PHONETIC CORRELATES OF GENDER IN PREPUBERTAL VOICES

Riccarda Funk[1], Melanie Weirich[2], Adrian P. Simpson[3]

[123]Friedrich Schiller University Jena, Germany
[1]riccarda.funk@uni-jena.de, [2]melanie.weirich@uni-jena.de, [3]adrian.simpson@uni-jena.de

ABSTRACT

This study investigates acoustic and perceptual correlates of gender in prepubertal voices. 60 German-speaking pupils (29 girls, 33 boys, aged 6-8 years) were recorded in first and again in second grade. A listening experiment was conducted to find out how female or male the voices are perceived to be. Stimuli of the five most female- and male-sounding children were played in a second experiment to two groups of listeners who judged each stimulus against six perceptual attribute pairs. Linear-mixed-models and Spearman's ρ correlations were run to evaluate the relationship between perception and acoustics.

There are significant acoustic differences between both genders and the most female- and male-sounding children. Significant correlations between perception ratings and acoustic parameters are reported. Furthermore, linear-mixed-models verify significant relationships between perception and acoustics within and across gender, indicating that certain phonetic parameters of the voice and speech are used by children to construct gender.

Keywords: Prepubertal voice, gender perception, acoustics, correlates of gender

1. INTRODUCTION

Many phonetic correlates of gender in the adult voice are due to anatomical differences in the larynx and the vocal tract. In prepubertal voices, these differences are marginal [1, 2]. Nevertheless, many studies have found acoustic differences between prepubertal boys and girls as young as two and a half years of age [3]. However, these differences are rarely systematic. Indeed, in some cases the results contradict each other:

- **f0:** girls > boys [4, 5], girls < boys [6], girls = boys [7]
- **F1:** girls > boys [8, 9, 10, 11]
- **HNR:** girls > boys [12, 13, 14, 15]
- **Tempo:** girls > boys [7, 15], girls < boys [7, 16]
- **/s/:** girls > boys in center of gravity (CoG) and girls < boys in skewness [17, 18, 19]

Despite such inconsistencies, other studies have shown that listeners are able to determine gender in children's voices at above-chance levels [14, 15, 20, 21, 22]. In most cases, the listeners' judgements coincide with biological sex. More interesting, however, are gender-identification rates of individual children. For some children identification rates are at chance level, whereas others receive a very clear gender assignment [14, 15]. The latter seem to exhibit a set of robust acoustic cues encoding gender that listeners overwhelmingly agree upon. When listeners rate stimuli of such children against a number of perceptual attribute pairs, e.g. fast–slow, high–low, strong correlations between perception ratings and acoustic measurements can be found [14].

The present study concentrates on identifying the acoustic and perceptual characteristics of children's voices which lead to a relatively unambiguous gender attribution by adult listeners. How do these voices differ from those with acoustic patterns that lead to more ambivalent, variable gender responses? Which acoustic parameters can be used to predict the perception of children's voices? Can we find relationships between perceptual attribute ratings and the corresponding acoustic parameters?

The study is part of a longitudinal project analysing recordings of ca. 60 German primary school children from the first to fourth grade (6-10 years). In this article, we focus on the results of the first and second graders.

2. METHOD

2.1 Listeners

Acoustic voice recordings were made of 62 German-speaking first-graders (29 girls, 33 boys) aged 6–7 years old ($\bar{x} = 6.16$). The recordings took place in September 2020 at two primary schools in neighbouring East German villages near Gotha. The same children were recorded again in October 2021. In this case, 60 children participated (28 girls, 32 boys), aged 7–8 years old ($\bar{x} = 7.18$). Two of the children could no longer participate due to relocation. In this study, we only analysed the data of the 60 children common to both recordings.

