicted acoustical ambiguity. It is plausible that this finding applies to further contexts which arise from the embedment of VBAs in real-life situations: Beyond talking about certain topics, VBAs also appear in certain environments. For example, male VBAs are trusted more in work settings, and females more in a home environment (Damen & Toh, 2019). Regarding the important role of contexts, HMC research can build on these findings, exploring further context effects on the perception of ambiguous VBAs. Taking gender ambiguity as an example, other ambiguity categorizations of voices (e.g., ethnical ambiguity) may be affected as well, according to prior research on other stimuli. Siri, for example, included a gender-ambiguous voice and two ethnically diverse options in its voice spectrum. These are perceived as Black or White by almost equal parts of the listeners, in contrast to the "old" voices which are mostly perceived as White, and already evoke racial stereotyping (Holliday, 2022). Categorization could thus also apply to these ethnically ambiguous voices.

Besides the perceived topic gender, age appeared to be the only influential factor, indicating that older people perceived the voice as more ambiguous, whereas younger people tended toward a more male assessment on average. A reason for this might be availability heuristics as described in the theory section: At increasing age, people have had more chances to encounter voices with acoustic parameters that do not fit into the prevailing genderism which might have led to the accommodation (Etzrodt & Engesser, 2021) of their gender scheme, enabling them to classify ambiguity.

Besides the significant effects, the non-effects also have implications as they contradict existing theory. People's personal stereotypes about men and women, for example, did not influence their gender perception of the ambiguous VBA. This may be caused by peculiarities concerning VBAs' gender stereotyping in general: In contrast to stereotype theory, the female stereotype of expressivity was constantly perceived lower than the theoretically male instrumentality for all voices, indicating that stereotypes for VBAs may not entirely resemble human gender stereotypes. Hence, it is plausible that VBAs' application as task-fulfilling assistants in everyday life and their artificiality cause this more instrumental bias. This strengthens previous reflections on the emergence of novel heuristics regarding artificial agents (e.g., Etzrodt & Engesser, 2021; Gambino et al., 2020; Guzman, 2020). If VBAs now have their own heuristics as this indicates, traditional gender stereotypes might not be as relevant for their classification anymore, causing the lack of stereotype effects. In the context of gender, our results point toward a VBA-specific stereotyping that lacks the traditional gender distinction regarding expressivity and instrumentality. However, it remains unclear for now if emerging human gender stereotypes or VBA-specific gender images are not applied as well.

A second indication for a specific VBA gender heuristic is the lack of impact of the theoretically indicated over-exclusion of the ambiguous voice from the participants' own gender. If the VBA was not perceived as a gendered person, but as a gendered *personified thing* (Etzrodt & Engesser, 2021) or *social thing* (Guzman, 2015), the VBA is already part of an outgroup, independent of its gender. A second possibility is the salience of the gender ascription cues serving as identifying features. As spoken information is so closely associated with VBAs and it was presented for around 30 seconds, its perceived stereotypicality might simply have been the more central gender cue in this experimental setting than the participants' own gender. However, this effect might be different in other experimental settings that concern the participants more personally and thus make their own characteristics more relevant for the situation, such as self-disclosure to a VBA in health care.

Limitations and Further Research

Due to the study's explorative character, we only investigated the direct effects of the topic and various covariates. However, literature suggests that moderating variables might also be significant (e.g., the interaction effect between topic and strength of their own stereotypes could possibly influence gender perception when persons with strong stereotypes about women hear the ambiguous voice speaking about a topic they consider as female). Future research could include this and other interactions to further differentiate existing effects.

The topic stimuli and the measurement of personal gender stereotypes were conceptualized in accordance with traditional stereotypes. However, in the past years, there were increasing discussions about sexism (especially against women), the visibility of women in language, public debates, prestigious jobs, and similar topics. For example, the #metoo movement has raised awareness toward gender stereotypes worldwide. Also, in Germany, leading media such as Süddeutsche Zeitung or Spiegel took over the debate with their own theme sites entitled "Sexismus-Debatte" (sexism debate), containing hundreds of articles. Thus, new sensitivity toward gender stereotypes might have caused the dissolution of traditional stereotypes. As a result, the traditionally stereotyped topic gender as well as the measured personal gender stereotypes did not have an effect on voice gender perception. However, the uncovered effect of perceived topic gender on voice gender perception implies that stereotypes are still a salient cue for the meaning-making of the voice's gender—but not in the traditional sense. Therefore, further research on gender stereotypes in general and in HMC needs to consider these new developments when measuring and investigating gender stereotype effects.

