The development of gendered speech in children: Insights from adult L1 and L2 perceptions

Priscilla Fung,[a] Jessamyn Schertz,[b] and Elizabeth K. Johnson

Department of Psychology, University of Toronto, Mississauga, ON L5L 1C6, Canada

priscilla.fung@mail.utoronto.ca,

elizabeth.johnson@utoronto.ca

Department of Language Studies, University of Toronto, Mississauga, ON L5L 1C6, Canada

jessamyn.schertz@utoronto.ca

Abstract: Past studies have shown that boys and girls sound distinct by 4 years old, long before sexual dimorphisms in vocal anatomy develop. These gender differences are thought to be learned within a particular speech community. However, no study has asked whether listeners' sensitivity to gender in child speech is modulated by language experience. This study shows that gendered speech emerges at 2.5 years old, and that L1 listeners outperform L2 listeners in detecting these differences. The findings highlight the role of language-specific sociolinguistic factors in both speech perception and production, and show that gendered speech emerges earlier than previously suggested.

Keywords: speech perception, children's voices, gender identification, language experience, acoustic information

[a] Author to whom correspondence should be addressed.

[b] Also at: Department of Linguistics, University of Toronto, Toronto, ON M5S 3G3, Canada.

1. Introduction

The voices of adult male and female humans are typically very easy to distinguish. For example, the average f0 for adult males is about half that of the average f0 for adult females (e.g., Baken & Orlikoff, 2000). This f0 difference, along with other salient acoustic differences such as lower spectral frequencies in males than females, is directly linked to anatomical traits: males tend to have long vocal tracts and more slowly vibrating vocal cords than women. But interestingly, past studies have shown that listeners can classify voices of children as young as 4-year-olds as belonging to a boy or girl – long before anatomical differences between the sexes emerge, and that these decisions are at least in part related to acoustic parameters such as f0 and formant frequencies (e.g., Perry et al., 2001). The most plausible explanation for why boys and girls sound distinct before adolescence is their early adoption of community-specific gender differences in the way children talk. Indeed, sociolinguistic studies have shown that sensitivity to gender differences in speaking style is early emerging (e.g., Ladegaard & Bless, 2003). In the current study, we use a set of longitudinal recordings to examine 1) when perceptible differences between boys' and girls' productions first emerge, and 2) whether listeners' sensitivity to the cues distinguishing boys and girls is affected by language experience.

Because gender differences in speech are at least in part learned from the speech community (i.e., not only due to anatomical differences; Johnson, 2005), the cues differentiating prototypical male versus female speech can differ across dialects and languages. For example, f0 differences between males and females have been shown to be more pronounced in Japanese speakers than in Dutch speakers (van Bezooijen, 1995). And Korean-English bilinguals change their f0 range to match the speech patterns of the language they are speaking, with males using a wider f0 range when speaking Korean, but a narrower range when speaking English (Cheng, 2020). Beyond f0, language-specific gender differences have been observed in the acoustic-phonetic realization of speech segments. For example, gender differences in the production of /s/ have been reported to be more pronounced in Japanese than in English (Heffernan, 2004). And gender-based differences in VOT have also been argued to vary by language. For example, female Seoul Korean speakers reportedly produce shorter VOTs for aspirated stops than males (Oh, 2011), whereas female English speakers reportedly produce longer VOTs than males (Robb et al., 2005; see, however, Morris et al., 2008).

Taken together, the examples given above demonstrate that gendered speech patterns are not solely physiological in nature, but are also learned through socialization within a particular community. Interestingly, evidence from the developmental literature has shown that children adopt these gender differences in speaking style early in life. In a cross-sectional study, native English adult listeners who were presented with recordings from 4-, 8-, 12-, and 16-year-olds, performed above chance at identifying gender in even the youngest of these children (Perry et al., 2001), despite the 4- and 8-year-olds being too young to exhibit sex differences in vocal tract anatomy (Tecumseh Fitch & Giedd, 1999). And some studies have reported gender differences in formant frequency values before puberty, arguably because boys attempt to produce a more masculine voice and speech pattern by lowering their jaw and modifying the extent of lip rounding (e.g., Lindblom & Sundberg, 1971; Bennett & Weinberg, 1979; Bennett, 1981). Other studies have suggested that gender differences in the realization of /s/ are present in children as young as 4 years of age (Li et al., 2016). And Yang and Mu (1989) reported acoustic differences in the speech of three years old, the youngest reported differences we are aware of. To summarize, perception studies suggest that boys and girls produce perceptible differences in their speech by the age of four, and production studies suggest that gender differences in speech may emerge even earlier – perhaps by the age of 3.

