Voice cues influence children's assessment of adults' occupational competence

Article (Accepted Version)

Voice cues influence children's assessment of adults' occupational competence

Authors: Valentina Cartei[1], Jane Oakhill[1], Alan Garnham[1], Robin Banerjee[1], David Reby[2,1]

[1] School of Psychology, University of Sussex, Brighton, UK

[2] Equipe Neuro-Ethologie Sensorielle, ENES / CRNL, CNRS UMR5292, INSERM UMR_S 1028, University of Lyon / Saint-Etienne, France

Corresponding author: Valentina Cartei, Pevensey 2, School of Psychology, University of Sussex, Falmer, BN1 9RH, UK (email address: v.cartei@sussex.ac.uk)

ORCID IDs:

Valentina Cartei https://orcid.org/0000-0001-8315-3595

Alan Garnham https://orcid.org/0000-0002-0058-403X

Robin Banerjee https://orcid.org/0000-0002-4994-3611

David Reby https://orcid.org/0000-0001-9261-1711

Acknowledgments

We are grateful to the Leverhulme Trust for funding this research (grant no. RPG-2016-396). We are also grateful to the children and their parents who agreed to take part in the study, and to the head teachers and staff of the schools who participated. Finally, we would like to thank Lucy Roberts and Dr Sophie Anns for their assistance with the data collection.

Abstract

The adult voice is a strong bio-social marker for masculinity and femininity. In this study we investigated whether children make gender stereotypical judgements about adults' occupational competence on the basis of their voice. Forty-eight 8- to 10- year olds were asked to rate the competence of adult voices that varied in vocal masculinity (by artificially manipulating voice pitch) and were randomly paired with 9 occupations (3 stereotypically male, 3 female, 3 gender-neutral). In line with gender stereotypes, children rated men as more competent for the male occupations and women as more competent for the female occupations. Moreover, children rated speakers of both sexes with feminine (high-pitched) voices as more competent for the female occupations. Finally, children rated men (but not women) with masculine (low-pitched) voices as more competent for stereotypically male occupations. Our results thus indicate that stereotypical voice-based judgements of occupational competence previously identified in adults are already present in children, and likely affect how they consider adults and interact with them in their social environment.

Keywords

occupational stereotypes, gender stereotypes, nonverbal communication, competence, child development

Introduction

The human voice is one of the main sources providing first impressions of a speaker's identity, including biological sex. The perceived biological sex of an adult speaker from their voice is primarily defined by mean fundamental frequency (F0, perceived as voice pitch) and, to a lesser extent, from vocal tract resonances (or formants), which in men are on average 50% and 20% lower, respectively, than women's (Titze 1989; Gelfer and Mikos 2005). In addition to signalling sex, these voice patterns (e.g. relatively lower pitch and resonance in men's voices and relatively higher pitch and resonance in women's voices) influence listeners' attributions of gender, that is the "roles, behaviours, activities, and attributes that any society considers appropriate for girls and boys, and women and men" (World Health Organisation 2020). For example, listeners judge men and women with low-frequency voices as physically bigger, stronger, more masculine, more physically and socially dominant than those with voices of relatively high-frequency voices (for reviews: Hall et al. 2005; Pisanski and

Bryant 2019). These associations can be partly explained in evolutionary terms, as voice pitch, at least in males, is inversely related to is testosterone (Cartei et al., 2020b; O'Connor, Re, & Feinberg, 2011), which in turn is positively associated with a host of physiological masculine characteristics, including physical strength and body size (Bhasin et al. 2001), as well as self-reported dominance (Puts et al., 2006). At the same time, listeners have a tendency to overgeneralise the sex dimorphism that characterises the voice of adult speakers, resulting in sex-stereotype biases in judgement patterns. For instance, the perceived association between pitch and body size may lead to misattributions of physical strength in adults (Feinberg et al., 2005; Fitch, 1997), and of sex in babies (e.g. low-pitched cries are more likely to be attributed to boys and high-pitched cries to girls, despite the absence of sex differences in pitch: Reby et al., 2016).

Although most of extant research focuses on the impact of vocal masculinity and femininity on listeners' perceptions of speakers within intrasexual competition or mate choice contexts, a few studies have helped uncover the wider socio-economic implications of speaker attributions. Like masculine-looking men and women (Todorov et al. 2005; Little 2014; Re and Rule 2017; Rule and Ambady 2009; Sczesny et al. 2006), speakers with masculine (e.g. lower-pitched) voices are often considered to have positive personality attributes including competence and leadership abilities. For instance, when asked to select political leaders, both men and women tend to select male and female leaders with more masculine (lower-pitched) voices and rate them as more competent than their higher-pitched counterparts (Klofstad et al. 2012; Klofstad et al. 2015). In addition, Tigue and colleagues (2012) showed voices from political candidates with artificially lowered pitch were associated with perceptions of ability and skill more often than were their higher-pitched versions, independent of whether the content spoken was political or neutral. Similarly, research on the impact of voice pitch within the business context, found that artificially lower-pitched voices of job candidates are associated with greater competence, regardless of applicant gender or résumé information (depicting either a stereotypically masculine or a stereotypically feminine applicant - Ko et al. 2009). Moreover, a lowered voice pitch from organisational spokespersons results in greater perceptions of competence and ability to restore organisational reputation compared to a raised voice pitch, particularly in times of crisis (Claeys and Cauberghe 2014).

While this research demonstrates that sex-related voice variation is sufficient to trigger stereotyping in adult listeners, an important theoretical question concerns whether auditory-based

stereotyping of adults is already present in childhood, paralleling evidence on children's gender stereotyped judgements of adults based on body shape and facial appearance (Montepare et al. 1989; Pine 2001). Our study aims to bridge this gap by directly examining how voice variation in masculinity and femininity impacts children's occupational stereotyping of adults. An investigation of this nature will provide valuable insights into the role of vocal cues in the early origins of stereotyping, paving the way for developmental investigations of stereotyping from multiple angles. Moreover, given that children's prior expectancies of other people bias their interactions with them (Harris et al. 1992; Gurland and Grolnick 2003), voice-based judgements may also have an impact on how children would engage with adults, with practical implications for understanding and improving such interactions.

