INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Perceptual and acoustic correlates of gender in the prepubertal voice

Adrian P. Simpson[1], *Riccarda Funk*[2], *Frederik Palmer*[1]

[1]Institute of German Linguistics, University of Jena, Germany
[2]Institute of Music, Media and Speech Studies, University of Halle-Wittenberg, Germany
adrian.simpson@uni-jena.de, riccarda.funk@web.de, fp@frederikpalmer.eu

Abstract

This study investigates the perceptual and acoustic correlates of gender in the prepubertal voice. 23 German-speaking primary school pupils (13 female, 10 male) aged 8–9 years were recorded producing 10 sentences each. Two sentences from each speaker were presented in random order to a group of listeners who were asked to assign a gender to each stimulus. Single utterances from each of the three male and three female speakers whose gender was identified most reliably were played in a second experiment to two further groups of listeners who judged each stimulus against seven perceptual attribute pairs. Acoustic analysis of those parameters corresponding most directly to the perceptual attributes revealed a number of highly significant correlations, indicating some aspects of the voice and speech (f0, harmonics-to-noise ratio, tempo) that children use to construct and adults use to identify gender in the prepubertal voice.

Index Terms: prepubertal voice, perceptual correlates of gender, acoustic correlates of gender

1. Introduction

Gender in the adult voice is comprised of an inevitable biological component arising from anatomical differences in the vocal apparatus, as well as a learned, social component. The combination of both enables listeners to identify the gender of an adult speaker with a degree of accuracy approaching 100% [1].

Studies of the prepubertal child's voice suggest that anatomical and physiological differences between the genders have a negligible effect on the acoustic output. Nevertheless, in the majority of studies that have investigated the perception of gender in content-neutral speech of children correct gender identification has generally found to be at a level better than chance [2, 3, 4, 5, 6, 7, 8]. While a degree of variaton in gender-identification rates have been found, a range of sex-specific phonetic differences have been found in children as young as 2½ years old:

- Voice onset time (VOT): girls > boys, but also greater variance than found for adults [9, 10, 11, 12]
- F1 in open vowels: girls > boys, i.e. girls have acoustically more open /a(ː)/ than boys [13, 14, 15, 7, 16, 17]
- F1, F2, F3: girls > boys, with significant differences across nearly all age groups from 4 to 16 year-olds [14, 7]
- Levels of syllable reduction: boys > girls [18].
- f0 patterns: decrease in male fundamental frequency from 7–8 years old, but also an increase in use of downward shifts [19]
- Voice quality, e.g. showing a higher contact quotient in EGG data for boys in /a/ [20].

Reasons for some of these differences are still unclear. So, while the use of different f0 patterns is undoubtedly behavioural, it is unclear whether systematic differences in formant frequencies are the result of sex-specific differences in the size, and in particular, in the volume of the resonators, which may already be present from an early age [21], or whether such differences are also the result of different articulatory behaviours [16].

Two studies have found a number of significant correlations between listeners' gender judgements and acoustic parameters [22, 7]. Most prevalent are correlations between higher formant frequencies and females, starting with F3 at 4 years of age. However, despite finding systematic sex-specific differences in a number of phonetic parameters, results from all perceptual studies show gender identification in children's speech to be an extremely difficult task. So, for instance, even in those studies claiming overall correct identification rates of over 70%, individual groups are often not recognised at levels better than chance [23]. This suggests that faced with the difficult task of identifying the gender of a child, a listener attends to any number of parameters, but may ultimately be swayed by the presence of a single feature, and children themselves exhibit a great deal of phonetic variation [9].

The present study explores the perceptual and acoustic correlates of gender in the prepubertal voice. However, rather than just analysing a random sample of prepubertal male and female speakers, we use the results of a preliminary listening test to use precisely those speakers whose gender is most reliably identified. In a second listening experiment, similar to that used by Günzburger and colleagues [4], we had listeners rate stimuli against a number of perceptual attribute pairs, e.g. clear–hoarse, montonous–melodious. However, in order to avoid a possible confound identified in [4], whereby listeners are likely to have identified a child's gender before rating the perceptual attributes, we used two different listener groups. One group thought they were listening exclusively to male voices, the other exclusively to females. In a final step, acoustic parameters thought to correspond most directly to the perceptual attributes are analysed.