Despite well-documented differences in production between male and female speech, no study to date has asked whether these translate into language-specific differences in the perception of speaker gender. In other words, no study has asked whether listeners are more accurate at identifying speaker gender in speakers of their own language community than speakers from another language community. Perhaps this is because physiological cues to speaker sex (e.g., average f0) are so salient in adult speech that they might overwhelm any differences due to language-specific experience. But as outlined above, gendered speech patterns in prepubertal children are thought to be learned (i.e., not due to physiological differences between boys and girls). Thus, one question that is ripe for investigation is whether a listeners' ability to distinguish boys and girls depends on the listener's language background.

In the current study, we ask when perceptible differences in boys' and girls' productions emerge, and if sensitivity to gender differences in children's speech is tied to a listener's language experience. To accomplish this, we use a set of recordings collected longitudinally, with the same set of native English-speaking children recorded at 2.5, 4, and 5.5 years of age. We then tested L1 English and L2 English speakers' accuracy at identifying these children's gender at these three ages. We hypothesize that adults should perform above chance at classifying speech productions as belonging to boys and girls by at least 4 years of age, and that performance in differentiating boys from girls should improve with child age. In addition, we hypothesize that language experience should modulate listeners' ability to classify children's gender. Thus, we predict L1 listeners will be more accurate than L2 listeners in both classifying child gender and gauging their performance (i.e., confidence rating more in line with accuracy).

2. Method

2.1 Participants

Forty-eight adults from the University of Toronto participated in the perception study. Half were L1 learners of English (4 males, M_{age} = 20.8) and half were L2 learners of English (12 males, M_{age} = 19.52). All L1 learners acquired English in Canada before the age of 6, and used English at least 80% of the time; all L2 learners moved to Canada after the age of 14 and had minimal to no classroom exposure to English before the age of 14. L2 learners' first languages were Mandarin (18) and Cantonese (6). Participants reported no hearing or vision impairments at the time of testing.

2.2 Stimuli

Stimuli were drawn from a child speech corpus consisting of isolated words, elicited using an experimenter-controlled video game (Cooper et al., 2020). Children viewed a target word on a computer screen (e.g., ball, duck) and were prompted to name the picture in citation form. Twenty-four words were chosen for use in the current study. All chosen words were produced by the same 12 Canadian English-learning children (6 assigned male at birth and 6 assigned female at birth) at each of three different ages: 2.5, 4, and 5.5 years old. See Supplementary Material[2] for a full set of experimental words. These children were identified as cisgender by their primary caregiver(s). Stimuli were normalized for root mean square amplitude in Praat 6.0.22 (Boersma & Weenink, 2020).

2.3 Procedure

Listeners were tested individually in a quiet room using PsychoPy3 (Peirce & MacAskill, 2018). On each trial, a word was presented once, and listeners were asked to indicate the child speaker's gender by clicking on a male or female icon (similar to the icons used to indicate bathrooms, and additionally marked by the colours blue and pink)[3]. To assess whether language experience affected listeners' confidence in their performance, participants were also asked to rate their confidence on a scale from 1 (not at all) to 5 (very confident).

To ensure that participants understood the task, the experimental trials were preceded by two practice trials (consisting of two children voices that were not included in the experimental trials). Across the experiment, participants heard 24 words with each word produced by a different boy/ girl pair at 2.5, 4, and 5.5 years of age. Thus, each listener heard tokens from all 12 children, each at three ages over the course of 144 trials (24 Words $\times$ 2 Gender $\times$ 3 Ages). The trial order was randomized for all participants. The study took approximately 10 minutes to complete.