Our study focuses on occupational competence, given that perceived competence is a key dimension (alongside warmth) underlying person and group perception (for a review: Fiske et al 2007). Although no research to date has directly examined how the voice impacts competence judgements of adults, recent evidence suggests that children may be sensitive to sex-related variation in voice frequency, and that this variation influences their assessment of speakers' traits in gender-stereotypical ways. For instance, children are sensitive to vocal masculinity and femininity in the voices of their peers, as they match stereotypically masculine and feminine descriptors of a child character with corresponding masculinised or feminised voices (Cartei et al. 2019a). Moreover, a recent study using a voice imitation paradigm has shown that children conform to gender-stereotyped expectations by masculinising and feminising their voices for traditionally male and female occupations (Cartei et al. 2020a). The present work aims to extend this literature by investigating for the first time whether child listeners use variation in voice masculinity and femininity (by artificially lowering/raising voice pitch) to make gender stereotypical predictions about the occupational competence of adult speakers.

We chose to focus on occupational competence in 8- to 10-year-olds as previous research has shown that from about 8 years children's range of stereotypes expands, and the nature of the gender associations becomes more abstract and multi-dimensional. For instance, they are able to use gender-related variation in behaviour and appearance in a stereotypical manner when making predictions of peers' future occupational career choices (Martin et al. 1990). Specifically, we hypothesise that children will assign higher competence to lower-pitched (more masculine) voices for stereotypically male occupations. Conversely, we expect that children will assign higher competence to higher pitched

(more feminine) voices for stereotypically female occupations. Finally, voices re-synthesised to a midline pitch should receive highest ratings when paired with gender neutral occupations.

Methods

Participants

Forty-eight children (20 females, mean age =9.46; SD =0.47, range: 8.6-10.4) took part in the study. The total sample size was based on a previous study of voice perception in child and adult listeners (Cartei et al. 2019) reporting significant effects of gender-role stereotype ratings based on variation in vocal masculinity and femininity in children's voices.

Children from UK Years 4 and 5 with no history of hearing impairments were prospectively recruited via school newsletters in two village primary schools, with informed consent by the headteachers. Parents were given a written study information sheet explaining the purpose and protocol of the study (that children would be asked to guess how good a person was at their job after listening through headphones to some men and women in specific occupations as they said a series of sentences). Parents were encouraged to ask any questions by contacting the researchers and were asked to give written consent for their child's participation. After parental consent, children were approached about the study on the day of the experiment. Researchers explained the main points of the consent/assent form verbally, adjusting the explanation to the child's age and comprehension level. Both written parental permission and child assent were obtained for child participants. Ethical approval was obtained from the University of Sussex Science and Technology Cross-Schools Research Ethics Committee (reference: ER/VC44/17).

Speaker selection

Eight adult speakers of British English (4 women, mean age =24; SD =0.32, range: 21-27) were selected from a database of 26 adults (13 women) reading out loud the following three sentences: "hello, it is nice to meet you", "thank you for your help", "no, I do not want to go" (see Appendix 1 for details on acoustic analysis). For each speaker, the three sentences were concatenated as a single voice stimulus with 50ms silence between sentences, creating 5-s "thin slices" (Ambady & Rosenthal 1992) to minimise task fatigue while eliciting listeners' judgements (see: Hughes & Harrison 2017; Tigue et

al. 2012 for examples of "thin slices" in voice research). These speakers were selected to maximise the variance in apparent vocal tract lengths (aVTL) from our original sample, which was estimated from formants 1 to 4 (aVTL is inversely correlated with the averaged distance between adjacent formants as well as absolute formant values: longer vocal tracts result in lower, more closely spaced formant frequencies, translating into a more resonant, or sonorous, voice - see Appendix 1). For males, the selected speakers had aVTLs of 15.4cm, 16.2cm, 16.7cm and 17.5cm. For women, the selected speakers had aVTLs of 14.2cm, 14.7cm, 15.0cm and 15.5cm.

Pitch re-synthesis

From each original recording, we used the PSOLA algorithm in PRAAT 6.0.28 (*change gender* command) to create three stimuli varying in pitch without altering other aspects of the sound. In one stimulus mean F0 was altered to fit the mean F0s for the men and women in our original speaker database (mid F0), while in the other two stimuli F0 was manipulated to be, respectively, 1 standard deviation (SD) lower (lowered F0) or higher (raised F0) than the mean values for men (mid F0: 115.2+/-12.8Hz) and women (mid F0: 204.4Hz +/- 29.4Hz) in our sample, following a similar procedure to Reby and colleagues (2016). Thus, the resulting F0 values for each of the selected male speakers were: 102.4Hz, 115.2Hz, 128.0Hz and for female speakers: 175.0Hz, 204.4Hz, 233.8Hz. To confirm the perceived naturalness of the voice stimuli, we asked 10 listeners (5 men, 5 women) to rate the speakers' voices from the database and the 24 resynthesised versions (3 x 8 speakers) on a 7-point scale (1=very unnatural, 2=unnatural, 3=somewhat unnatural,4=neither, 5=somewhat natural, 6=natural, 7=very natural). One-way ANOVAs were separately run on the ratings of male and female speakers, treating the ratings from 1 to 7 as continuous. The within-subjects factor was stimulus type (four levels: original, raised, lowered and mid resynthesised variants). Listeners' average scores for the original and resynthesised stimuli were above 6 "natural" and there was no significant difference between unmanipulated and resynthesised voices, female: $F(3,24)= .663$, $p>.05$, male: $F(3,24)=.277$, $p>.05$.