2. Method

2.1. Subjects and speech recordings

Subjects were drawn from a single primary school class in the German town of Wandersleben, located in the state of Thuringia, Germany. Prior consent for the recordings had been obtained from the parents and permission from the local authority. In all, 23 pupils (13 female, 10 male) were recorded in a quiet room at the school producing the short set of sentences shown in Table 1 together with the spontaneous description of a simple picture containing a well-known cartoon figure ("Das

http://dx.doi.org/10.21437/Interspeech.2017-1055

Sams")[1]. Subjects were 8–9 years of age (mean = 8.73 years), so just still within the limits of prepuberty [24]. Recordings were made using a Zoom H4 recorder with the built-in microphones placed at approximately 50 cm from the speaker's mouth. Recordings were made at a sampling frequency of 44.1 kHz and 16-bit amplitude resolution. The sentence list was read once from a single sheet of paper.

Table 1: *Sentence list*

	German	Gloss
1.	*Tim isst gerne Zitroneneis.*	Tim likes lemon ice-cream.
2.	*Die Oma mag Urlaub am Meer.*	Grandma likes to holiday by the sea.
3.	*Im Sommer blühen die Blumen.*	Flowers flower in the summer.
4.	*Beate fhrt mit dem Auto nach Köln.*	Beate is driving to Cologne by car.
5.	*Am Anfang war es bewölkt.*	In the beginning it was cloudy.
6.	*Im Garten fliegen die Bienen.*	Bees are flying in the garden.
7.	*Heute Morgen war ich zu spät.*	I was too late this morning.
8.	*Die Mama fährt morgen nach Jena.*	Mum is travelling to Jena tomorrow.
9.	*Der Schneemann hat eine rote Nase.*	The snowman has a red nose.
10.	*Maler malen einen Mann.*	Painters are painting a man.

2.2. Listening experiment – gender identification

Two sentences from each child were chosen at random to produce a total of 46 stimuli (=23 subjects $\times$ 2 sentences). Amplitude of all stimuli was normalised using a *praat* script [25][2]. The 46 stimuli were presented via headphones once in randomised order to 28 listeners (15 male, 13 female; age range: 19–54 y) using the *praat* experiment environment *ExperimentMFC*. None of the listeners reported having any hearing problems. Listeners were asked to decide whether a stimulus was produced by a boy or girl and also, on a five-point scale, how sure they were about their decision (1: unsure – 5: sure). We will not be examining the certainty ratings further here, since the subgroup of interest from the sample are those subjects whose gender was robustly identified. However, in further studies, when considering the whole sample, it will be interesting to see whether correct gender identification correlates with judgement certainty.

2.3. Listening experiment – perceptual attribute rating

From the results of the gender identification task (see Results section) stimuli from the three girls and three boys whose gender had been most successfully identified were used for a further listening experiment. 18 listeners (9 female, 9 male), not used in the previous experiment, were asked to assess different attributes of the voices producing the stimuli. To avoid a possible experimental confound identified above [4], listeners were split into two equally sized groups. Both groups heard exactly the same stimuli, but the first group was told they were listening to boys, the second that they were listening to girls. Listeners were asked to rate each stimulus against seven perceptual attribute pairs suitable for lay listeners: quiet–loud, fast–slow, clear–hoarse, monotonous–melodious, excited–calm, mumbled–clear. high–low. Each pair was rated on a 4-point scale. The stimuli were again presented over headphones using *praat*'s ExperimentMFC environment, but the listeners marked their ratings on paper. Listeners were allowed to listen to each stimulus as often as they liked.

[1]The spontaneous data was not used in the present study.

[2]*Equalizing amplitude (db)* from Shigeto Kawahara: http://user.keio.ac.jp/ kawahara/resource.html, accessed 25.06.2016

2.4. Acoustic parameters

The stimuli from the second listening experiment were subjected to acoustic analysis. A number of acoustic parameters considered to be most directly related to the perceptual attributes were analysed using *praat*. The *Voice report* facility was used to calculate mean fundamental frequency and standard deviation of each stimulus together with the mean harmonics-to-noise ratio of the final 200 ms of voicing. Tempo was calculated using the canonical number of syllables in each stimulus divided by the total duration of the stimulus, both with and without the total duration of any pauses produced. Due to the fact that a different sentence from each speaker was used as a stimulus with varying vowel category content, no measures of formant frequency were made.