3. Results

3.1 Perception

To assess L1 and L2 listeners' performances across the three children's ages, we fit a generalized mixed-effects model to our data using the *glmer* function of the *lme4* package (Bates et al., 2015) in R. The binary response variable was Accuracy (1 = correct response). The independent variables, listener Group and children's Age, were entered as fixed effects. An interaction term between the two fixed effects was not included because, for the purpose of the study, we are primarily focused on the difference between L1 and L2 listeners' abilities (not the relative magnitude of this difference across age). We included random intercepts for Word, and a random slope for children's Age by Participant. Listener Group was simple-coded (with L1 listeners as the reference level). In addition, because we expected listeners would be more accurate with older children than with younger children, we coded children's Age with Helmert contrasts: 1) 2.5-year-olds vs. 4- and 5.5-year-olds combined, and 2) 4-year-olds vs. 5.5-year-olds. The β-coefficient for the intercept represents the log odds of a correct response averaged across all ages, and the β-coefficient corresponding to each effect represents the difference in log odds of a correct response between the two levels of that comparison, collapsed over all levels of the other factor.

As predicted, a significant intercept was found, indicating listeners' overall performance above chance (see Table 1 and Fig. 1a). Moreover, the model revealed a significant main effect of listener Group, with L1 listeners performing better than L2 listeners (irrespective of children's age)[4]. The model also revealed that listeners' performance differed significantly between 2.5-year-olds vs. older children, but no difference was found between 4- and 5.5-year-olds. Importantly, a subsequent follow-up test, using the same model but with 2.5-year-olds coded as the reference level, shows that listeners' performances with 2.5-year-olds were significantly above chance (β = 0.57, SE = 0.09, z = 6.67, p < 0.001). Note that this is the first study to date to demonstrate that children this young already produce perceptible gender differences in their speech.

Table 1. Summary of results from a logistic regression model for gender classification.

	β	SE	z	p
Intercept	0.68	0.08	8.67	< 0.001
Group: L2 listeners (vs. *L1 listeners*)	-0.22	0.05	-4.11	< 0.001
Age 2.5 (vs. *Age 4 & Age 5.5*)	0.16	0.06	2.88	0.004
Age 4 (vs. *Age 5.5*)	0.07	0.07	1.01	0.31

Note. Model: glmer(accuracy ~ group + age + (1 | word) + (age | participant)). Italics indicates the reference level.

In addition, we analyzed whether listeners were able to gauge their performance accurately. To assess this, we used a linear mixed-effects model to predict Confidence rating from Accuracy, listener Group, and their interaction. Listener Group was simple coded (with L1 listeners as the reference level) and Confidence rating was centred. The model showed that both Accuracy (β = 0.39, *SE* = 0.04, t = 9.91, p < 0.001) and listener Group (β = 0.17, *SE* = 0.04, t = 3.90, p < 0.001) significantly predicted Confidence rating. A significant interaction was also found (β = -0.18, *SE* = 0.06, t = -3.33, p < 0.001), with L1 listeners (*M* = 3.37, *SE* = 0.05) reported lower confidence than L2 listeners (*M* = 3.53, *SE* = 0.05) for incorrect responses, but equally high confidence for correct responses (L1 listeners: *M* = 3.78, *SE* = 0.04; L2 listeners: *M* = 3.77, *SE* = 0.05)[5]. The results indicate that, overall, listeners were able to gauge their performance accurately, but the ability to gauge their performance when incorrect was modulated by language experience (see Fig. 1b).