2.2 Recordings

All recordings were made using a USB microphone with integrated audio interface (Røde NT-USB) located in a portable sound wall (Marantz Pro Sound Shield) using *audacity* [23]. The microphone was positioned 20-30 cm from the speaker's mouth. The children were recorded individually in a quiet room at the school. Both spontaneous and read (copied) speech was elicited. Children were asked to repeat versions of ten simple prose sentences pre-recorded by an adult female speaker (first author). We recognise that children copying pre-recorded utterances may accommodate to aspects of the adult speaker's voice. However, given the level of (il)literacy in first graders, this was the only way to elicit tightly controlled sentence-length utterances. In this study, we focus on the recordings of the first two sentences:

1. *Im Sommer blühen die Blumen.*
   (Flowers blossom in summer.)
2. *Kannst du die Rose riechen*?
   (Can you smell the rose?)

2.3 Acoustic analysis

The stimuli of the children were segmented and annotated using *Web-MAUS* [24]. The *praat* [25] text grids were manually checked and corrected. All acoustic analyses were also performed in *praat* using different scripts.

Mean fundamental frequency was measured in both sentences and semitone range was calculated with the formula

$$(1) \quad st = \frac{12}{ln(2)} ln\left(\frac{fmax}{fmin}\right).$$

Mean formant values were estimated in the vowels /ɔ/, /yː/ and /uː/ of sentence 1 and /a/, /oː/ and /iː/ of sentence 2 (time step 0.01; window length 0.025 s; maximum formant 5500 Hz). A two-dimensional acoustic vowel space (AVS) was created using the F1 and F2 of these vowels. The area of this polygon was calculated and for legibility and ease of comparison will be stated in kHz^2. The same vowels were used for determining voice quality (HNR). Tempo was calculated in canonical syllables per second (without pauses). center of gravity (CoG) and skewness were calculated from the sibilant spectra of /z/ in “Sommer” (sentence 1) and “Rose” (sentence 2) and /s/ in “kannst (sentence 2).

2.4 Listening experiment 1 – gender identification

The aim of the first listening experiment was to find out how female or male the voices of the children are perceived to be, i.e. which children produce robust phonetic correlates of gender. For this, listeners heard the recordings of sentences 1 and 2. The stimuli were normalised to 70 dB.

The listening experiment was conducted online using *Percy* [26] – one experiment for the voices of the first graders and another one for the second-graders. Stimuli were presented in random order. Each stimulus could be listened to twice. Listeners decided on a seven-point scale how boy-like or girl-like a stimulus was. The duration of the experiment for each subject was shortened by dividing the stimuli into three approximately equally sized groups.

167 listeners participated in the experiment with first-graders (119 female, 46 male, 2 diverse; age: 16–71 years, $\bar{x} = 33.2$, $\sigma = 13.7$). 117 different subjects listened to the voices of the second-graders (94 female, 21 male, 2 diverse; age: 16-70 years, $\bar{x} = 33.2$, $\sigma = 13.4$). None of the listeners reported any hearing problems and the mother tongue of all was German.

2.5 Listening experiment 2 – perceptual attribute rating

The second listening experiment explored possible relationships between perceptual attribute ratings and the corresponding acoustic parameters. Stimuli of the five girls and five boys with the highest and lowest perception ratings (see 3.1) were used. Listeners judged perceptual attribute pairs for sentences 1 (recording from first grade) and 2 (recording from second grade) of each child on a seven-point-scale: fast–slow, high–low, clear–mumbled, hoarse–clear, loud–quiet, monotonous–melodious. The stimuli were normalised to 70 dB using *praat* [25] and the experiment was conducted using *SoSci survey* [27].

To avoid a possible influence of gender stereotypes on the perception ratings, the listeners were split into two approximately equally sized groups. As in [14], both groups heard the same stimuli, but the first group was told they were listening to boys, the second that they were listening to girls. 102 listeners took part in this listening experiment (80 female, 21 male, 1 diverse; 16–77; $\bar{x} = 34.8$, $\sigma = 12.8$). None of the participants reported any hearing problems and the mother tongue of all was German.