Whereas the acoustic voice genders for the male, female, and ambiguous voice were analyzed for all 343 participants, the analysis of the perceived gender of the ambiguous voice was reduced to 113 participants. Especially when investigating the perceived topic gender's impact, this resulted in relatively small group sizes—even more so when taking into account that participants were not evenly distributed across these groups. Although robust methods supplemented the analysis, the detection of smaller effects might have been prevented, even when these could actually be valid. Differences in the effect of perceived topic gender on perceived voice gender could only be detected in the pairwise comparisons of the scale's poles, even though mean values and an accompanying regression analysis suggested a significant linear effect. Thus, we suggest to validate the results with a bigger sample.

Last, a lack of robust testing methods limited our data analysis as a result of the lack of normal distribution in the testing groups. As ANCOVAs with more covariates cannot be conducted robustly and robust regressions do not yield results which are interpretable in the exact same way, we could not validate our results completely. However, as non-normal distributions are neither uncommon in social science research (Wilcox, 2017) nor in human-machine communication (Author), we are looking forward to further developments of robust methods.

Conclusion

In contrast to previous research—primarily located in linguistics and phonetics and, thus, focused on acoustical factors—we adopted a social scientific perspective by investigating the *perception* of a VBA's acoustically gender-ambiguous voice in contrast to its male and female voices and the contextual influences on this perception. We found that, although people were more unsure about the acoustically ambiguous voice of the VBA than its gendered voices, most tried to accommodate their existing gender categories. Only some expressed genuine gender ambiguity. We uncovered that neither the listeners' gender nor their personal gender stereotypes, but rather their age and the topic's perceived gender were influential on the VBA's gender stereotypization. Whereas increasing age supported a more ambiguous assessment, the embedment of the ambiguous VBA voice into a perceived topic gender led more often to the ascription of this respective gender. This indicates that, although there is evidence of gender stereotyping of VBAs, traditional human gender stereotypes and role images cannot be entirely applied to them. In contrast, research on gender stereotypes in HMC needs to consider possible ontological differences between different communicators and resulting new heuristics when investigating the communication with and ascription of (stereotypical) traits to VBAs.

Author Biographies

Sandra Mooshammer (MA) is a PhD student at the Chair of Economic and Political Communication at the Technical University Dresden and a fellow at Schaufler Kolleg@ TU Dresden, where she is currently granted a scholarship. Her research focuses on HMC, concerning aspects of communication with voice-based agents as well as automation and AI in journalism.

https://orcid.org/0000-0003-3556-6517

Katrin Etzrodt (MA) is a research assistant and PhD student at the chair of Science and Technology Communication at the Technical University Dresden. Her research interest is in HMC, aiming to understand several aspects of ambiguity in emerging artificial agents.

https://orcid.org/0000-0001-6515-9985

References

Abercrombie, G., Curry, A. C., Pandya, M., & Rieser, V. (2021). Alexa, Google, Siri: What are your pronouns? Gender and anthropomorphism in the design and perception of conversational assistants. *ArXiv Preprint.* Advance online publication. https://doi.org/10.48550/arXiv.2106.02578

Allen, T. J., Sherman, J. W., Conrey, F. R., & Stroessner, S. J. (2009). Stereotype strength and attentional bias: Preference for confirming versus disconfirming information depends on processing capacity. *Journal of Experimental Social Psychology*, *45*(5), 1081–1087. https://doi.org/10.1016/j.jesp.2009.06.002

Allport, G. W. (1954). *The nature of prejudice* (4th ed.). Addison-Wesley. https://www.semanticscholar.org/paper/The-Nature-of-Prejudice-Allport/03a1bd45ca48412cd2401abb9a1adbe65da0e20c

Altstötter-Gleich, C. (2004). Expressivität, Instrumentalität und psychische Gesundheit. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, *25*(3), 123–139. https://doi.org/10.1024/0170-1789.25.3.123

Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (S. 1–37). L. Erlbaum Associates.

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Rand McNally.

Baker, P. (2014). *Using corpora to analyze gender*. Bloomsbury Publishing.