Procedure

Children sat individually in a quiet room at their school with the researcher. Voice stimuli were played back one at the time from a laptop through high-quality child-safe headphones (PURO

Labs BT2200). For each voice, the experimenter read out loud the speaker's occupation, followed by a brief description of the occupation. Next, children listened to the speaker's voice and were asked to rate how good or bad they thought that person (children were told whether it was a man or a woman) was at their job on the basis of their voice. Children marked their answer by putting a cross on a paper-based, picture-aided Likert-scale (1=very bad, 2=bad, 3=not bad nor good, 4=good, 5=very good, with corresponding smiley faces ranging from "unhappy" to "happy" (see Appendix 2)). We selected nine occupations, three stereotypically female (Babysitter, Beautician, Nurse) three gender-neutral (Doctor, Student, Writer), and three stereotypically male (Builder, Lorry driver, Mechanic). Our choice of occupations for each of the three categories was guided by the Office of National Statistics (2019) and by findings from a questionnaire with UK children aged 6-10 on perceived occupational gender ratio and competence (Cartei et al. 2020a).

Each child rated all the voice stimuli in two successive blocks, one with all 12 male voice stimuli from the 4 male speakers, and one with all 12 female voice stimuli from the 4 female speakers (8 speakers X 3 pitch conditions X 1 out of 9 occupations randomised within each child, and counter-balanced between children). Children were told the speakers' sex for each stimulus, and the order in which the blocks were presented was alternated between participants to control for order effects. Before each block, children practised the task twice by listening to a man's and woman's voice from the original database of 26 speakers, but not from the 8 selected speakers. This pre-test allowed the experimenter to make sure the child understood the task, as well as to adjust the playback volume to a comfortable level.

Statistical analyses and results

To investigate the effects of occupation type and F0 variant on children's ratings of men and women speakers, we ran two Linear Mixed Models (LMM) separately for the male and female speakers, with occupation type (male-typed, female-typed, gender-neutral) F0 variant (lowered F0, mid F0, raised F0), listener sex and their 2-way interactions as fixed factors. Apparent Vocal Tract Length (aVTL) and occupation (nested within occupation type) were random factors. Both LMMs also included listener identity as a random factor, with a separate intercept for each listener (Table 1). Pairwise comparisons (Bonferroni corrected) were used to detect significant differences between group

means for significant main and interaction effects. Standard estimates of effect sizes (Cohen's ds) are reported, with values of 0.2, 0.5, and 0.8 represent small, medium, and large effects (Cohen 1988).

Occupational competence ratings of women speakers

There was a significant main effect of occupation type on ratings of women speakers: across F0 variants, women were slightly, but significantly, rated as more competent for the gender-neutral occupations than the female ($d=.28$, $p<.05$) or male occupations ($d=.59$, $p<.05$). Women were rated significantly more competent for the stereotypically female occupations than the male occupations, $d=.31$, $p=.025$ (see Fig.1a).

Fig.1 effect of occupation type on children's mean competence ratings of (a) women and (b) men speakers

There was also a significant interaction effect between occupation type and F0 variant (Fig.2). When paired with the stereotypically female occupations, women's raised pitch voices received the highest competence ratings ($M=3.9$, $SE=.15$), compared to the mid pitch voices ($M=3.4$, $SE=.15$), $d=.65$, $p<.05$, and lower pitch voices, $d=1.1$, $p<.05$. Women's lower pitch voices also received lower ratings ($M=2.9$, $SE=.15$) than mid pitch voices, $d=.39$, $p<.05$. For the stereotypically male occupations, women's raised pitch voices received the lowest ratings ($M=2.6$ $SE=.16$) compared to the mid pitch voices ($M=3.4$, $SE=.15$), $d=.80$, $p<.05$, and lower pitch voices, ($M=3.3$, $SE=.15$) $d=.61$, $p<.05$. However, women's lowered pitch voices did not receive higher ratings than mid pitch voices, $p>.05$. No significant difference in ratings was found amongst women's F0 variants in the gender-neutral occupations, $p>.05$.

Occupational competence ratings of men speakers

There was a significant main effect of occupation type on ratings of men speakers: pairwise comparisons revealed that, across F0 variants, men were rated less competent for the female occupations than for the gender-neutral, $d=.48$,$p<.05$, and male occupations, $d=.61$, $p<.05$. Mean ratings were highest for the male occupations compared to the gender-neutral occupations, though not significantly so, $p>.05$ (see also Fig.1b).

There was a significant interaction of occupation type and F0 variant (Fig.3). When paired with the stereotypically female occupations, children rated men's lowered pitch voices as significantly less competent (M= 2.2, SE=.15) than mid F0 (M=2.9, SE=.15), d=.70,p<.05, or raised pitch versions (M=3.6, SE=.14), d=1.3, p<.05, while the latter received higher competence ratings than mid pitch voices d=.77, p<.05. For the stereotypically male occupations, children rated men's lowered pitch voices as significantly more competent (M=4.2, SE=.14) than the mid pitch (M=3.4, SE=.14), d=.80, p<.05, and raised pitch (M=3.2, SE =.15) versions, d=.97, p<.05. For the gender-neutral occupations, no significant differences were found amongst F0 variants, all ps>.05.

Fig.2 Occupation type (female, neutral, male) by F0 variant (raised (yellow), mid (green), lowered(blue)) for women speakers.

Fig.3 Occupation type (female, neutral, male) by F0 variant (raised (yellow), mid (green), lowered(blue)) for male speakers.

Discussion

This is the first study to show that children make gender-stereotypical judgements of adult speakers on the basis of speaker's variation in vocal masculinity and femininity, complementing prior research that focused exclusively on adults. Specifically, in line with our predictions, we found that feminised voices received the highest ratings when paired with stereotypically female occupations, and the lowest ratings when paired with stereotypically male occupations. Also consistent with our predictions, masculinised voices received the lowest ratings when paired with stereotypically female occupations, and male (but not female) masculinised voices received the highest ratings when paired with stereotypically male occupations. Overall, our results show that variation in adults' vocal masculinity and femininity (manipulated by artificially lowering or raising mean voice pitch) affects children's ratings of speakers' occupational competence in gender-stereotypical ways, though ratings for stereotypically male occupations were also influenced by speakers' sex.