3. Results

3.1. Gender identification

Figure 1 summarises mean gender identification rates for each speaker. Speakers are grouped by gender and ranked by correct identification within each group. At 69.2%, the mean overall gender identification rate is significantly above chance, and in line with results of other studies cited above. However, the mean identification rate belies considerable intra-group variation, with some speakers being correctly identified at rates well above 80%, while others are being incorrectly identified in the majority of cases, or rather assigned to the other gender.

3.2. Perceptual attribute ratings

The lefthand side of Table 2 summarises the results of the second listening experiment in which listeners were asked to rate a single utterance from each of the six most successfully identified male and female speakers against seven perceptual attribute pairs. Mean ratings for each attribute pair and speaker are shown in the table, as are the mean ratings for each gender. The 18 listeners had been split into two groups, with one half being told the stimuli had been produced by boys, the other half that they had been produced by girls. This was designed to cancel out any rating bias that may have arisen from listeners identifying the gender and then making a perceptual attribute rating. Comparison of the ratings for the two listener groups, however, showed this fear to be unwarranted. None of the differences between the mean ratings for any of the perceptual attribute pairs for the group thinking they were listening exclusively to boys and those thinking they were listening exclusively to girls reached significance, so the means shown in the table have been calculated over all listeners, regardless of the speaker gender they thought they were listening to.

The bolded figures in Table 2 mark those differences between male and female ratings for a perceptual attribute pair that were found to be significantly different. Individual t-tests were applied separately to the male and female ratings for each pair and a Bonferroni correction of α (.05) was used to correct for the number of comparisons being made ($.007 = .05/7$). Surprisingly, five of the seven attribute pairs were found to be significantly different. Put into words, the three male stimuli were perceived to be significantly slower, hoarser, less melodious, more mumbled and lower pitched.

3.3. Acoustic analyses

The righthand side of Table 2 summarises the results of the acoustic analyses for each of the stimuli used in the percep-

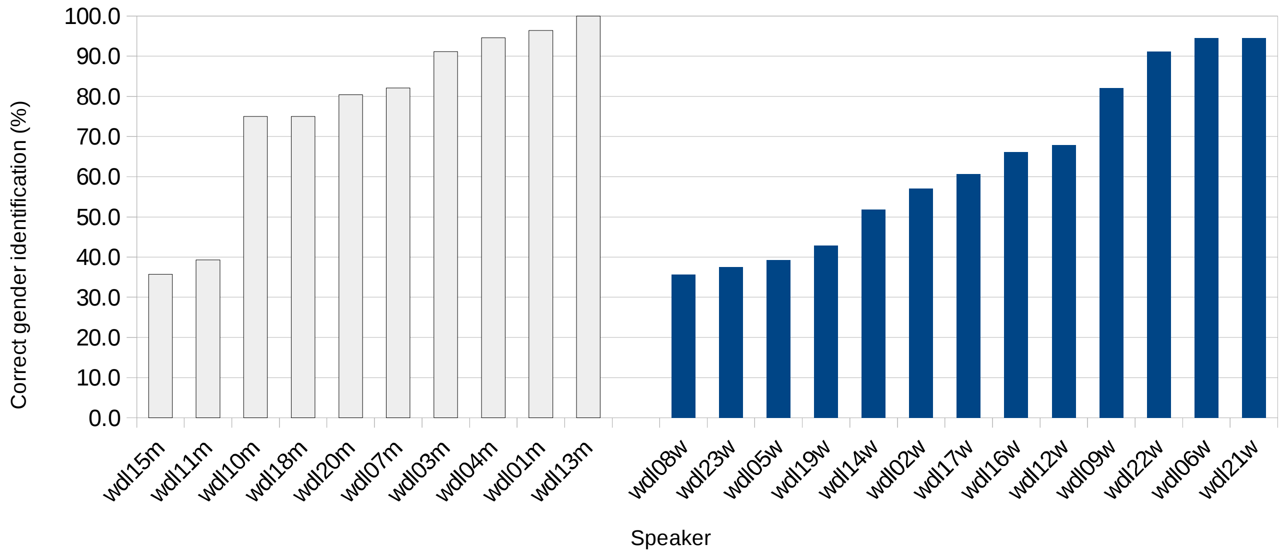


Figure 1: *Mean correct gender identification rate for each speaker (male, left; female, right). Speakers in each gender group are ranked according to the correct gender identification. The three most successfully identified speakers within each gender group were used to provide stimuli for the perceptual attributes experiment (male: wdl01m, wdl04m, wdl13m; female: wdl06w, wdl21w, wdl22w).*