3. RESULTS

3.1 Gender identification

In listening experiment 1, the gender identification score of the individual children is between 1.55 and 6.56 on the seven-point scale (mean of both sentences and grade levels). As expected, boys and girls differ significantly in a Wilcoxon rank-sum test (girls: $\bar{x} = 5.09$, $\sigma = 1.01$; boys: $\bar{x} = 3.04$, $\sigma = 1.18$; $W = 85$, $p < .01$*). When we take a score less than 2 or greater than 6 as a sign of systematic gender assignment, 15 children are producing phonetic patterns that receive unambiguous gender attribution by the listeners. The remaining 45 children are producing acoustic patterns that lead to more ambivalent gender responses. Therefore it is not surprising that

	all (N = 60)							selected (N = 10)						
	girls		boys					females		males				
	x̄	σ	x̄	σ	p	W	D	x̄	σ	x̄	σ	p	W	d
F0 (Hz)	250	21	249	24	1.0	6845	.08	256	22	232	16	.01*	78	1.0
Semitones	6.82	2.58	6.66	1.91	.99	7170	.01	7.10	1.81	6.42	1.67	1.0	162	.33
AVS (kHz^2)	1.44	1.10	1.30	0.94	1.0	6853	.08	1.64	0.99	1.67	1.19	.99	201	.01
HNR (dB)	19.10	3.01	20.10	3.02	.04*	8620	.35	18.74	2.75	18.45	3.77	1.0	196	.03
Tempo (syll/s)	3.92	0.58	4.15	0.57	.01*	8914	.42	4.04	0.48	4.08	0.65	1.0	196	.03
CoG (Hz)	4198	1987	3599	2013	.10	5909	.30	4209	2381	3699	2036	1.0	172	.23
Skewness	0.59	1.49	1.14	2.14	.08	8408	.30	0.64	1.35	1.33	2.06	1.0	245	.38

Table 1: Arithmetic means and standard deviations of acoustic parameters (girls and boys / selected female- and male-sounding children). Wilcoxon rank-sum tests (p-values Bonferroni-Holm-corrected) and Cohen's d on the right.

the intraclass correlation coefficient (ICC) including the ratings of all children shows a poor inter-rater reliability ($ICC\ (1, 1) = .50, p < .01^*$), while the inter-rater reliability of the other 15 children is quite good ($ICC\ (1, 1) = .82, p < .01^*$).

For listening experiment 2 and the acoustic analysis, we choose the five most female- and male-sounding children (mean of both grades). In this case, perception coincides with biological gender.

3.2 Acoustic analysis

Table 1 shows the acoustic differences between girls and boys (mean of both sentences and both grade levels). When all 60 children are included, significant differences in the Wilcoxon rank-sum tests can be found in HNR and speech tempo with higher speech rates and HNR values for boys. Cohen's d was also calculated to evaluate the effect sizes. Here we can assume medium effects of gender for HNR, speech tempo, CoG and skewness. In line with earlier studies [17, 18, 19], girls have higher frequencies of CoG and lower values for skewness than boys. However, the standard deviations of both parameters are quite high, which suggests a high interindividual variability. By contrast to other studies [12, 13, 14, 15], HNR values for the boys are higher than for the girls indicating a higher modality in the boys' voices. The girls have a slightly larger AVS than the boys; the fundamental frequency of both genders is almost the same.

Interestingly, if we only compare the most female- and male sounding children (selected from listening experiment 1, see 3.1), only the fundamental frequency differs significantly between both groups with a large effect – the female children speak with higher fundamental frequency than the male children. The semitone range of the females is also higher than it is for the males (medium effect). AVS, HNR and speech tempo are almost the same. As in the whole group, the girls show higher values for CoG and lower values for skewness than the male children, with a small effect for CoG and a medium effect for skewness.

3.3 Acoustic analysis and perception

To find out which acoustic parameters predict the perception of the children's voices, linear-mixed-models were run using the *lme4 package* [28] in *R* [29] for all children and separately for girls and boys. Perception was the dependent variable acoustic parameters of sentence 1 and 2 were tested as fixed factors. We entered a random intercept for each child. P-values were obtained using Likelihood ratio tests comparing the model with the acoustic parameters with the model without them. Interactions between each acoustic parameter and grade level or sentence have been calculated.

