How do children mark gender phonetically? An analysis of voice quality

Eugene Wong, Benjamin Munson

University of Minnesota
wong0703@umn.edu, munso005@umn.edu

ABSTRACT

Children assigned female at birth (AFAB) have identical vocal tracts and larynxes to children assigned male at birth (AMAB) before puberty. Nonetheless, previous studies have found that adult listeners perceive the speech of AFAB children as more girl-like than that of AMAB children, suggesting that some gendered speech characteristics are learned before the development of vocal tract sex dimorphism. AFAB and AMAB children differ in many segmental acoustic measures. This study examined differences in voice quality between the two groups of children, given the well documented voice quality differences between adult men and women. We examined various acoustic indices of voice quality of 110 English-speaking children at three and at five years-of-age. We found only minimal differences in the voice quality between the groups at both time-points. We found that these measures can predict gender ratings of children's speech, suggesting that people apply stereotypes about adult speech when rating children.

Keywords: sociophonetics, voice quality, gender, selective learning, language acquisition

1. INTRODUCTION

It is well documented that adult men and women speak differently. Some of the phonetic differences found between cisgender men and women can be attributed to the sexual dimorphism in their speech-production mechanisms. As a group, cisgender women tend to speak with higher fundamental frequencies and resonant frequencies than cisgender men. One reasonable hypothesis is that these differences are a passive consequence of sex dimorphism in the vocal tract. However, there are numerous phonetic differences between men and women that are not attributable to sex dimorphism. Studies of children's speech lend support to the conjecture that gendered speech represents culturally and linguistically specific learned ways of speaking.

Prior to puberty, there are no consistent group-level differences in vocal-tract sizes between children assigned female at birth (AFAB) and children assigned male at birth (AMAB) [1, 2]. Throughout this paper, we use AFAB and AMAB instead of 'girl' and 'boy', as we reserve the latter set of terms for studies in which children were asked directly for their gender identity, and in which options more than just 'boy' and 'girl' were given. The speech of AFAB and AMAB children is not identical: the speech of children as young as 2.5 years of age can be robustly distinguished based on sex assigned at birth (SAB) by adult listeners [3]–[5]. In these studies, the speech of AFAB children were rated as more girl-like and AMAB children as more boy-like. These findings suggest that children have learned to express gender phonetically by age 2.5. However, it is unclear what phonetic features distinguish the speech of AFAB and AMAB children, and what acoustic cues allow adults to distinguish the two groups of children.

Previous studies examining acoustic features of gender in children's speech have found that some differences in vowel formant frequency values between AFAB and AMAB children emerge at around 4 years of age, but are not markedly and consistently different until 8 years of age. Moreover, there is no gender difference in habitual f0s until 8 years of age [3, 6, 7]. One study [8] found differences in the overall scaling of vowels' formant frequencies differed between AFAB and AMAB children at age 5, but not at age 3. The measure in [8] used the average distance between adjacent formants to estimate the length of the vocal tract during vowel production, which was adapted from [9]. Another analyses of the same group of children's /s/ productions found gender differences in the spectral mean frequency at both 3 and 5 years of age, consistent with differences between adult men and women [10]. Taken together, these studies show that children mark features related to gender early in life, and that their expression of gender change across early development.

One phonetic domain that has rarely been discussed in previous literature on gender differences in children's speech is voice quality. It is generally thought that there are two dimensions of voice quality. One of these reflects the habitual voice settings of an individual. In adults, part of this habitual voice quality is tied to sex dimorphism of the laryngeal structures between adult men and women, given that individuals can still manipulate their voices to a certain degree. This dimension of voice quality variation also reflects upper respiratory health, and explains why voice quality is often discussed in the