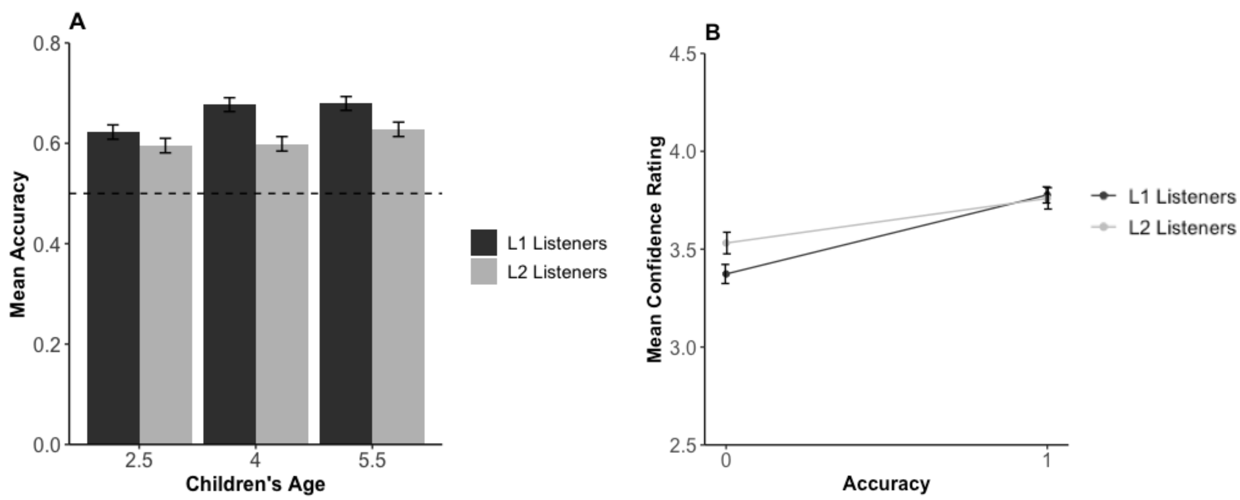

Fig. 1. a: Mean accuracy of gender classification of 2.5-, 4-, and 5.5-year-olds by L1 and L2 listeners. Dashed line at 0.5 indicates the chance level. b: Mean confidence rating of L1 and L2 listeners when responses were correct (Accuracy = 1) and incorrect (Accuracy = 0). Error bars indicate SE based on by-participant means.

3.2 Acoustics

We conducted follow-up tests to examine 1) whether children in our sample exhibited gender and age differences on several acoustic measures (f0, F1, F2), and 2) which, if any, of these acoustic measures are predictive of L1 and L2 listeners' responses, focusing on monosyllabic words to reduce variability. We manually annotated the vocalic portion of all tokens (including the vowel as well as preceding/following /l/ or /r/ when present). Given the inherent difficulty in estimating f0 and formant values in children's speech using static parameters, we manually inspected each token and chose the optimal parameters for obtaining accurate acoustic measures for f0 and formants for each token independently. Using these parameters, measures of f0, F1, and F2 were taken at the midpoint of the vocalic interval. Five words with diphthongs or coda /r/ were excluded from formant analyses due to the dynamic nature of the formants. Tokens for which was not possible to obtain an accurate pitch (8) and/or formant track (27) were omitted from the relevant analyses, and mispronounced tokens (2) were omitted entirely.

To examine whether acoustic measures differed by gender and age, we performed separate analyses for f0, F1, and F2 (see Fig. 2) using linear mixed-effects models (implemented using the *lmer* function of the *lme4* package (Bates et al., 2015) whereas *p*-values were computed using the *lmerTest* package; Kuznetsova et al., 2017). In each model[6], the acoustic dimension was the response variable, with Gender, Age, and their interaction as fixed effects. Both predictors were simple-coded, with reference levels as Male for Gender and 4-year-olds for Age. β-coefficients represent mean differences in the acoustic value between levels of the predictor factor. Only significant and trending main effects or interactions ($p < .010$) are reported here. As expected, the model for f0 revealed a significant decrease across the three ages ($\beta = 18.79$, $SE = 9.89$, $z = 1.90$, $p = 0.03$). Although there was no significant effect of Gender, a significant Age $\times$ Gender interaction was found in the comparison of 4- and 5.5-year-olds ($\beta = -45.10$, $SE = 18.58$, $z = -2.43$, $p = 0.02$). Wald $\chi 2$ test was performed using the phia package to explore this significant interaction (De Rosario-Martinez, 2015). A follow-up test indicated that 5.5-year-olds, but not 4-year-olds, showed a marginal gender difference (5.5-year-olds: $\chi 2 = -23.49$, $p =$

0.056; 4-year-olds: $\chi2 = -5.21$, $p = 0.37$). For F1, a marginally significant difference in the expected direction was found between 4- and 5.5-year-olds ($\beta = -72.54$, $SE = 32.66$, $z = -2.22$, $p = 0.07$). In summary, we found only marginal gender differences, and only in 5.5-year-olds, with both in the expected direction. However, it is important to note that because stimuli were chosen for ease of production by 2.5-year-olds, these analyses were therefore based on a relatively small amount of data; given the low power, these null results should be interpreted with caution.