Table 2: *Left: Mean perceptual attribute ratings for each of the most successfully identified male and female subjects. Right: Results of acoustic analysis of each stimulus.*

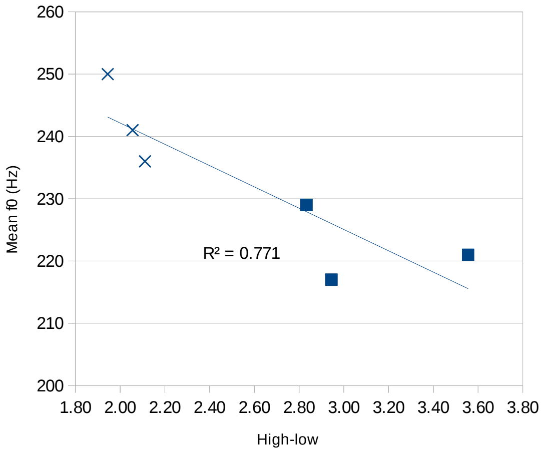
	Mean perceptual attribute pair ratings							Acoustic parameters				
	quiet –loud	fast –slow	clear –hoarse	monotonous –melodious	excited –calm	mumbled –clear	high –low	f0 M (Hz)	f0 CV (%)	HNR (dB)	T+ (syll/s)	T− (syll/s)
Male speakers												
wdl01m	2.67	2.61	2.28	2.33	3.39	3.28	2.95	217	9.1	7.13	3.74	4.00
wdl04m	2.22	3.50	2.56	1.73	3.39	3.00	2.84	229	15.0	6.51	2.39	2.82
wdl13m	2.73	2.39	2.84	1.95	2.84	2.84	3.56	221	6.2	9.37	4.19	4.19
Mean	2.54	**2.84**	**2.56**	**2.00**	3.21	**3.04**	**3.11**					
SD	0.66	0.77	0.96	0.73	0.86	0.82	0.82					
Female speakers												
wdl06w	2.84	1.89	1.73	3.00	2.78	3.45	2.06	241	13.6	12.06	4.73	4.73
wdl21w	2.78	2.22	1.39	2.84	3.45	3.89	1.95	250	9.6	13.38	4.68	4.68
wdl22w	2.67	2.89	1.44	2.38	3.50	4.00	2.11	236	8.9	14.08	3.76	3.76
Mean	2.77	**2.33**	**1.53**	**2.74**	3.23	**3.78**	**2.06**					
SD	0.58	0.80	0.64	0.85	0.82	0.50	0.61					

tual attribute listening experiment. The first two columns show mean fundamental frequency of each stimulus together with the coefficient of variation (*SD/M*) represented in percent. The third column shows the average harmonics-to-noise ratio calculated over the final 200 ms of voicing in each utterance. The final two columns show tempo in syllables/second, including (T+) and excluding (T−) the duration of any pauses. Means for each gender group have been omitted here, as we will concentrate on relating the individual acoustic values of each speaker-stimulus to the perceptual attribute ratings.