context of speech pathologies. The other dimension is intentional variation in voice quality to signal lexical contrasts, indexical meanings, or both. For instance, creaky voice is a variable commonly used by young Americans for social indexical purposes [11]. As the laryngeal structure in prepubertal children are the same [1], any acoustic differences found in the voice quality between AMAB and AFAB children should be indicative of their learning and expression of gender. In previous studies, [12] found that five-years-old AMAB and AFAB children exhibit differences in one index of voice quality, contact quotients in the English /a/ vowel, yet the gendered pattern did not mirror those found in adults. The authors argued that such difference might arise from the larger variability in contact quotients in AMAB children, and concluded there were only minimal differences in laryngeal features between the two groups of children. Another study [13] found that voices of nine-years-old German-speaking AMAB children were perceived as less breathy than AFAB children. The findings from [12, 13] suggest that children might learn ways of producing voice quality that reflect their gender. Nevertheless, recent studies from [4, 5] found that gendered speech emerges much earlier than the age range in [12, 13]. To date, there is no literature that examines gender differences in voice quality in children of 3 to 5 years of age with a large group of participants. To address this gap, we examine the acoustic voice quality differences between AMAB and AFAB children. We also examine whether voice quality measurements predict adult listeners' gender ratings of children's speech, using data reported previously in [5].

2. METHODS

2.1. Participants

Speech samples were collected from 55 AMAB and 55 AFAB children who are monolingual speakers of American English. The children participated in a longitudinal study of vocabulary development, where speech samples were collected at the three time-points at one-year intervals. The first time-point (FTP) was at 2.5 to 3.5 years of age, the second time-point was at 3.5 to 4.5 years-of-age, and the last time-point (LTP) being 4.5-5.5 years of age. In this study, we only examined the speech samples collected in FTP and LTP. The children were recruited in Minneapolis, MN or Madison, WI. Among the 110 participants, seven children (5 AFAB) speak African American English (AAE), and the remaining 103 children speak the local white variety of American English (wAE, 50 AFAB). All children reported no history of speech, language nor hearing disorders. All caregivers provided consent for their children to participate in the study.

2.2. Productions

The productions were a subset of the approximately 100 productions elicited in the larger study. That word set was designed to elicit a variety of word-initial consonants in a variety of vowel contexts. The words were chosen because they were likely to be familiar to children of the age being tested, and hence were not identical across time-points. In this study, we examined a subset of productions consisting of the low vowels /æ/ and /a/ in a primary-stressed syllable for acoustic measurement of children's voice quality.

2.3. Data collection

The children completed a real-word-repetition task at each time-points. In this task, an experimenter played the recorded target words one at a time and asked the child to repeat the words, while showing a picture of the target words, at a randomized order. Some words were repeated twice to elicit sufficient tokens for analysis. The prompts used with the wAE speaking children was a speaker of the local wAE variety, while those used with AAE speaking children was a native speaker of AAE.

In addition to the production experiment, gender ratings of these children were collected from 80 naïve adult listeners. The listeners ratings were the same as those obtained in [5]. In short, the adult listeners were presented the children's productions one-by-one, and were asked to rate on a visual-analogue scale from "definitely a boy" to "definitely a girl". A mean gender rating was then obtained by averaging the listeners' numerical responses for each child. The numerical values ranged from 0 "definitely a boy" to 1 "definitely a girl".

2.4. Data Analysis

Textgrids were created using Praat [14] and aligned with the Montreal Forced Aligner [15]. The alignments were manually inspected and corrected when necessary. Acoustic measurements were performed on the vowels using VoiceSauce [16]. Voice quality of the vowels was gauged using two spectral tilt parameters (H1*-H2* and H2*-H4*), five spectral slopes parameters (H1*-A1*, H1*-A2*, H1*-A3*, H4*-H2k* and H2k*-H5k*), Cepstral Peak Prominence (CPP) and Harmonic-to-noise ratios at four frequency ranges 0-500Hz (HNR05), 0-1500Hz (HNR15), 0-2500Hz (HNR25), 0-3500Hz (HNR35). Asterisks denote the correction of bandwidth and formants. Results of the acoustic measurements were then time-normalized into three

segments of equal duration. Only the middle segment was used for subsequent analysis. In addition, vowel tokens with total duration shorter than 50ms were removed. As a result, our final data consisted of 1,661 vowels for FTP, and 2,071 vowels for LTP.