Behrens, S. I., Egsvang, A. K. K., Hansen, M., & Møllegård-Schroll, A. M. (2018). Gendered robot voices and their influence on trust. In T. Kanda, S. Ŝabanović, G. Hoffman, & A. Tapus (Eds.), *Hri'18 companion, March 5–8, 2018, Chicago, IL, USA* (pp. 63–64). ACM Press. https://doi.org/10.1145/3173386.3177009

Berg, M., Fuchs, M., Wirkner, K., Loeffler, M., Engel, C., & Berger, T. (2017). The speaking voice in the general population: Normative data and associations to sociodemographic and lifestyle factors. *Journal of Voice*, *31*(2), 257. https://doi.org/10.1016/j.jvoice.2016.06.001

Bodenhausen, G. V., & Peery, D. (2009). Social categorization and stereotyping in vivo: The VUCA challenge. *Social and Personality Psychology Compass*, *3*(2), 133–151. https://doi.org/10.1111/j.1751-9004.2009.00167.x

Bralley, R. C., Bull, G. L., Gore, C. H., & Edgerton, M. T. (1978). Evaluation of vocal pitch in male transsexuals. *Journal of Communication Disorders*, *11*(5), 443–449. https://doi.org/10.1016/0021-9924(78)90037-0

Brambilla, M., Biella, M., & Freeman, J. B. (2018). The influence of visual context on the evaluation of facial trustworthiness. *Journal of Experimental Social Psychology*, *78*, 34–42. https://doi.org/10.1016/j.jesp.2018.04.011

Bryant, D., Borenstein, J., & Howard, A. (2020). Why should we gender? In T. Belpaeme, J. Young, H. Gunes, & L. Riek (Eds.), *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 13–21). ACM. https://doi.org/10.1145/3319502.3374778

Cartei, V., Oakhill, J., Garnham, A., Banerjee, R., & Reby, D. (2020). "This is what a mechanic sounds like": Children's vocal control reveals implicit occupational stereotypes. *Psychological Science*, *31*(8), 957–967. https://doi.org/10.1177/0956797620929297

Cheng, N., Chandramouli, R., & Subbalakshmi, K. P. (2011). Author gender identification from text. *Digital Investigation*, *8*(1), 78–88. https://doi.org/10.1016/j.diin.2011.04.002

Couch, J. V., & Sigler, J. N. (2001). Gender perception of professional occupations. *Psychological Reports*, *88*(3 Pt 1), 693–698. https://doi.org/10.2466/pr0.2001.88.3.693

Cunningham, S. J., & Macrae, C. N. (2011). The colour of gender stereotyping. *British Journal of Psychology*, *102*(3), 598–614. https://doi.org/10.1111/j.2044-8295.2011.02023.x

Damen, N., & Toh, C. (2019). Designing for trust: Understanding the role of agent gender and location on user perceptions of trust in home automation. *Journal of Mechanical Design*, *141*(6), 1–11. https://doi.org/10.1115/1.4042223

Deloitte. (2018). *Beyond touch: Voice-commerce 2030.* https://web.archive.org/web/20221010155602/https://www2.deloitte.com/content/dam/Deloitte/de/Documents/consumer-business/Deloitte-Beyond_Touch-Voice_Commerce_2030.pdf

D'haeseleer, E., Depypere, H., Claeys, S., Wuyts, F. L., Baudonck, N., & van Lierde, K. M. (2011). Vocal characteristics of middle-aged premenopausal women. *Journal of Voice*, *25*(3), 360– 366. https://doi.org/10.1016/j.jvoice.2009.10.016

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The Sage handbook of prejudice, stereotyping and discrimination* (pp. 3–28). Sage.

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. *John M. MacEachran memorial lecture series*. Erlbaum. https://web.archive.org/web/20220119083109/http://catdir.loc.gov/catdir/enhancements/fy0742/86032921-d.html

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. van Lange, A. Kruglanski, & E. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 458–476). SAGE Publications Ltd. https://doi.org/10.4135/9781446249222.n49

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, *69*, 275–298. https://doi.org/10.1146/annurev-psych-122216-011719

Etzrodt, K., & Engesser, S. (2021). Voice-based agents as personified things: Assimilation and accommodation as equilibration of doubt. *Human-Machine Communication*, *2*, 57–79. https://doi.org/10.30658/hmc.2.3

Eyssel, F., & Hegel, F. (2012). (S)he's got the look: Gender stereotyping of robots. *Journal of Applied Social Psychology*, *42*(9), 2213–2230. https://doi.org/10.1111/J.1559-1816.2012.00937.X

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.