Next, we examined how well the acoustic measures discussed above predicted L1 and L2 listeners' responses. We performed separate analyses for f0, F1, and F2 using a generalized mixed-effects model[7] in R. The model predicted listeners' Responses of "female" (vs. "male") from the acoustic measures (f0/F1/F2) and listener Group as well as their interaction. Listener Group was simple coded (with L1 listeners as the reference level). As above, we included random intercepts for Word and Participant. The β-coefficient represents the difference in log odds of a "female" response between the two levels of that comparison, collapsed over all levels of the other factor. The model revealed that, overall, listeners' responses are predicted by f0 ($\beta = 0.01$, $SE = 0.001$, $z = -7.86$, $p < 0.001$), F1 ($\beta = 0.003$, $SE = 0.0003$, $z = 8.23$, $p < 0.001$), and F2 ($\beta = 0.001$, $SE = 0.0001$, $z = 6.88$, $p < 0.001$), with low values eliciting mostly "male" response and high values eliciting mostly "female" response. This is consistent with the literature that lower pitch and formants are associated with males' voices. However, no difference was found between L1 and L2 listeners for any of the three dimensions (i.e., there was no significant interaction between Group and any of the acoustic measures).

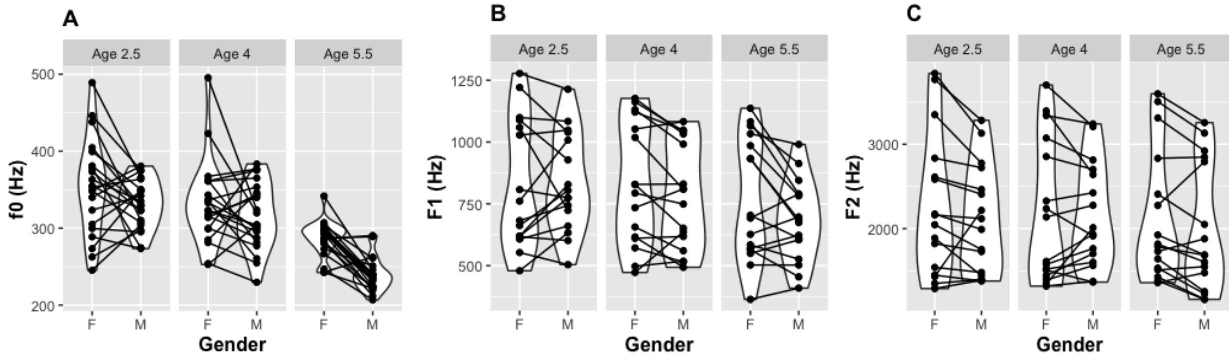

Fig 2. a-c: Gender differences in f0, F1, and F2 values across the three ages in children.

4. Discussion

Past work has demonstrated that gender differences in children's speech are perceptible by 4 years of age, long before anatomical differences in children's vocal tract anatomy are generally thought to have developed (see Munson & Babel, 2019, for discussion). These gender differences in speech are thought to be learned early in life, with what constitutes typical male or female speech patterns differing to some degree across language communities. In the current study, we investigate the possibility that gendered speech patterns can be detected in children even younger than 4 years of age, and also ask whether adult listeners' sensitivity to these patterns is dependent on language experience.

In line with our predictions, we found that L1 listeners outperformed L2 listeners in their gender classification of child voices. Moreover, both L1 and L2 listeners performed well above chance in classifying the gender at all three age groups tested, including the 2.5-year-olds. As far as we are aware, this is the first published study to demonstrate an apparent role of language experience in gender classification, and also the first study to show that children as young as 2.5 years of age show perceptible gender differences in their speech productions. Our findings with L1 listeners fit well with other longitudinal studies on the development of gendered speech patterns (Munson et al., 2019). In our follow-up acoustic analyses, we also found that all three acoustic measures (f0, F1, F2) influenced listeners' judgments of children's gender, although we found only marginal evidence of gender-based acoustic differences. This suggests that the task might have tapped into participants' stereotypes about gender differences in adults, rather than to their knowledge of differences between boys' and girls' speech.