A principal component analysis (PCA) was then performed. Since the initial data showed a bimodal distribution, which was related neither to vowel differences nor to the productions of particular child or subset of children, the PCA analysis would potentially reduce the bimodality of the data. Moreover, as the acoustic parameters also showed moderate correlation with each other, the PCA analysis was expected to yield a smaller set of acoustic variables. All parameters were centred and scaled before entering the computation of PCA.

3. RESULTS

3.1. Principal components analysis

There were four principal components (PCs) that had eigenvalues greater than 1. In total, they accounted for 84.8% of the variance in the data. Acoustic parameters with loadings over 0.32 in absolute values were considered to form a component. Table 1 shows the acoustic parameters that form the four PCs, together with the percentage of variance explained.

Table 1: Components of PC1 to PC4

PC	Acoustic parameters	% of variance explained
PC1	HNR05, HNR15, HNR25, HNR35	35%
PC2	H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*, H2k*-H5k*	30%
PC3	H1*-H2*, H1*-A1*, H4*-H2k*, H2k*-H5k*	12%
PC4	H1*-H2*, H2*-H4*, H4*-H2k*	9%

For PC1, the four HNRs shared a positive loading to this component. For PC2, all parameters except H2k*-H5k* shared a positive loading towards the component. This suggest that for PC2, a higher H2*-H4*, and the three higher H1*-An* parameters were associated with a lower H2k*-H5k* value. For PC3, only H4*-H2k* had a positive loading, whereas H1*-H2*, H1*-A1* and H2k-H5k* shared a negative loading. In other words, a higher H4*-H2k* was associated with lower values in the other three parameters. Finally, for PC4, H1*-H2* and H4*-H2k* shared a positive loading, whereas H2*-H4* had a negative loading. This indicates that higher H1*-H2* and H4*-H2k* were associated with a lower H2*-H4* value.

3.2. Voice quality differences between AMAB and AFAB children

After we reduced the acoustic parameters to four PCs, we examined whether the AFAB and AMAB children differ in these measures at each time-point. Linear mixed effect models were constructed using the fitted values obtained from each of the four PCs as response variables. The models included a random intercept for participant. Coefficient-level results are presented in Table 2. AFAB children are used as reference level.

Table 2: Coefficient-level results of the LMER models on voice quality predicted by SAB.

PC	Predictor	β (SE)	t	p
First time-point (FTP)				
PC1	(Intercept)	-0.37 (0.17)	-2.23	0.03
	SAB	0.32 (0.23)	1.35	0.18
PC2	(Intercept)	-0.09 (-0.01)	-0.83	0.41
	SAB	-0.011 (0.15)	-0.07	0.94
PC3	(Intercept)	-0.13(0.06)	-2.21	0.03
	SAB	0.18 (0.08)	2.18	0.03
PC4	(Intercept)	0.06 (0.05)	1.33	0.19
	SAB	-0.06 (0.07)	-0.83	0.41
Last time-point (LTP)				
PC1	(Intercept)	0.20 (0.18)	1.12	0.27
	SAB	-0.086 (0.25)	-0.35	0.73
PC2	(Intercept)	-0.012 (0.11)	-0.12	0.91
	SAB	0.15 (0.15)	0.99	0.32
PC3	(Intercept)	0.028 (0.06)	0.50	0.62
	SAB	0.025 (0.08)	0.32	0.75
PC4	(Intercept)	-0.06 (0.04)	-1.45	0.15
	SAB	0.077 (0.06)	1.22	0.23

As revealed in Table 2, at FTP, only PC3 was found to have a main effect from SAB. Recall that PC3 consisted of a negative loading of H1*-H2*, H1*-A1*, H2k*-H5k* and a positive loading of H4*H2k*. The coefficient intercept of -0.13 indicates that AFAB children had a higher H1*-H2*, H1*-A1* and H2k*-H5k*, and a lower H4*-H2k* than the AMAB counterpart. In other words, AFAB children had laxer voice quality than AMAB children at FTP. At LTP, none of the PCs were found to have a main effect from SAB. Overall, the statistical results suggest that there was at best only a minimal difference in voice quality between the two groups of children.