When all children are tested, the model is best when skewness is added as fixed factor ($\chi^2\ (1) = 4.26, p = .02^*$). There is no interaction with grade level ($\chi^2\ (2) = 2.69, p = .26$) or sentence ($\chi^2\ (2) = 1.51, p = .47$). For the boys, we get the best result when we only add f0 as fixed factor ($\chi^2\ (1) = 5.36, p = .02^*$). There is no interaction with grade level ($\chi^2\ (2) = 4.10, p = .13$) or sentence ($\chi^2\ (2) = 2.87, p = .24$). For the girls, the best result can be achieved when AVS ($\chi^2\ (1) = 4.07, p = .04^*$), HNR ($\chi^2\ (1) = 6.92, p = .01^*$) and tempo ($\chi^2\ (1) = 4.35, p = .04^*$) are included as fixed factors. Interactions can be found between grade level and AVS ($\chi^2\ (2) = 7.74, p = .02^*$) and grade level and tempo ($\chi^2\ (2) = 6.82, p = .02^*$), whereby the influence of AVS on gender perception is greater in the first grade and tempo in the second grade. There is no interaction between grade level and HNR ($\chi^2\ (2) = 5.90, p = .05$). For sentence, interactions can be found with AVS ($\chi^2\ (2) = 8.82, p = .01^*$) and tempo, ($\chi^2\ (2) = 9.58, p = .01^*$). The influence of AVS on gender perception is greater in sentence 1, whereas tempo has a greater effect in sentence 2. There is no interaction between sentence and HNR ($\chi^2\ (2) = 3.93, p = .14$).

3.4 Perceptual attribute ratings

The perceptual attributes high–low (pitch), monotonous–melodious (range), fast–slow (tempo), hoarse–clear (hoarseness), loud–quiet (volume) and clear–mumbled (precision) were assessed on a seven-point-scale. The first term has the value 1 (e.g.

"high"), the second the value 7 (e.g. "low"). If we compare these perception ratings (mean of both stimuli) of the most female- and male-sounding children, significant differences can be found for pitch, range and precision, which means that the female children sound higher, more melodic and seem to have a more precise articulation than the males (see table 2). In these parameters, large effects can be found. Furthermore, there is a large effect of hoarseness with males being perceived as hoarser than females.

	female		male				
	x̄	σ	x̄	σ	p	t	d
Pitch	2.98	0.58	5.10	0.30	.001*	9.53	4.3
Range	4.94	0.29	3.56	0.37	.001*	-4.40	2.0
Tempo	4.06	0.56	3.99	0.86	.87	-0.17	.08
Hoarsen.	5.11	0.54	3.49	0.63	.27	-5.03	2.2
Volume	3.76	0.33	4.03	0.49	.46	1.25	.56
Precis.	2.85	0.40	4.47	0.64	.001*	4.57	2.0

Table 2: Arithmetic means and standard deviations of attribute ratings. T-tests (p-values Bonferroni-Holm-corrected) and Cohen's d on the right.

It makes no significant difference whether listeners thought they were listening to boys or girls ($p > .40$ in all cases).

Spearman's ρ correlations between attribute ratings and corresponding acoustic parameters (mean of both stimuli) were run. Perceived precision was correlated with acoustical vowel space and tempo. There are strong significant correlations between perceived pitch and f0 ($\rho = -.84, p < .01$*) and between perceived tempo and measured tempo ($\rho = -.83, p < .01$*): With increasing f0 and tempo, the children are perceived as higher pitched and faster speaking (see figure 1). No significant correlations between semitone range and perceived range ($\rho = .27, p = .45$) or between HNR and perceived hoarseness ($\rho = .53, p = .12$) can be found. For precision, there is neither a significant correlation with AVS ($\rho = -.27, p = .44$) nor with tempo ($\rho = .25, p = .48$).