Given the weak relationship between our acoustic measures and child gender, how did listeners perform above chance in our gender classification task? One possibility is that listeners based their classifications on aspects of the signal that we did not measure. This would be consistent with claims that early gender classifications might be based on more holistic qualities such as voice quality. For example, Günzburger and Keurs (1987) found that listeners judged girls' voices to be significantly clearer, softer, shriller, high-pitched, melodious and precise than boys' voices. In contrast, these same listeners described boys' voices as duller, louder, deeper, more monotonous and more careless than girls' voices. These descriptions are in line with a rating study we are currently running with these same recordings, where participants tend to describe boys' and girls' voices with similarly stereotyped labels. It is possible that these labels reflect community norms for gendered speaking styles, and that they could have driven listeners' classification decisions in our study. Additional support for this view comes from previous studies also failing to find robust spectral differences in children under 6 years of age (e.g., Lee et al., 1999; Perry et al., 2001; Vorperian et al., 2019).

Our finding that listeners' language experience impacts how accurately they can classify boys' and girls' voices fits well with the hypothesis that gendered speech in children is learned (i.e., not simply due to biological maturation). This is also consistent with the view from the linguistic literature that variations in speech patterns reflect the process of social differentiation (Eckert, 2012; Foulkes et al., 2010). Interestingly, we found that L1 and L2 listeners both relied on f0, F1, and F2 when making gender judgments. Therefore, other characteristics of the speech signal that are not measured here may be responsible for this performance differential between L1 and L2 listeners. A likely explanation, given the literature on the language-specific nature of gender differences in speech, is that the L1 listeners had knowledge of language-specific cues to talker gender that the L2 listeners lacked. Presumably, these cues were not captured in our acoustic measures. Future studies could address this issue more directly by manipulating specific cues to talker gender that differ between two language populations (e.g., see if Japanese speakers are more likely to classify children with a wider f0 as boys whereas English speakers are more likely to classify them as girls), as opposed to previous perception studies which only focused on English speakers. However, another possible explanation for performance differences between L1 and L2 listeners is that the latter were distracted by attempts to understand the children in this study, giving them few processing resources to devote to gender classification. We find this explanation unlikely, however, because our task did not require listeners to comprehend child talkers. Nonetheless, future studies could test this possibility by having the same target word or phrase used on every gender classification trial.

To conclude, gender differences in children's speech clearly emerge even earlier than previous work has demonstrated, but L1 listeners are better at detecting these differences than L2 listeners. Logically speaking, this may be because at least some of the realization of gender in children's speech is language-specific. These findings generate many hypotheses about the sociolinguistics of early speech acquisition, and make exciting predictions regarding communication and development in linguistically-diverse communities.

Acknowledgements

This research was supported by grants awarded to E.K.J. from the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada.

Footnotes

[1] Note that the cited literature on this topic uses the terms gender and sex loosely, often failing to include a robust measure of gender identity, and not give an explanation for how talker sex was determined.

[2] See supplementary material at [URL will be inserted by AIP] for list of experimental words.

[3] We are aware that forcing participants to make a binary decision does not reflect the complex realities of gender identities in the real world, but this binary classification task was appropriate for addressing our question of interest in the current study.

[4] Although it was not included in the initial analysis, based on a reviewer's comment, we tested whether there might be a children's Age $\times$ listener Group interaction. We found a trending interaction ($p = 0.07$). Upon breaking down the interaction, we found that the effect was significant for 4- and 5.5-year-olds but not for 2.5-year-olds. Although we cannot draw firm conclusions given that the interaction was only trending, this suggests that looking into the trajectory of this effect across ages is an interesting topic for future work.

[5] The same main results were found when the data was modeled using a Poisson distribution.

[6] Model: lmer(f0/F1/F2 ~ gender $\times$ age + (1|word) + (1|participant))

[7] Model: glmer(female.response ~ f0/F1/F2 $\times$ group + (1|word) + (1|participant))

References and links

Bennett, S., and Weinberg, B. (1979). Sexual characteristics of preadolescent children’s voices. *The Journal of the Acoustical Society of America*, *66*, 989–1000. https://doi.org/10.1121/1.382234

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*. Singular Thomson Learning.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Package “lme4” v. 1.1.7. https://cran.r-project.org/web/packages/car/index.html

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females, *The Journal of the Acoustical Society of America*, *69*, 321–238. https://doi.org/10.1121/1.385343

Boersma, P. & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.27: http://www.praat.org/

Cheng, A. (2020). Cross-linguistic f 0 differences in bilingual speakers of English and Korean. *The Journal of the Acoustical Society of America*, *147*(2), EL67–EL73. https://doi.org/10.1121/10.0000498