In terms of the overall pattern of results, the observed ratings are largely consistent with psychoacoustic studies with adult listeners, showing that (re-synthesised and natural) male voices with lower pitch are preferentially attributed stereotypically male characteristics, such as masculinity

(Pisanski et al. 2012), physical and social dominance (Vukovic et al. 2011; Puts et al. 2007; Hall et al. 2005), authority (Sorokowski et al. 2019), and leadership (Klofstad et al. 2012; Tigue et al. 2012), though perceivers associated higher pitch more strongly with high- than with low-rank behaviours in at least one study (Ko et al. 2015). On the other hand, women with higher-pitched voices are known to be preferentially attributed stereotypically female characteristics, such as femininity (Röder et al. 2013), friendliness (Tsuji 2004; Ohara 1999), and submissiveness (Borkowska and Pawlowski 2011).

Although, as expected, our results show that feminised voices from speakers of both sexes received the highest competence ratings for stereotypically female jobs, psychoacoustic studies report that adult listeners rate lower-pitched individuals as more competent than higher-pitched individuals both from speakers' recordings that are neutral (ratings of speakers reading out loud vowels and sentences of gender-neutral content: Krahé et al. 2020; Oleszkiewicz et al. 2016) or politically relevant (e.g. ratings of hypothetical political candidates: Klofstad et al. 2012). However, none of these studies asked listeners to make judgements in the context of female-typed occupations, whereas our study did. Because professions that are dominated by women tend to be stereotyped as more feminine, and requiring more "female-like" traits (e.g. warmth: Eagly and Carli 2003; friendliness: Wharton 1999; helpfulness and cooperation: Cejka and Eagly 1999) competence on these jobs is likely to be judged on these traits, and thus may drive the higher competence ratings for the higher-pitched voices observed in the present study. While the present study did not directly assess whether high-pitched voices triggered these types of inferences, in partial support of this hypothesis, Oleszkiewicz and colleagues (2016) report that adult listeners make positive associations between high pitch and warmth in women's voices (though not in men's). Also, Halper and Stopeck (2019) report that perceptions of warmth primarily drive the relationship between job candidate gender and both likeability and job hireability for female-dominated domains such as the caregiving professions.

Both speakers' biological characteristics and listeners' socialisation processes may contribute to the observed overall pattern of results. Lower-pitched male voices positively correlate with salivary testosterone levels in childhood and adulthood (Cartei et al. 2014; Cartei et al. 2020b), and testosterone is a primary driver of physiological masculine features, such as increased muscle size and strength (Bhasin et al. 2001), and physical fitness (Fink et al. 2006; Manning and Taylor 2001), which are valued traits in physically demanding jobs that are male-dominated (Colker 1985). As well as negatively correlating with testosterone, higher-pitched voices in men are preferred by women seeking

291 greater perceived parental and relationship investment (Apicella and Feinberg 2009). Moreover,
292 higher-pitched voices in women positively correlate with level of oestrogen, which is positively linked
293 to maternal behaviour in numerous species, including rats, mice, sheep, and possibly non-human
294 primates (Bridges 2015). Thus, a high voice pitch may advertise greater actual or perceived propensity
295 for nurturing and care-taking roles, which are stereotypically seen as women's jobs (Guy and Newman
296 2004). While the observed ratings may partially reflect children's sensitivity to voice cues underlying
297 qualities of speakers, many such attributions are nowadays irrelevant to job competence. For instance,
298 there is considerable overlap in men's and women's physical strength, and many heavy manual jobs are
299 now machine-operated, which means that many women are physically capable of doing such work
300 (Ness 2012).

301 Moreover, the idea that voice pitch is a reliable cue to biosocial dimensions fails to account
302 for the fact that children and adults typically develop stereotypic views and prejudices concerning
303 groups that are unjustified (and thus uncorrelated with any observable traits or behaviours, e.g.
304 Zebrowitz 1996; Bereczkei and Mesko 2006; Bigler and Liben 2007). Specifically, socialisation
305 research has shown that, consistent with the general principle of correspondence bias (Gilbert and
306 Malone 1995), individuals tend to ascribe gender-stereotypic attributes to job holders that are in line
307 with occupational sex ratios, even if those attributes are irrelevant to those jobs (Cejka and Eagly
308 1999). Given that sex-segregation is still a predominant feature of many jobs (Office of National
309 Statistics 2019), the observed ratings could emerge from children's observations of the vocal
310 characteristics of the sex that is numerically dominant in the occupation (males' voices being, on
311 average, lower-pitched than females'), even if those correspondences are irrelevant to competence.

312 An additional possibility for children's higher ratings of feminised voices in female-typed
313 roles is based on children's prior experience. From infancy, children learn to associate higher pitch
314 voices with relational and affective skills, which are important in many stereotypically female
315 occupations, including the ones in the present study (Guy and Newman 2004). Indeed, raised pitch
316 appears to communicate caregivers' affect and intentions non-verbally, and caregivers routinely
317 increase their pitch when speaking to children as opposed to adults (Broesch and Bryant 2015; Grieser
318 & Kuhl 1988). For instance, when mothers speak with a heightened pitch (and expanded melodic
319 contours) they are more able to elicit and maintain infant attention, independent of what they are saying

(Papoušek et al. 1990). High-pitch is also common in caregivers' speech when conveying emotional information to children compared to speaking to adults (Kitamura & Burnham 2003).

Contrary to our hypothesis, we also found that women's masculinised voices were not rated as more competent than the mid F0 variant for the masculine occupations. Specifically, to the extent that F0 cues for physiological masculinity in women (e.g. decreased oestrogen, lower fertility: Prelevic 2013; Bryant and Haselton 2009, but not testosterone: Dabbs and Mallinger 1999), more masculine female voices were expected to be rated as more competent in male jobs, but this is not what we observed. An alternative explanation for our findings, is that children's competence ratings of low-pitched women's voices resulted from a (conscious or unconscious) compromise between perceived masculinity and overall preference for high-pitched voices in females. Previous research with adult listeners indicates that, while low-pitched voices in both men and women are perceived as more masculine (Krahé and Papakonstantinou 2020), and are preferred over high-pitched voices in male speakers, they are not preferred over high-pitched voices in female speakers (Tsantani et al. 2006). In fact, women speaking with lower-pitched voices are rated as less vocally attractive (Feinberg et al. 2008) and as having fewer favourable personality traits than higher-pitched women (e.g., Scherer 1974, 1978). Lending support to this argument, a recent study looking at job hiring preferences (Phelan et al. 2008) found that fictitious female job applicants with masculine traits were judged by adult raters as more competent, but lacking in social skills compared to applicants with feminine traits, while no such bias was found in male applicants.