3.3. Predicting gender ratings

We then examined whether perceived gender ratings collected form adult listeners were predicted by the

measurements of children's voice quality. To facilitate this analysis, for each child we computed average values for the four PCs using the fitted values from the mixed models. We then constructed simple regression models using the four average PC values as predictors, and the gender ratings as the response variable. The regression models were constructed separately for each group of children at each time-points. The initial model included all four PCs, since together they represented the voice quality of the children. The models then went through a backward stepwise procedure to select the best model, based on AIC values. The results showed that only one PC remained in each of the models. This suggests that voice quality parameter might be an acoustic cue to children's SAB, and adult listeners rely on one PC of voice acoustic cue. Table 3 and 4 summarize the coefficient-level results of the best fitted model for each SAB and for each time-point.

Table 3: Coefficient-level results of linear models of gender ratings predicted by voice quality at FTP.

SAB/ Predictor	*β* (SE)	*t*	*p*
AFAB			
(Intercept)	0.586 (0.015)	39.76	<0.001
PC3	0.051 (0.035)	1.49	0.143
AMAB			
(Intercept)	0.498 (0.015)	33.4	<0.001
PC2	-0.041 (0.019)	-2.1	0.041

As shown in Table 3, PC2 had a main effect on gender ratings of AMAB children. The coefficient estimate of -0.041 suggests that a lower PC2 (i.e. a lower H2*-H4*, H1*-An* and higher H2k*-H5k*) was associated with a lower gender ratings, i.e. more towards "definitely a boy". PC3, on the other hand, was retained in the model for predicting gender ratings of AFAB children, suggesting that PC3 might contribute to gender ratings. However, the p-value for the coefficient in the model itself was not statistically significant. Together, these findings indicate that the effect of PC3 on ratings was, at best, small.

Table 4: Coefficient-level results of linear models of gender ratings predicted by voice quality at LTP.

SAB/ Predictor	*β* (SE)	*t*	*p*
AFAB			
(Intercept)	0.584 (0.016)	36.89	<0.001
PC1	0.026 (0.012)	2.24	0.029
AMAB			
(Intercept)	0.45 (0.015)	30.31	<0.001
PC1	0.019 (0.012)	1.52	0.134

In LTP, both the models for predicting gender ratings of AFAB children and of AMAB children consist of PC1. The positive coefficient estimates from both models (0.019 and 0.026) indicate that higher HNR values (the components of PC1) were associated with an increase in gender ratings (towards "definitely a girl").

4. DISCUSSION

We found that the two groups of children differed in one PC (PC3) at 3 years of age, but not at 5 years of age. Specifically, the two groups of children showed speech differences in their spectral shape. A lower H1*-H2*, H1*-A1* and H2k*-H5k*, but higher H4*-H2k*, were associated with the group of AMAB children. To our knowledge, this finding is just the second speech parameter, next to analyses of /s/ production in [10], that distinguished children's SAB at 3 years of age. However, the overall variance explained by PC3 was relatively low (the r^2 is only 12%) in this set of data. This may suggest that only a small subset of children was utilizing voice quality cues in their expression of gender. However, as we found no SAB difference in the voice quality of children at 5 years of age, we argue that there were overall minimal differences in voice quality between the groups of children.