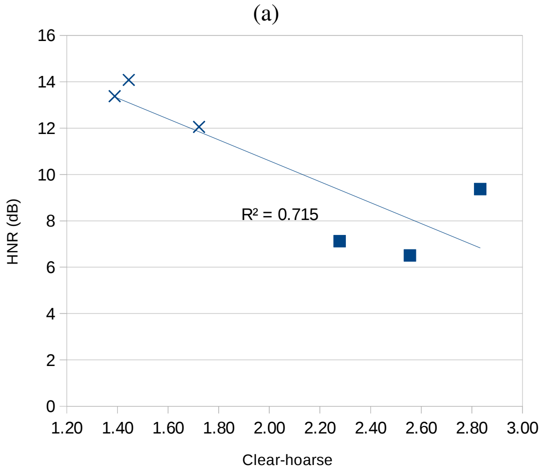
Despite the small number of data points in the acoustic analysis, there are a number of marked differences between the male and the female values. The female stimuli have higher average fundamental frequency values and lower harmonics-to-noise ratios with corresponding ratings for perceptual attributes *monotonous–melodious*, *high–low* and *clear–hoarse*. There are several highly significant correlations between perceptual attribute ratings and their corresponding acoustic parameters. Three of these, representing different phonetic aspects (pitch, voice quality and tempo), are visualised in Figure 2 with female (crosses) and male (squares) separated with different symbols. In Figure 2(a) mean fundamental frequency is plotted as a function of ratings for the perceptual attribute pair *high–low*, in (b) mean HNR is plotted as a function of *clear–hoarse*, and in (c) tempo excluding pauses is plotted as function of *fast–slow*. In each case, the linear component of the correlation accounts for a large proportion of the variance. Indeed, the strongest correlation is between perceived and measured tempo (excluding pause duration), with the linear component accounting for 96% of the variance ($R^2 = 0.963$). This value is reduced only marginally if *fast-slow* is correlated with a measure of tempo that includes pause duration ($R^2 = 0.937$). However, it is also interesting to note, that while perceived and measured tempo exhibit a very strong correlation, this is also the one case, where the male and female values are not discretely grouped.

4. Discussion

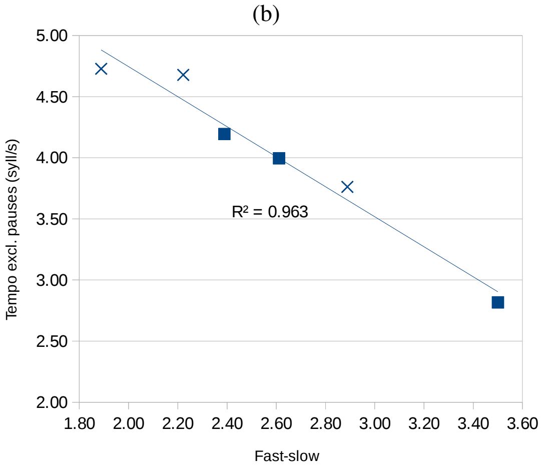
This study has investigated perceptual and acoustic correlates of the prepubertal voice somewhat differently from the path



(a)



(b)



(c)

Figure 2: *Acoustic parameters plotted as a function of perceptual attribute ratings: (a) "high–low" × mean f0; (b) "clear–hoarse" × mean HNR; (c) "fast–slow" × tempo excluding pauses.* R^2 *shows the amount of variation accounted for by the linear component of the correlation. Female values are plotted as crosses, male values as squares.*

usually chosen. When looking at acoustic differences between samples of male and female adult speakers, we typically measure a number of central acoustic parameters that characterise expected phonatory and articulatory differences, such as fundamental frequency, or formant frequencies. Studies examining gender differences in children's voices have often found relatively few differences between male and female speakers with respect to these core acoustic parameters. Our study has used the results of two successive listening experiments to inform the acoustic phonetic analysis. In the first experiment, listeners identified those voices that were most female- or male-sounding. The second experiment had listeners rate these voices against a number of perceptual attributes. Several significant differences between the ratings allowed us to look directly at temporal and spectral correlates that we might reasonably assume to be responsible for the perceptual ratings. We identified some strong relationships between perceptual attribute ratings and the corresponding acoustic parameters. These indicate several ways in which gender can be phonetically constructed in the prepubertal voice before sex-specific anatomical and physiological changes during puberty begin to play a role. At the same time, the results indicate the cues that adult listeners use and appear to require to make a correct gender assignment to a prepubertal voice.