4. DISCUSSION

Studies have shown that the gender perception of boys and girls differs significantly in a listening experiment. Similar to [14, 15], 15 of the 60 children seem to produce unequivocal acoustic patterns and achieve very high or low perception values. When all 60 children are compared to each other, significant differences between girls and boys can be found in HNR and tempo, with girls having higher levels of noise and lower tempo. This is contrary to the results of [14]. It is possible that the girls of the present study were shyer and as a result used quieter, breathier voice qualities than the boys of the same age. In [14], the sentences were read by the children. Therefore it is probable that reading competence rather than speech tempo was reflected there.

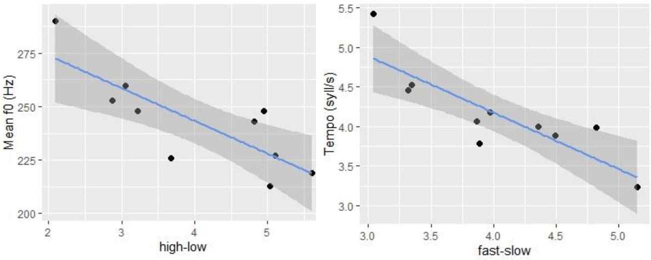

Figure 1: Acoustic parameters plotted as a function of perceptual attribute ratings. Left: "high–low" and mean f0; Right: "fast–slow" and tempo.

When we only compare the five most female- and male-sounding children, only f0 differs significantly. This would suggest that f0 plays a strong role in gender perception. Linear-mixed-models confirm this for boys, whereas skewness is also decisive for perception across gender. For girls, AVS, HNR and tempo could be crucial. Interactions between AVS, tempo and grade level or sentence show that in different stimuli, different parameters could be used for gender construction and perception in the female group.

A listening experiment with perceptual attribute ratings shows significantly higher ratings of pitch, pitch range and precision for the female-sounding children. In addition, there are strong correlations between perception and corresponding acoustic parameters, especially for pitch and speech tempo. In other words, listeners are able to recognize acoustic parameters associated with gender perception.

From these results we can conclude that there is a range of acoustic parameters that influence gender perception. Furthermore, it seems that a particular high or low gender rating does not rely on a single acoustic parameter. In further listening experiments using sophisticated voice morphing techniques [30] we will be able to examine in more detail the role played by individual acoustic parameters. In addition, due to the longitudinal nature of the project from which this study is a part, we will be able to analyse how the vocal expression of gender in individual children develops over the first years in primary education before reaching puberty.

5. ACKNOWLEDGMENTS

The project is supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG grant SI 743/9-1, http://www.dfg.de/). We would like to thank Christoph Draxler (MLU Munich) for his help in constructing and creating the first listening experiment. Most of all we are very grateful to the children and their parents agreeing to take part in this project.