Fiske, S. T. (2017). Prejudices in cultural contexts: Shared stereotypes (gender, age) versus variable stereotypes (race, ethnicity, religion). *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*(5), 791–799. https://doi.org/10.1177/1745691617708204

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037//0022-3514.82.6.878

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In A. Unknown (Ed.), *Advances in Experimental Social Psychology: Volume 23* (1st ed., pp. 1–74). Elsevier textbooks. https://doi.org/10.1016/S0065-2601(08)60317-2

Fiske, S. T., & Taylor, S. E. (2020). Social cognition evolves: Illustrations from our work on intergroup bias and on healthy adaptation. *Psicothema*, *32*(3), 291–297. https://doi.org/10.7334/psicothema2020.197

Fitch, J. L. (1990). Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech. *The Journal of Speech and Hearing Disorders*, *55*(2), 360–363. https://doi.org/10.1044/jshd.5502.360

Fortunati, L., Edwards, A., Edwards, C., Manganelli, A. M., & de Luca, F. (2022). Is Alexa female, male, or neutral? A cross-national and cross-gender comparison of perceptions of Alexa's gender and status as a communicator. *Computers in Human Behavior, 137*(3), 107426. https://doi.org/10.1016/j.chb.2022.107426

Freeman, J. B., Ma, Y., Barth, M., Young, S. G., Han, S., & Ambady, N. (2015). The neural basis of contextual influences on face categorization. *Cerebral Cortex*, *25*(2), 415–422. https://doi.org/10.1093/cercor/bht238

Freeman, J. B., Ma, Y., Han, S., & Ambady, N. (2013). Influences of culture and visual context on real-time social categorization. *Journal of Experimental Social Psychology*, *49*(2), 206–210. https://doi.org/10.1016/j.jesp.2012.10.015

Gallena, S. J. K., Stickels, B., & Stickels, E. (2018). Gender perception after raising vowel fundamental and formant frequencies: Considerations for oral resonance research. *Journal of Voice*, *32*(5), 592–601. https://doi.org/10.1016/j.jvoice.2017.06.023

Gambino, A., Fox, J., & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*, 71–86. https://doi.org/10.30658/hmc.1.5

García-Mainar, I., Montuenga, V. M., & García-Martín, G. (2018). Occupational prestige and gender-occupational segregation. *Work, Employment and Society*, *32*(2), 348–367. https://doi.org/10.1177/0950017017730528

Gelfer, M. P., & Bennett, Q. E. (2013). Speaking fundamental frequency and vowel formant frequencies: Effects on perception of gender. *Journal of Voice*, *27*(5), 556–566. https://doi.org/10.1016/j.jvoice.2012.11.008

genderlessvoice. (2022). *Meet Q. The first genderless voice.* www.genderlessvoice.com

Glick, P., Wilk, K., & Perreault, M. (1995). Images of occupations: Components of gender and status in occupational stereotypes. *Sex Roles*, *32*(9–10), 565–582. https://doi.org/10.1007/BF01544212

Google. (2022). *Text-to-speech: Lifelike speech synthesis*. https://cloud.google.com/text-to-speech

Graham, S., & Lowery, B. S. (2004). Priming unconscious racial stereotypes about adolescent offenders. *Law and Human Behavior*, *28*(5), 483–504. https://doi.org/10.1023/b:lahu.0000046430.65485.1f

Guevremont, A., & Grohmann, B. (2015). Consonants in brand names influence brand gender perceptions. *European Journal of Marketing*, *49*(1/2), 101–122. https://doi.org/10.1108/EJM-02-2013-0106

Guzman, A. L. (2015). *Imagining the voice in the machine: The ontology of digital social agents*. Dissertation. University of Illinois.