Male-female differences in fundamental frequency (girl > boy) suggest that the subjects are already approximating the gender-specific differences of the adult voice. The significant differences we found are interesting because prepubertal gender-specific differences in f0 are generally absent [7]. Interestingly, however, our findings are in line with the prepubertal subjects in [26] who were asked to produce isolated words like a boy or a girl. While there was no f0-differences between the male and female voices speaking in their natural voice, boys significantly raised their f0 while imitating a girl's voice and girls significantly lowered their f0 when imitating a boy's voice. With other parameters, it is less apparent that children are approximating gender-specific adult patterns. The male stimuli were perceived to be significantly hoarser and exhibited correspondingly lower HNR values. This lines up well with a finding that has been replicated in English, Swedish and German samples, that there is a higher proportion of boys judged to be hoarse than girls [27, 28, 29]. By contrast, for adult voices, female voices have been found to be breathier [30, 31, 32]. Likewise, there is an indication in our results for the boys to have slower speech rates than the girls. Again, temporal measures of this in the adult population suggest that males have higher speech rates than females [33, 34]. So, while f0 differences seem to be approximating to adult-like patterns, this is less obviously the case for differences in voice quality and speech rate.

There are certain weaknesses in the present study that will be remedied in future work. Despite the strong correlations found between perceptual and acoustic categories, the speaker sample needs to be increased substantially, not least to increase the sample of male and female speakers with reliably identified gender. Furthermore, interspeaker comparison of acoustic parameters will be enhanced by using stimuli having the same verbal content, i.e. using the same sentence(s).

Finally, further investigations will need to include an auditory impressionistic assessment of certain features that are undoubtedly associated with the perception of clarity or tempo, such as the presence or absence of final plosive release, or the realisation of final unstressed /-ən/ as reduced [-n̩], or the unreduced [-ən]. However, a reliable comparison of these features also requires a larger speaker sample with a more homogeneous stimulus set.

5. References

[1] R. M. Krauss, R. Freyberg, and E. Morsella, “Inferring speakers’ physical attributes from their voices,” *Journal of Experimental Social Psychology*, vol. 38, no. 6, pp. 618–625, 2002.

[2] S. Bennett and B. Weinberg, “Sexual characteristics of preadolescent childrens’ voices,” *Journal of the Acoustical Society of America*, vol. 65, no. 1, pp. 179–189, 1979.

[3] S. Curtin and S. F. Kiesling, “Cues to gender in children’s speech,” *Journal of the Acoustical Society of America*, vol. 115, p. 2607, 2004.

[4] D. Günzburger, A. Bresser, and M. t. Keurs, “Voice identification of prepubertal boys and girls by normally sighted and visually handicapped subjects,” *Language and Speech*, vol. 30, pp. 47–58, 1987.

[5] I. Karlsson and M. Rothenberg, “Inter-cultural variations in gender-based language differences in young children,” *STL-QPSR*, vol. 1/1992, pp. 1–17, 1992.

[6] M. J. Nairn, “The perception of gender differences in the speech of 4½–5½ year old children,” in *Proc. XIIIth ICPhS*, vol. 2, Stockholm, 1995, pp. 302–305.

[7] T. L. Perry, R. N. Ohde, and D. H. Ashmead, “The acoustic bases for gender identification from children’s voices,” *Journal of the Acoustical Society of America*, vol. 109, no. 6, pp. 2988–2998, 2001.

[8] J. Sachs, P. Lieberman, and D. Erickson, “Anatomical and cultural determinants of male and female speech,” in *Language attitudes*, R. W. Shuy and R. W. Fasold, Eds. Washington, D.C.: Georgetown University Press, 1973, pp. 74–83.

[9] L. Koenig, “Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds,” *Journal of Speech, Language and Hearing Research*, vol. 43, pp. 1211–1228, 2000.

[10] F. Karlsson, E. Zetterholm, and K. P. H. Sullivan, “Development of a gender difference in voice onset time,” in *Proc. 10th Australian International Conference on Speech Science and Technology*, Sydney, 2004, pp. 316–321.

[11] S. P. Whiteside and J. Marshall, “Developmental trends in Voice Onset Time: some evidence for sex differences,” *Phonetica*, vol. 58, pp. 196–210, 2001.

[12] S. P. Whiteside, L. Henry, and R. Dobbin, “Sex differences in voice onset time: A developmental study of phonetic context effects in British English,” *Journal of the Acoustical Society of America*, vol. 116, pp. 1179–1183, 2004.

[13] S. Bennett, “Vowel formant frequency characteristics of preadolescent males and females,” *Journal of the Acoustical Society of America*, vol. 69, no. 1, pp. 231–239, 1981.

[14] P. A. Busby and G. L. Plant, “Formant frequency values of vowels produced by preadolescent boys and girls,” *Journal of the Acoustical Society of America*, vol. 97, pp. 2603–2607, 1995.