Although variation in voice pitch within the two sexes influenced children's ratings stereotypically, children rated men as significantly more competent than women in male jobs and less competent than women in female jobs, regardless of our pitch manipulations. These results suggest that speaker gender may be a stronger contributor to stereotyping than vocal variation in masculinity and femininity. It is also possible that this effect was heightened by our paradigm, given that children knew in advance the sex of the speaker and rated all speakers of the same sex in one block. Indeed, hiring bias research demonstrates that when occupational assessors are told the sex of hypothetical job candidates, stereotype-congruent associations (e.g. female/male applicants being considered for a stereotypically female/male jobs), are given more favourable evaluations than when stereotype incongruent associations are primed (e.g. female/male applicants being considered for stereotypically male/female jobs), even when applicants are equally qualified (Rice and Barth 2015).

In summary, our study shows that children use within-sex variation in vocal masculinity and femininity when making gender-stereotypical judgements of adults, as previously found in judgements of other children (Cartei et al. 2019a). Our findings also complement those of a recent voice imitation study, which showed that children link vocal masculinity/femininity to stereotypically male/female occupations (Cartei et al. 2020a), by showing that gender-linked variation influences beliefs about competence. Together these observations highlight the fact that the voice is an important aspect of children's gender stereotyping and indicate that it can be easily used as a versatile, implicit measure of children's gender stereotyping, through voice perception or production tasks.

To further trace the developmental trajectory of children's occupational stereotyping (stereotype flexibility and stereotype knowledge), the present paradigm could be used with a wider range of occupations and ratings of relevant traits other than competence (e.g. dominance, friendliness). It could also be extended to younger children and adolescents to assess the degree to which voice stereotypes correlate with a child's classification skills, knowledge about job requirements, and gender stereotype flexibility, all of which develop with age (Liben et al. 2002). Moreover, cross-cultural comparisons with our study should establish the extent to which our findings can be generalised to diverse cultural contexts, outside that of Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies (Henrich et al. 2010). Our paradigm could also be used in conjunction with inter-individual measures, to investigate how individual differences in children's occupational stereotyping may emerge. For instance, differences in exposure to division of labour in the family (Serbin et al. 1993; Fulcher et al. 2008), and on television (O'Bryant et al. 1978), both affect children's occupational stereotyping. It would be interesting to know if and how the patterns observed in the present work would be subject to this kind of environmental influence.

Finally, given that children use gender-related voice variation to make judgements about adults in occupations, an important next step would be to explore the relative contributions of these judgements to child-adult interpersonal processes. Specifically, future studies could explore whether voice masculinity and femininity do affect children's interactions with men and women in these roles, by using confederates and recording children's behavioural responses during and after the interactions (e.g. asking children if they felt more comfortable to be treated by a nurse having a feminine rather than masculine voice).

References

Apicella, C. L., & Feinberg, D. R. (2009). Voice pitch alters mate-choice-relevant perception in hunter–gatherers. *Proceedings of the Royal Society B: Biological Sciences*, 276(1659), 1077-1082.

Bereczkei, T., & Mesko, N. (2006). Hair length, facial attractiveness, personality attribution: A multiple fitness model of hairdressing. *Review of psychology*, 13(1), 35-42.

Bhasin, S., Storer, T. W., Berman, N., Callegari, C., Clevenger, B., Phillips, J., ... & Casaburi, R. (1996). The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. *New England Journal of Medicine*, 335(1), 1-7.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, 16(3), 162-166.

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82(1), 55-59.

Bridges, R. S. (2015). Neuroendocrine regulation of maternal behavior. *Frontiers in neuroendocrinology*, 36, 178-196.

Broesch, T. L., & Bryant, G. A. (2015). Prosody in infant-directed speech is similar across western and traditional cultures. *Journal of Cognition and Development*, *16*(1), 31-43.

Bryant, G. A., & Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biology Letters*, 5(1), 12-15.

Cartei, V., Banerjee, R., Hardouin, L., & Reby, D. (2019a). The role of sex-related voice variation in children's gender-role stereotype attributions. *British Journal of Developmental Psychology*, 37(3), 396-409.

Cartei, V., Oakhill, J., Garnham, A., Banerjee, R., & Reby, D. (2020a). “This is what a mechanic sounds like.” Children’s vocal control reveals implicit occupational stereotypes. Psychological Science.

Cartei, V., Banerjee, R., Garnham, A., Oakhill, J., Roberts, L., Anns, S., ... & Reby, D. (2020b). Physiological and perceptual correlates of masculinity in children's voices. *Hormones and behavior*, 117, 104616.

Cartei, V., Bond, R., & Reby, D. (2014). What makes a voice masculine: Physiological and acoustical correlates of women's ratings of men's vocal masculinity. *Hormones and Behavior*, 66(4), 569-576.

Cartei, V., Garnham, A., Oakhill, J., Banerjee, R., Roberts, L., & Reby, D. (2019b). Children can control the expression of masculinity and femininity through the voice. *Royal Society open science*, 6(7), 190656.

Cejka, M. A., & Eagly, A. H. (1999). Gender-Stereotypic Images of Occupations Correspond to the Sex Segregation of Employment. *Personality and Social Psychology Bulletin*, 25(4), 413–423.

Claeys, A. S., & Cauberghe, V. (2014). What makes crisis response strategies work? The impact of crisis involvement and message framing. *Journal of Business Research*, 67(2), 182-189.

Colker, R. (1985). Rank-order physical abilities selection devices for traditionally male occupations as gender-based employment discrimination. *UC Davis Literature Review*, 19, 761.