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication*, *1*, 37–54. https://doi.org/10.30658/hmc.1.3

Hancock, A. B., Colton, L., & Douglas, F. (2014). Intonation and gender perception: Applications for transgender speakers. *Journal of Voice*, *28*(2), 203–209. https://doi.org/10.1016/j.jvoice.2013.08.009

Hancock, A. J., Clarke, H. M., & Arnold, K. A. (2020). Sexual orientation occupational stereotypes. *Journal of Vocational Behavior*, *119*, 103427. https://doi.org/10.1016/j.jvb.2020.103427

Hess, A. C., & Melnyk, V. (2016). Pink or blue? The impact of gender cues on brand perceptions. *European Journal of Marketing*, *50*(9/10), 1550–1574. https://doi.org/10.1108/EJM-11-2014-0723

Higgins, E. T., Bargh, J. A., & Lombardi, W. J. (1985). Nature of priming effects on categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 59–69. https://doi.org/10.1037/0278-7393.11.1.59

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f(0) and formant frequencies in distinguishing the voices of men and women. *Attention, Perception & Psychophysics*, *71*(5), 1150–1166. https://doi.org/10.3758/APP.71.5.1150

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5 Pt 1), 3099–3111. https://doi.org/10.1121/1.411872

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, *47*, 237–271. https://doi.org/10.1146/annurev.psych.47.1.237

Holliday, N. (2022). *Siri, you've changed! Acoustic properties and racialized judgments of voice assistants*. https://www.dropbox.com/s/aqow5vypj0pam3g/poster_lsa_2022_siri.pdf?dl=0

Holtgraves, T. M., & Leaper, C. (2014). Gender similarities and differences in language. In T. M. Holtgraves (Ed.), *The Oxford Handbook of Language and Social Psychology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199838639.013.002

Howansky, K., Wilton, L. S., Young, D. M., Abrams, S., & Clapham, R. (2019). (Trans)gender stereotypes and the self: Content and consequences of gender identity stereotypes. *Self and Identity*, 1–18. https://doi.org/10.1080/15298868.2019.1617191

Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An introduction to voice assistants. *Medical Reference Services Quarterly*, *37*(1), 81–88. https://doi.org/10.1080/02763869.2018.1404391

Huart, J., Corneille, O., & Becquart, E. (2005). Face-based categorization, context-based categorization, and distortions in the recollection of gender ambiguous faces. *Journal of Experimental Social Psychology*, *41*(6), 598–608. https://doi.org/10.1016/j.jesp.2004.10.007

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science*, *15*(5), 342–345. https://doi.org/10.1111/j.0956-7976.2004.00680.x

Hutchings, P. B., & Haddock, G. (2008). Look Black in anger: The role of implicit prejudice in the categorization and perceived emotional intensity of racially ambiguous faces. *Journal of Experimental Social Psychology*, *44*(5), 1418–1420. https://doi.org/10.1016/j.jesp.2008.05.002

Janssen, S., & Backes-Gellner, U. (2016). Occupational stereotypes and gender-specific job satisfaction. *Industrial Relations: A Journal of Economy and Society*, *55*(1), 71–91. https://doi.org/10.1111/irel.12126

Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2019). Classifying smart personal assistants: An empirical cluster analysis. In T. Bui (Ed.), *Proceedings of the Annual Hawaii International Conference on System Sciences, Proceedings of the 52nd Hawaii International Conference on System Sciences.* Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2019.245

Krosch, A. R., Berntsen, L., Amodio, D. M., Jost, J. T., & van Bavel, J. J. (2013). On the ideology of hypodescent: Political conservatism predicts categorization of racially ambiguous faces as Black. *Journal of Experimental Social Psychology*, *49*(6), 1196–1203. https://doi.org/10.1016/j.jesp.2013.05.009

Lee, S., Ratan, R., & Park, T. (2019). The voice makes the car: Enhancing autonomous vehicle perceptions and adoption intention through voice agent gender and style. *Multimodal Technologies and Interaction*, *3*(1), 1–11. https://doi.org/10.3390/mti3010020

Lieven, T., Grohmann, B., Herrmann, A., Landwehr, J. R., & van Tilburg, M. (2015). The effect of brand design on brand gender perceptions and brand preference. *European Journal of Marketing*, *49*(1/2), 146–169. https://doi.org/10.1108/EJM-08-2012-0456

Ma, E. P.-M., & Love, A. L. (2010). Electroglottographic evaluation of age and gender effects during sustained phonation and connected speech. *Journal of Voice*, *24*(2), 146–152. https://doi.org/10.1016/j.jvoice.2008.08.004