Cooper, A., Fecher, N., & Johnson, E.K. (2020). Identifying children's voices. *The Journal of the Acoustical Society of America*, *148*, 324-333. https://doi.org/10.1121/10.0001576

De Rosario-Martinez, H. R. (2015). Package “Phia”. https://CRAN.R-project.org/package=phia

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology, 41*, 87-100. https://doi.org/10.1146/annurev-anthro-092611-145828

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511–1522. https://doi.org/10.1121/1.427148

Foulkes, P., Docherty, G., Khattab, G., & Yaeger-Dror, M. (2010). Sound judgements: perception of indexical features in children’s speech. In D. Preston, & N. Niedzielski (Eds.), *A Reader in Sociophonetics* Mouton de Gruyter.

Günzburger, D., Bresser, A., & Keurs, M. Ter. (1987). Voice identification of prepubertal boys and girls by normally sighted and visually handicapped subjects. *Language and Speech*, *30*(1), 47–58. https://doi.org/10.1177/002383098703000104

Heffernan, K. (2004). Evidence from HNR that /s/ is a social marker of gender. *Toronto Working Papers in Linguistics*, *23*.

Johnson, K. (2008). Speaker normalization in speech perception. In D. B. Pisoni, & R. E. Remez (Eds.), *The handbook of speech perception* (pp.363389, Chapter xi, 708 Pages). Blackwell Publishing, Malden.

Kuznetsova A., Brockhoff P. B., Christensen R. H. B. (2017) lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v082.i13

Ladegaard, H. J., & Bleses, D. (2003). Gender differences in young children's speech: the acquisition of sociolinguistic competence. *International Journal of Applied Linguistics*, *13*(2), 222-233. https://doi.org/10.1111/1473-4192.00045

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America, 105*, 1455-1468.

Li, F., Rendall, D., Vasey, P. L., Kinsman, M., Ward-Sutherland, A., & Diano, G. (2016). The development of sex/gender-specific /s/ and its relationship to gender identity in children and adolescents. *Journal of Phonetics*, *57*, 59–70. https://doi.org/10.1016/j.wocn.2016.05.004

Lindbiota, B., and Stmdberg, J. (1971). Acoustical Consequences of Lip, Tongue, Jaw and Larynx Movements, *The Journal of the Acoustical Society of America*, *50*, 1166-1179. https://doi.org/10.1121/1.1912750

Morris, R., McCrea, C.R., & Herring, K.D. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics*, *36*, 308-317. https://doi.org/10.1016/j.wocn.2007.06.003

Munson, B., Lackas, N., & Koeppe, K. (2019). *The Longitudinal development of gendered speech production in children: patterns and predictors.* Oral presentation at the Boston University Conference on Language Development, Boston, MA.

Munson, B., & Babel, M. (2019). The phonetics of sex and gender. In W. Katz & P. Assmann (Eds.), Routledge Handbook of Phonetics.

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics, 39*, 59-67. https://doi.org/10.1016/j.wocn.2010.11.002

Peirce, J. W., & MacAskill, M. R. (2018). Building Experiments in PsychoPy. London: Sage.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*(6), 2988–2998. https://doi.org/10.1121/1.1370525

Robb, M., Gilbert, H., & Lerman, J. (2005). Influence of gender and environmental setting on voice onset time. *Folia Phoniatrica Et Logopaedica: International Journal of Phoniatrics, Speech Therapy and Communication Pathology*, *57*(3), 125-133. https://doi.org/10.1159/000084133

van Bezooijen, R. (1995). Sociocultural Aspects of Pitch Differences between Japanese and Dutch Women. *Language and Speech*, *38*(3), 253–265. https://doi.org/10.1177/002383099503800303

Vorperian, H. K., Kent, R. D., Lee, Y., & Bolt, D. M. (2019). Corner vowels in males and females ages 4 to 20 years: Fundamental and F1–F4 formant frequencies. *The Journal of the Acoustical Society of America*, *146*(5), 3255–3274. https://doi.org/10.1121/1.5131271

Yang, S., & Mu, L. (1989). An investigation of the third formant of /a/ in prepubertal children. *Journal of Voice*, *3*(4), 321–323. https://doi.org/10.1016/S0892-1997(89)80054-2