Dabbs Jr, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitchamong men. *Personality and individual differences*, 27(4), 801-804.

Eagly, A. H., & Carli, L. L. (2003). The female leadership advantage: An evaluation of the evidence. *Leadership Quarterly*, 14(6), 807-834.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The Role of Femininity and Averageness of Voice Pitch in Aesthetic Judgments of Women's Voices. *Perception*, 37(4), 615–623.

Fink, B., Thanzami, V., Seydel, H., & Manning, J. T. (2006). Digit ratio and hand-grip strength in German and Mizos men: Cross-cultural evidence for an organizing effect of prenatal testosterone on strength. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, 18(6), 776-782.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 11(2), 77-83.

Fulcher, M., Sutfin, E. L., & Patterson, C. J. (2008). Individual differences in gender development: Associations with parental sexual orientation, attitudes, and division of labor. *Sex Roles*, *58*(5-6), 330-341.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental psychology*, 24(1), 14.

Hall, J. A., Coats, E. J., & LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychological bulletin*, 131(6), 898.

Halper, L. R., Cowgill, C. M., & Rios, K. (2019). Gender bias in caregiving professions: The role of perceived warmth. *Journal of Applied Social Psychology*, 49(9), 549-562.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences*, 33(2-3), 61-83.

Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice*, 19(4), 544-554.

Gurland, S. T., & Grolnick, W. S. (2003). Children's expectancies and perceptions of adults: Effects on rapport. *Child development*, 74(4), 1212-1224.

Guy, M. E., & Newman, M. A. (2004). Women's jobs, men's jobs: Sex segregation and emotional labor. *Public administration review*, 64(3), 289-298.

Harris, M. J., Milich, R., Corbitt, E. M., Hoover, D. W., & Brady, M. (1992). Self-fulfilling effects of stigmatizing information on children's social interactions. *Journal of Personality and Social Psychology*, 63(1), 41.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, 4(1), 85-110.

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, 279(1738), 2698-2704.

Klofstad, C. A., Anderson, R. C., & Nowicki, S. (2015). Perceptions of competence, strength, and age influence voters to select leaders with lower-pitched voices. *PloS one*, 10(8), e0133779.

Ko, S. J., Judd, C. M., & Stapel, D. A. (2009). Stereotyping based on voice in the presence of individuating information: Vocal femininity affects perceived competence but not warmth. *Personality and Social Psychology Bulletin*, 35(2), 198-211.

Ko, S. J., Sadler, M. S., & Galinsky, A. D. (2015). The sound of power: Conveying and detecting hierarchical rank through voice. *Psychological Science*, 26(1), 3-14.

Krahé, B., & Papakonstantinou, L. (2020). Speaking like a Man: Women’s Pitch as a Cue for Gender Stereotyping. *Sex Roles*, 82(1-2), 94-101.

Liben, L. S., Bigler, R. S., Ruble, D. N., Martin, C. L., & Powlishta, K. K. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the society for research in child development*, i-183.

Little, A. C. (2014). Facial appearance and leader choice in different contexts: Evidence for task contingent selection based on implicit and learned face-behaviour/face-ability associations. *The Leadership Quarterly*, 25(5), 865-874.

Manning, J. T., & Taylor, R. P. (2001). Second to fourth digit ratio and male ability in sport: implications for sexual selection in humans. *Evolution and Human Behavior*, 22(1), 61-69.

Martin, C. L., Wood, C. H., & Little, J. K. (1990). The development of gender stereotype components. *Child development*, 61(6), 1891-1904.

Montepare, J. M., & Zebrowitz-McArthur, L. (1989). Children's perceptions of babyfaced adults. *Perceptual and Motor Skills*, 69(2), 467-472.

Ness, K. (2012). Constructing masculinity in the building trades:'Most jobs in the construction industry can be done by women'. *Gender, Work & Organization*, 19(6), 654-676.

O'Bryant, S. L., & Corder-Bolz, C. R. (1978). The effects of television on children's stereotyping of women's work roles. *Journal of Vocational Behavior*, 12(2), 233-244.

Office of National Statistics (2019). Annual Population Survey - Employment by occupation by sex. Retrieved from: https://www.nomisweb.co.uk/datasets/aps168/reports/employment-by-occupation?compare=K02000001

Oleszkiewicz, A., Pisanski, K., Lachowicz-Tabaczek, K., & Sorokowska, A. (2016). Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. *Psychonomic Bulletin & Review*, 24, 856–862.

Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, 13(4), 539-545.

Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting hiring criteria reflects backlash toward agentic women. *Psychology of Women Quarterly*, 32, 406–413.

Pine, K. J. (2001). Children's perceptions of body shape: A thinness bias in pre-adolescent girls and associations with femininity. *Clinical Child Psychology and Psychiatry*, 6(4), 519-536.

Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior*, 33(5), 509-519.

Pisanski, K., & Bryant, G. A. (2019). The evolution of voice perception. In N. S. Eidsheim & K. L. Meizel (Eds.), *The oxford handbook of voice studies* (pp. 269–300). New York, NY: Oxford University Press.

Prelevic, G. M. (2013). The effects of sex hormones on the female voice. *Physical and Emotional Hazards of a Performing Career: A special issue of the journal Musical Performance*, 2, 93.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, 28(5), 340-344.

Re, D. E., & Rule, N. O. (2017). Distinctive facial cues predict leadership rank and selection. *Personality and Social Psychology Bulletin*, 43, 1311-1322.

Reby, D., Levréro, F., Gustafsson, E., & Mathevon, N. (2016). Sex stereotypes influence adults' perception of babies' cries. *BMC psychology*, 4(1), 19.

Rice, L., & Barth, J. M. (2016). Hiring decisions: The effect of evaluator gender and gender stereotype characteristics on the evaluation of job applicants. *Gender Issues*, 33(1), 1-21.

Röder, S., Fink, B., Feinberg, D. R., & Neave, N. (2013). Facial visualizations of women's voices suggest a cross-modality preference for femininity. *Evolutionary Psychology*, 11(1), 147470491301100119.