MacLellan, L. (2019, March 22). *This AI voice is gender-neutral, unlike Siri and Alexa.* Quartz. https://web.archive.org/web/20220806114324/https://qz.com/work/1577597/this-ai-voice-is-gender-neutral-unlike-siri-and-alexa/

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, *52*(2), 464–488. https://doi.org/10.3758/s13428-019-01246-w

Mason, M. F., Cloutier, J., & Macrae, C. N. (2006). On construing others: Category and stereotype activation from facial cues. *Social Cognition*, *24*(5), 540–562. https://doi.org/10.1521/soco.2006.24.5.540

McGinn, C., & Torre, I. (2019). Can you tell the robot by the voice? An exploratory study on the role of voice in the perception of robots. In *Hri'19: The 14th ACMIEEE International Conference on Human-Robot Interaction: March 11–14, 2019, Daegu South Korea* (pp. 211–221). IEEE.

Mills, M. J., Culbertson, S. S., Huffman, A. H., & Connell, A. R. (2012). Assessing gender biases. *Gender in Management: An International Journal*, *27*(8), 520–540. https://doi.org/10.1108/17542411211279715

Mooshammer, S., & Etzrodt, K. (2022). Social research with gender-neutral voices in chatbots—The generation and evaluation of artificial gender-neutral voices with Praat and Google WaveNet. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, E. Luger, M. Goodwin, & P. B. Brandtzaeg (Eds.), *Lecture Notes in Computer Science. Chatbot Research and Design* (Vol. 13171, pp. 176–191). Springer International Publishing. https://doi.org/10.1007/978-3-030-94890-0_11

Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *The Journal of the Acoustical Society of America*, *98*(6), 3080–3095. https://doi.org/10.1121/1.413832

Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. *Computer-human interaction*. MIT Press.

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, *27*(10), 864–876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x

National Public Media. (2022). *The smart audio report*. https://web.archive.org/web/20220901042912/https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, *45*(3), 211–236. https://doi.org/10.1080/01638530802073712

Pätzold, M., & Simpson, A. P. (1997). Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. *Arbeitsberichte Des Instituts Für Phonetik Und Digitale Sprachverarbeitung Universität Kiel*(32), 215–247.

Pauker, K., Carpinella, C. M., Lick, D. J., Sanchez, D. T., & Johnson, K. L. (2018). Malleability in biracial categorizations: The impact of geographic context and targets' racial heritage. *Social Cognition*, *36*(5), 461–480. https://doi.org/10.1521/soco.2018.36.5.461

Perez, S. (2022). *Siri gains a new gender-neutral voice option in latest iOS update*. https://web.archive.org/web/20220305010257/https://techcrunch.com/2022/02/24/siri-gains-a-new-gender-neutral-voice-option-in-latest-ios-update/

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184. https://doi.org/10.1121/1.1906875

Piaget, J. (1997). *The principles of genetic epistemology. Jean Piaget: Selected works* (Original work published 1970) (VII). Routledge.

Pilcher, J. (2017). Names and "doing gender": How forenames and surnames contribute to gender identities, difference, and inequalities. *Sex Roles*, *77*(11), 812–822. https://doi.org/10.1007/s11199-017-0805-4

Praat. (2020, April 6). *Sound: Change gender*. https://web.archive.org/web/20200429120202/https://www.fon.hum.uva.nl/praat/manual/Sound__Change_gender___.html

Rudman, L. A., & Glick, P. (2008). *The social psychology of gender: How power and intimacy shape gender relations*. *Texts in social psychology*. Guilford Press.

Schneider, D. J. (2004). *The psychology of stereotyping. Distinguished contributions in psychology*. Guilford Press. https://web.archive.org/web/20100717230024/http://catdir.loc.gov/catdir/bios/guilford051/2003008819.html

Sey, A., & Fesalbon, L. (2019). "OK Google: Is AI gendered?" In A. Sey & N. Hafkin (Eds.), *Taking stock: Data and evidence on gender equality in digital access, skills and leadership* (pp. 144–145). United Nations University Institute on Computing and Society/International Telecommunications Union: Macau.