6. REFERENCES

[1] Kahane, J. C. 1978. A morphological study of the human prepubertal and pubertal larynx. Journal of Anatomy 151, 11–20.
[2] Fitch, W. Tecumseh & J. Giedd. 1999. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. Journal of the Acoustical Society of America 106(3), 1511–1522.
[3] McCormack, P. F. & T. Knighton. 1996. 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, 337–342.
[4] Hasek, C. S., S. Singh, & T. Murry. 1980. Acoustic attributes of children's voices. Journal of the Acoustical Society of America 68, 1262–1265.
[5] Ferrand, C.T. & R. L. Bloom. 1996. Gender differences in children's intonational patterns. Journal of Voice 10(3), 284–29.
[6] Glaze, L. E., D. M. Bless, P. Milenkovic, & R. D. Susser. 1988. Acoustic characteristics of children's voice. Journal of Voice 2(4), 312–319.
[7] Whiteside, S. P. & C. Hodgson. 2000. Speech patterns of children and adults elicited via a picture-naming task: An acoustic study. Speech Communication 32, 267–285.
[8] Bennett, S.. 1981. Vowel formant frequency characteristics of preadolescent males and females. Journal of the Acoustical Society of America 69(1), 231–239.
[9] Busby, P. A. & G. L. Plant. 1995. Formant frequency values of vowels produced by preadolescent boys and girls. Journal of the Acoustical Society of America 97, 2603–2607.
[10] Pettinato, M., O.Tuomainen, S. Granlund, & V. Hazan. 2016. Vowel space area in later childhood and adolescence. Effects of age, sex and ease of communication. Journal of Phonetics 54, 1–14.
[11] Whiteside, S. P. 2001. Sex-specific fundamental and formant frequency patterns in a cross-sectional study. Journal of the Acoustical Society of America 110, 464–478.
[12] Ferrand, C. T. 2000. Harmonics-to-noise ratios in normally speaking prepubescent girls and boys. Journal of Voice 14(1), 17–21.
[13] Kallvik, E., E. Lindström, S. Holmqvist, J. Lindman, & S. Simberg. 2015. Prevalence of hoarseness in school-aged children. Journal of Voice 29(2), 260–e1.
[14] Simpson, A. P., R. Funk, & F. Palmer. 2017. Perceptual and acoustic correlates of gender in the prepubertal voice. In Interspeech 2017. Stockholm, 914–918.
[15] Funk, R., S.Voigt-Zimmermann, & A. P. Simpson. 2018. Junge oder Mädchen? Zur Geschlechtsidentifikation präpubertärer Stimmen. In Proc. Phonetik und Phonologie 14. Vienna.
[16] Walker, J. F., L. M. D. Archibald, S. R. Cherniak, & V. G. Fish. 1992. Articulation rate in 3-and 5-year-old children. Journal of Speech, Language and Hearing Research 35(1), 4–13.
[17] Li, F., D. Rendall, P. L. Vasey, M. Kinsman, A. Ward-Sutherland, & G. Diano. 2016. The development of sex/gender-specific /s/ and its relationship to gender identity in children and adolescents. Journal of Phonetics 57, 59–70.
[18] Shadle, C. H., L. L. Koenig, & J.L. Preston. 2011. Acoustic characterization of /s/ spectra of adolescents: moving beyond moments. In Proceedings of Meetings on Acoustics 12. Seattle.
[19] Flipsen, P., L. Shriberg, G. Weismer, H. Karlsson, & J. McSweeny. 1999. Acoustic characteristics of /s/ in adolescents. Journal of Speech, Language, and Hearing Research 42, 663–677.
[20] Günzburger, D., A. Bresser, & M. ter Keurs. 1987. Voice identification of prepubertal boys and girls by normally sighted and visually handicapped subjects. Language and Speech 30, 47–58.
[21] Karlsson, I. 1987. Sex differentiation cues in the voices of young children of different language background. Journal of the Acoustical Society of America 81, 68–69.
[22] Kaya, H., A. A. Salahb, A. Karpovc, O. Frolovae, A. Grigoreve, & E. Laykso. 2017. Emotion, age, and gender classification in children's speech by humans and machines. Computer Speech and Language 46, 268–283.
[23] Audacity Team. 2021. Audacity(R): Free Audio Editor and Recorder. Version 3.0.0 retrieved March 17th 2021 from https://audacityteam.org/.
[24] Kisler, T., Reichel U. D., & F. Schiel. 2017. Multilingual processing of speech via web services. Computer Speech & Language 45, 326–347.
[25] Boersma, Paul & David Weenink. 2018. Praat, doing phonetics by computer. http://www.praat.org, version 6.0.40.
[26] Draxler, Christoph. 2014. Online experiments with the Percy software framework – experiences and some early results. Reykjavik, Iceland, 235–240.
[27] Leiner, D. J. (2019). SoSci Survey. Version 3.1.06. Available at https://www.soscisurvey.de
[28] Bates D, Maechler M, Bolker B & Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4, Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.
[29] R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.
[30] Kawahara, Hideki, M. Morise, Toru Takahashi, R. Nisimura, T. Irino, & Hideki Banno. 2008. Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 3933–3936.