Rule, N. O., & Ambady, N. (2009). She's got the look: Inferences from female chief executive officers' faces predict their success. *Sex Roles*, 61(9-10), 644-652.

Scherer, K. R. (1974). Voice quality analysis of American and German speakers. *Journal of Psycholinguistic Research*, *3*(3), 281-298.

Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. European *Journal of Social Psychology*, 8(4), 467-487.

Sczesny, S., Spreemann, S., & Stahlberg, D. (2006). Masculine= competent? Physical appearance and sex as sources of gender-stereotypic attributions. *Swiss Journal of Psychology*, 65(1), 15-23.

Serbin, L. A., Powlishta, K. K., Gulko, J., Martin, C. L., & Lockheed, M. E. (1993). The development of sex typing in middle childhood. *Monographs of the society for research in child development*, i-95.

Sorokowski, P., Puts, D., Johnson, J., Żółkiewicz, O., Oleszkiewicz, A., Sorokowska, A., ... & Pisanski, K. (2019). Voice of authority: professionals lower their vocal frequencies when giving expert advice. *Journal of Nonverbal Behavior*, 43(2), 257-269.

Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, 33(3), 210-216.

Titze, I. R. (1989), Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of* America, 85, 1699–1707.

Todorov A., Mandisodza A.N., Goren A., Hall C.C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308, 1623–1626.

Tsantani, M. S., Belin, P., Paterson, H. M., & McAleer, P. (2016). Low Vocal Pitch Preference Drives First Impressions Irrespective of Context in Male Voices but Not in Female Voices. *Perception*, 45(8), 946–963.

Tsuji, A. (2004). The case study of high pitch register in English and in Japanese: Does high pitch register relate to politeness. *Seijo English Monographs*, 37, 227-260.

Vukovic, J., Jones, B. C., Feinberg, D. R., DeBruine, L. M., Smith, F. G., Welling, L. L., & Little, A. C. (2011). Variation in perceptions of physical dominance and trustworthiness predicts individual differences in the effect of relationship context on women's preferences for masculine pitch in men's voices. *British Journal of Psychology*, 102(1), 37-48.

Wharton, A. S. (1999). The psychosocial consequences of emotional labor. *The Annals of the American Academy of Political and Social Science*, 561(1), 158-176.

World Health Organisation (2020). *Gender*. Retrieved from: https://www.who.int/health-topics/gender

Zebrowitz, L. A. (1996). Phsicial appearance as a basis of Stereotyping. In Macrae, C. N., Stangor, C., & Hewstone, M. (Eds.). *Stereotypes and stereotyping* (pp. 79-119). New York: Guilford Press.

546 Appendix 1

547 The original database included 26 adult speakers (aged 18 to 35, 13 women). Each speaker was
548 recorded while reading out the sentences: “hello, it is nice to meet you”, “thank you for your help”,
549 “no, I do not want to go”. These sentences were chosen because they were gender-neutral in content,
550 familiar, relatively short and grammatically simple for adults to say and for children to understand, and
551 because they included the main vowels of British English. For each speaker, the three sentences were
552 scaled at 60dB and then concatenated in the order presented above, with 50ms silence in between.
553 Acoustic measurements for each speaker were taken from the entire sequence using PRAAT software
554 (version 4.6, Boersma 2001). Pitch values were obtained using PRAAT’s pitch-tracking function with a
555 range setting of 75-300Hz for males and 100-500Hz for females (Boersma 2009). Formant values were
556 obtained using PRAAT’S formant-tracking function, setting *maximum formant* to 5000Hz for males
557 and 5500Hz for females, and *number of formants* to 5. Averaged across male speakers, mean F_0 was
558 115.2 Hz (SD = 12.8Hz, range: 98-144 Hz) and mean ΔF was 1061.4 Hz (SD =51.9Hz, range: 999-
559 1138Hz), corresponding to an apparent Vocal Tract Length of 16.42cm (SD = 0.8cm, range: 15.4-
560 17.5cm). Averaged across female speakers, mean F0 was 204.4 Hz (SD =29.4Hz range: 171-274 Hz),
561 and mean ΔF was 1192.1Hz (SD =30.3Hz, range: 1229-1131Hz), corresponding to an apparent Vocal
562 Tract Length of 14.46cm (SD = 0.4cm, range: 14.2-15.5cm). These values are in line with those of
563 previous samples of speakers of British and American English (e.g. Bachorowski and Owren 1999;
564 Rendall et al. 2005; Cartei et al. 2012). Mean formant spacing (ΔF) and apparent Vocal Tract Length
565 (aVTL), its inverse acoustic correlate measured in cm, were computed from the mean centre
566 frequencies of F1-F4, using the method described by Reby and McComb (2003). Given that the vocal
567 tract can be approximated to a straight uniform tube that is closed at one end and open at the other
568 (Titze 1994), formant frequencies can be calculated as: (1) $Fi = \frac{(2i-1)c}{4aVTL}$
569 Where *i* is the formant number, c is the speed of sound in a mammal vocal tract (350m/s), aVTL is the
570 apparent vocal tract length and Fi is the frequency of ith formant. Formant spacing can be defined as
571 the spacing between any two successive formants, $\Delta F = F_{i+1}-F_i$. Thus from (1), it follows that:
572 (2) $\Delta F = F_{i+1}-F_i = \frac{c}{2aVTL}$
573 And thus Fi can also be expressed as: (3) $Fi = \frac{(2i-1)}{2}\ \Delta F$
574 We can therefore estimate ΔF by seeking the best fit for equation (3).

575
576

Appendix 2

577

Today, I will ask you to listen through these headphones to some people as they say a series of sentences: "Hello, it is nice to meet you", "Thank you for your help", "No, I do not want to go".

For each person, I will tell you what their job is. I will then ask you to guess how good that person is at their job. If you are unsure what the job is, just ask me and I will tell you.

Let's practice together. (*research to test headphones first, adjust volume and play sound*).

If it's a bit too quiet, or it's a bit too loud, please let me know and we can adjust the volume of the sound.