Our results showed that adult listeners seemed to use some acoustic cues of voice quality to rate children's gender, although there were minimal differences in voice quality between the two groups of children. This reflects that adult listeners were generalizing acoustic cues for gender differences in adult talkers when rating children's gender. This may be because f0 cues, which is typically used to distinguish adult voice gender, are absent in young children. In LTP, PC1 (which consists of HNR parameters) predicted children's gender ratings. The use of HNR was also found in previous study in which older children use breathiness to express gender [13]. In FTP, PC2 and PC3 (consists of spectral shape parameters) predicted adults' gender ratings, but not PC1. These pieces of evidence seem to suggest that adults, when they are asked to identify speaker gender in the absence of f0 differences, turn to noise cue (PC1) before spectral cues (PC2 and PC3). The noise cue may be a more salient cue for gender expression in older children and adults. When noise cues are also not salient, adults turn to spectral shape cues. However, both PC1 and PC2 were not the acoustic cues that children use in their gendered speech. The mismatch in production and perception cues suggest that voice quality is not a robust cue for gendered speech in children. Our results also confirm that gender can be expressed and perceived with multiple variables and idiosyncrasy exhibit across individuals. Future studies may explore the set of acoustic cues for perception of gendered speech.

5. REFERENCES

[1] Fitch, W. T., Giedd, J. 1999. Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J. Acoust. Soc. Am.* 106(3), 1511–1522.

[2] Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., Yandell, B. S. 2005. Development of vocal tract length during early childhood: a magnetic resonance imaging study. *J. Acoust. Soc. Am.* 117(1), 338–350.

[3] Perry, T. L., Ohde, R. N., Ashmead, D. H. 2001. The acoustic bases for gender identification from children's voices. *J. Acoust. Soc. Am.* 109(6), 2988–2998.

[4] Fung, P., Schertz, J., Johnson, E.K. 2021. The development of gendered speech in children: Insights from adult L1 and L2 perceptions. *JASA Express Lett.* 1, p. 014407.

[5] Munson B., Lackas, N., Koeppe, K. 2022. Individual Differences in the Development of Gendered Speech in Preschool Children: Evidence From a Longitudinal Study. *J. Speech Lang. Hear. Res.*, 65(4), 1311–1330.

[6] Lee, S., Potamianos, A., Narayanan, S. (1999) Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am.* 105(3), 1455–1468.

[7] Vorperian, H.K., Kent, R.D. 2007. Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. *J. Speech Lang. Hear. Res.* 50 (6), 1510–1545.

[8] Wong, E., Munson, B. 2021. A longitudinal study of gender-specific characteristics of children's vowels. J. *Acoust. Soc. Am.* 150(4), A152.

[9] Lammert, A. C., Narayanan, S. S. 2015. On Short-Time Estimation of Vocal Tract Length from Formant Frequencies. *PLOS ONE* 10(7).

[10] Koeppe, K. 2021. The emergence of gendered phonetic variation in preschool children: Findings from a longitudinal study. M.A. thesis, University of Minnesota, Minneapolis. Available: https://hdl.handle.net/11299/224487.

[11] Podesva, R. J., Callier, P. 2015. Voice quality and identity. *Annual Review of Applied Linguistics* 35, 173–194.

[12] Robb, M. P., Simmons, J. O. 1990. Gender comparisons of children's vocal fold contact behavior. *J. Acoust. Soc. Am.* 88(3), 1318–1322.

[13] Simpson, A. P., Funk, R., Palmer, F. 2017. Perceptual and Acoustic Correlates of Gender in the Prepubertal Voice. *INTERSPEECH 2017*, 914–918.

[14] Boersma, P., Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 6.2.14. Accessed: May 2022. [Online]. Available: https://www.praat.org

[15] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M. 2017. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *INTERSPEECH 2017*, 498–502.

[16] Shue, Y.-L., Keating, P. A., Vicenik, C., Yu, K. 2011. VoiceSauce: A program for voice analysis. *Proc. 17th ICPhS* Hong Kong,1846–1849.