[15] S. Lee, A. Potamianos, and S. Narayanan, “Acoustics of children’s speech: Developmental changes of temporal and spectral parameters,” *105*, pp. 1455–1468, 1999.

[16] S. P. Whiteside, “Sex-specific fundamental and formant frequency patterns in a cross-sectional study,” *Journal of the Acoustical Society of America*, vol. 110, pp. 464–478, 2001.

[17] C. Klein, “Acoustic and perceptual gender characteristics in the voices of pre-adolescent children,” *Phonus (Forschungsbericht Institut für Phonetik, Universität des Saarlandes)*, vol. 9, pp. 221–329, 2005.

[18] P. F. McCormack and T. Knighton, “Gender differences in the speech patterns of two and a half year old children,” in *Proc. Sixth Australian International Conference on Speech Science and Technology*, Adelaide, 1996, pp. 337–342.

[19] C. T. Ferrand and R. L. Bloom, “Gender differences in children’s intonational patterns,” *Journal of Voice*, vol. 10, no. 3, pp. 284–291, 1996.

[20] M. P. Robb and J. O. Simmons, “Gender comparisons of children’s vocal fold contact behavior,” *Journal of the Acoustical Society of America*, vol. 88, pp. 1318–1322, 1990.

[21] E. King, “A roentgenographic study of pharyngeal growth,” *The Angle Orthodontist*, vol. 22, pp. 23–37, 1952.

[22] S. Bennett and B. Weinberg, “Acoustic correlates of perceived sexual identity in preadolescent childrens’ voices,” *Journal of the Acoustical Society of America*, vol. 66, no. 4, pp. 989–1000, 1979.

[23] P. Foulkes, G. J. Docherty, G. Khattab, and M. Yaeger-Dror, “Sound judgements: perception of indexical features in childrens speech,” in *A reader in sociophonetics*, D. R. Preston and N. Niedzielski, Eds. New York: De Gruyter Mouton, 2010, ch. 14, pp. 327–356.

[24] A.-S. Parent, G. Teilmann, A. Juul, N. E. Skakkebaek, J. Toppari, and J.-P. Bourguignon, “The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration,” *Endocrine Reviews*, vol. 24, no. 5, pp. 668–693, 2003.

[25] P. Boersma, “Praat, a system for doing phonetics by computer,” *Glot International*, no. 5:9/10, pp. 341–345, 2001.

[26] V. Cartei, W. Cowles, R. Banerjee, and D. Reby, “Control of voice gender in pre-pubertal children,” *British Journal of Developmental Psychology*, vol. 32, no. 1, pp. 100–106, 2014.

[27] E. Yairi, L. H. Currin, N. Bulian, and J. Yairi, “Incidence of hoarseness in school children over a 1 year period,” *Journal of Communication Disorders*, vol. 7, no. 4, pp. 321–328, 1974.

[28] E. Sederholm, “Prevalence of hoarseness in ten-year-old children,” *Scandinavian Journal of Logopedics and Phoniatrics*, vol. 20, no. 4, pp. 165–173, 1995.

[29] E. Kallvik, E. Lindström, S. Holmqvist, J. Lindman, and S. Simberg, “Prevalence of hoarseness in school-aged children,” *Journal of Voice*, vol. 29, no. 2, pp. 260–e1, 2015.

[30] C. G. Henton and R. A. W. Bladon, “Breathiness in normal female speech: Inefficiency versus desirability,” *Language and Communication*, vol. 5, pp. 221–227, 1985.

[31] M. Södersten and P.-Å. Lindestad, “Glottal closure and perceived breathiness during phonation in normally speaking subjects,” *Journal of Speech and Hearing Research*, vol. 33, no. 3, pp. 601–611, 1990.

[32] D. H. Klatt and L. C. Klatt, “Analysis, synthesis, and perception of voice quality variations among female and male talkers,” *Journal of the Acoustical Society of America*, vol. 87, pp. 820–857, 1990.

[33] D. Byrd, “Relations of sex and dialect to reduction,” *Speech Communication*, vol. 15, pp. 39–54, 1994.

[34] S. P. Whiteside, “Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences,” *Journal of the International Phonetic Association*, vol. 26, pp. 23–40, 1996.