Sieverding, M., & Alfermann, D. (1992). Instrumentelles (maskulines) und expressives (feminines) Selbstkonzept: ihre Bedeutung für die Geschlechtsrollenforschung. *Zeitschrift Für Sozialpsychologie*(23), 6–15. https://doi.org/10.11588/HEIDOK.00019637

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640. https://doi.org/10.1111/j.1749-818X.2009.00125.x

Singh, S. (2001). A pilot study on gender differences in conversational speech on lexical in richness measures. *Literary and Linguistic Computing*, *16*(3), 251–264. https://doi.org/10.1093/llc/16.3.251

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language, and Hearing Research*, *57*(1), 285–296. https://doi.org/10.1044/1092-4388(2013/12-0314)

Slepian, M. L., & Galinsky, A. D. (2016). The voiced pronunciation of initial phonemes predicts the gender of names. *Journal of Personality and Social Psychology*, *110*(4), 509–527. https://doi.org/10.1037/pspa0000041

Smith, E. R., & Mackie, D. M. (2010). Affective processes. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The Sage handbook of prejudice, stereotyping and discrimination* (pp. 131–145). Sage.

SoftBank Robotics. (2022). *Design—Pepper is a humanoid-robot*. https://web.archive.org/web/20210706024656/https://developer.softbankrobotics.com/pepper-qisdk/design/pepper-humanoid-robot

Son Hing, L. S., & Zanna, M. P. (2010). Individual differences. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The Sage handbook of prejudice, stereotyping and discrimination* (pp. 163–178). Sage.

Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America*, *115*(4), 1791–1807. https://doi.org/10.1121/1.1687832

Sugiyama, S. (2021). The apparatgeist of pepper-kun: An exploration of emerging cultural meanings of a social robot in Japan. In J. Katz, J. Floyd, & K. Schiepers (Eds.), *Springer eBook collection. Perceiving the future through new communication technologies: Robots, AI and everyday life* (1st ed., pp. 113–128). Springer International Publishing; Imprint Palgrave Macmillan. https://doi.org/10.1007/978-3-030-84883-5_8

Sutton, S. J. (2020). Gender ambiguous, not genderless. In M. I. Torres, S. Schlögl, L. Clark, & M. Porcheron (Eds.), *Proceedings of the 2nd conference on Cconversational user interfaces* (pp. 1–8). ACM. https://doi.org/10.1145/3405755.3406123

Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 83–114). L. Erlbaum Associates.

Teig, S., & Susskind, J. E. (2008). Truck driver or nurse? The impact of gender roles and occupational status on children's occupational preferences. *Sex Roles*, *58*(11–12), 848–863. https://doi.org/10.1007/s11199-008-9410-x

Tilburg, M., Lieven, T., Herrmann, A., & Townsend, C. (2015). Beyond "Pink it and shrink it" perceived product gender, aesthetics, and product evaluation. *Psychology and Marketing*, *32*(4), 422–437. https://doi.org/10.1002/mar.20789

Trovato, G., Lucho, C., & Paredes, R. (2018). She's electric—The influence of body proportions on perceived gender of robots across cultures. *Robotics*, *7*(3), 1–13. https://doi.org/10.3390/robotics7030050

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, *1*(2), 179–195. https://doi.org/10.1017/S0047404500000488

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

van Borsel, J., Pot, K. de, & Cuypere, G. de (2009). Voice and physical appearance in female-to-male transsexuals. *Journal of Voice*, *23*(4), 494–497. https://doi.org/10.1016/j.jvoice.2007.10.018

West, M., Kraut, R., & Ei Chew, H. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*. EQUALS and UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000367416

White, M. J., & White, G. B. (2006). Implicit and explicit occupational gender stereotypes. *Sex Roles*, *55*(3-4), 259–266. https://doi.org/10.1007/s11199-006-9078-z

Whiteside, S. P. (1998). The identification of a speaker's sex from synthesized vowels. *Perceptual and Motor Skills*, *87*(2), 595–600. https://doi.org/10.2466/pms.1998.87.2.595

Wilcox, R. (2017). *Modern statistics for the social and behavioral sciences*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315154480

Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multination study* (Rev. ed.). *Cross-cultural research and methodology series* (Vol. 6). Sage. https://web.archive.org/web/20221010160558/http://catdir.loc.gov/catdir/enhancements/fy0655/90008295-d.html

Wu, K., & Childers, D. G. (1991). Gender recognition from speech. Part I: Coarse analysis. *The Journal of the Acoustical Society of America*, *90*(4 Pt 1), 1828–1840. https://doi.org/10.1121/1.401663