This man is a **poet**.

How good or bad do you think he is at his job?

1	Poet *Writes poems*		 **1** **Very bad**	 **2** **Bad**	 **3** **Not Bad nor Good**	 **4** **Good**	 **5** **Very good**

578

579

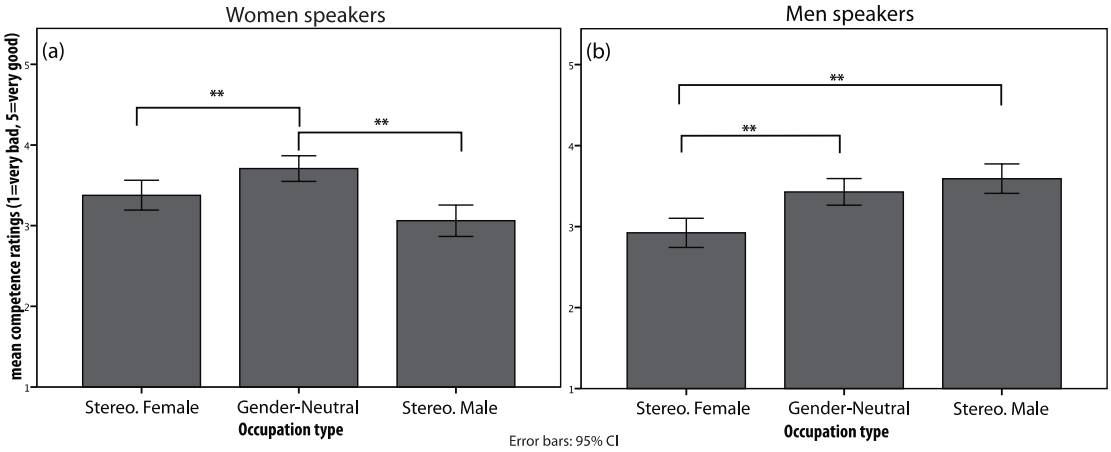

Fig 1. effect of occupation type on children's mean competence ratings of (a) women and (b) men speakers

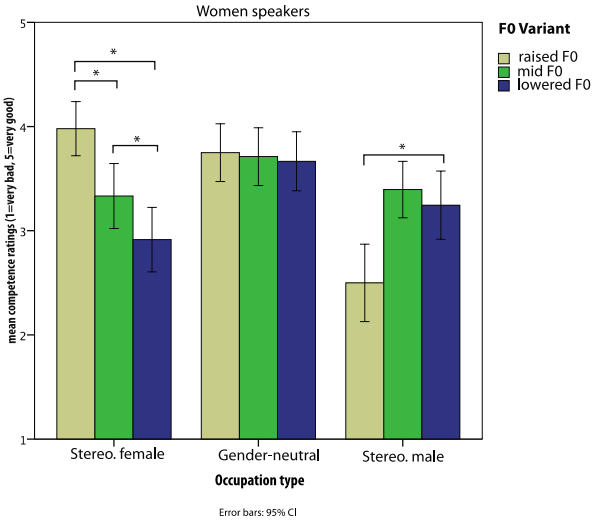

Fig.2 Occupation type (female, neutral, male) by F0 variant (raised (yellow), mid (green), lowered(blue)) for women speakers.

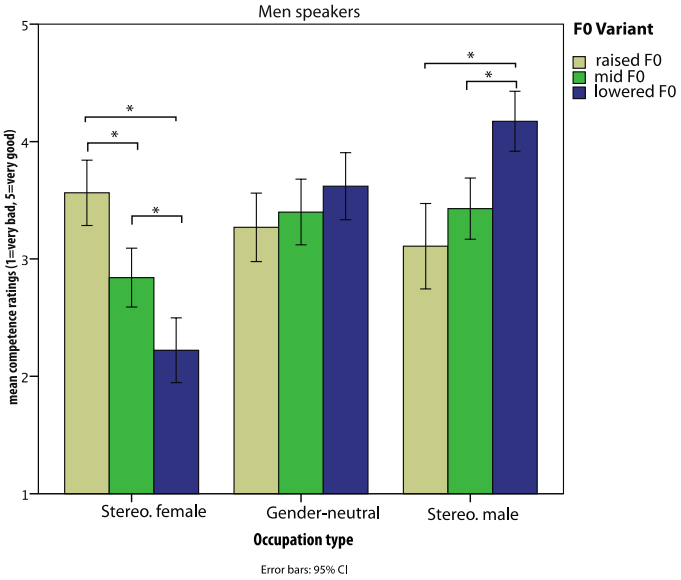

Fig.3 Occupation type (female, neutral, male) by F0 variant (raised (yellow), mid (green), lowered(blue)) for male speakers.

Table 1. Effect of occupation type and speaker resynthesised F0 on occupational competence ratings.

SpeakerSex	Source	df1	df2	F	p
Women	Intercept	1	46.851	3229.096	<.001**
	F0Variant	2	299.408	1.925	0.148
	Occupation type	2	375.422	13.139	<.001**
	Listener sex	1	46.709	3.141	0.083
	F0Variant * Occupation type	4	398.116	10.932	<.001**
	F0Variant * Listener sex	2	298.001	0.943	0.391
	Occupation type * Listener sex	2	374.522	1.814	0.164
Men	Intercept	1	46.996	3404.147	<.001**
	F0Variant	2	294.563	1.002	0.369
	Occupation Type	2	374.048	21.801	<.001**
	Listener sex	1	46.787	2.818	0.1
	F0Variant * Occupation type	4	404.458	19.769	<.001**
	F0Variant * Listener sex	2	295.105	2.647	0.073
	Occupation type * Listener sex	2	373.506	2.744	0.066

Linear mixed models (LMM) testing the (main and interaction) effects of occupation type (stereotypically male, gender-neutral, stereotypically female), listener sex, and speaker resynthesised F0 (lowered F0, mid F0, raised F0) on child listeners' ratings of occupational competence in voices presented as belonging to men and women (both rated along a 5 point Likert scale).