INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Developmental articulatory and acoustic features for six to ten year old children

Vishwas M. Shetty[1], *Steven M. Lulich*[2], *Abeer Alwan*[1]

[1]Electrical and Computer Engineering, University of California, Los Angeles, California, USA
[2]Speech Language and Hearing Sciences, Indiana University, Bloomington, Indiana, USA

shettyvishwas@ucla.edu, slulich@indiana.edu, alwan@ee.ucla.edu

Abstract

In this paper, we study speech development in children using longitudinal acoustic and articulatory data. Data were collected yearly from grade 1 to grade 4 from four female and four male children. We analyze acoustic and articulatory properties of four corner vowels: /æ/, /i/, /u/, and /ɑ/, each occurring in two different words (different surrounding contexts). Acoustic features include formant frequencies and subglottal resonances (SGRs). Articulatory features include tongue curvature degree (TCD) and tongue curvature position (TCP). Based on the analyses, we observe the emergence of sex-based differences starting from grade 2. Similar to adults, the SGRs divide the vowel space into high, low, front, and back regions at least as early as grade 2. On average, TCD is correlated with vowel height and TCP with vowel frontness. Children in our study used varied articulatory configurations to achieve similar acoustic targets.
Index Terms: speech development, children's speech, subglottal resonances, acoustic-articulatory analysis.

1. Introduction

The elementary school years form a critical period in speech and language development. As children's speech utterances become longer and more complex, their speech becomes more adult-like, and they develop basic literacy skills. At the same time, language, phonological, and articulation disorders are difficult to differentiate from each other and from normal developmental variation until this period, when the window for early intervention is rapidly closing and therapy effectiveness is diminishing. Moreover, relevant speech-based technologies such as mispronunciation detectors [1], Automatic Speaker Verification (ASV) systems [2], and Automatic Speaker Recognition (ASR) systems [3] perform less well with younger children because of a general reliance on adult acoustic data or acoustic models trained from adult speech.

Although a number of important studies have investigated speech production by school-age children, most of these studies have been cross-sectional in design [4, 5, 6, 7, 8, 9]. Longitudinal information about the development of vowel formant frequencies [10, 11], vocal tract anatomy [12], and articulation among children in elementary school is needed to characterize patterns of individual variation in developmental trajectories, as well as the timing of developmental milestones (c.f. [13], regarding longitudinal voice changes in adolescent boys).

For example, sex-based differences in acoustics have been reported to appear by four years of age [9, 14], although differences in the vocal tract length configuration have been observed only from the onset of puberty [15]. Articulatory development may provide the 'missing link' that mediates between the absence of observed anatomical dimorphism and the presence of acoustic dimorphism before puberty.

In addition to vowel formants, the subglottal resonances (SGRs) of children have been studied. Longitudinal evidence from 2- and 3-year-old children suggests that the development of vowel formants interacts with the subglottal resonances (SGRs)[16] in accordance with the quantal theory of speech production [17, 18, 19, 7], such that the SGRs divide the vowel space into high, low, front, and back regions. Although SGRs from several children in the elementary school period have previously been reported [7, 20], these studies have been cross-sectional in design, and are therefore not able to reveal either individual developmental trajectories for SGRs or potential developmental milestones in the pre-pubertal years.

We are unaware of any longitudinal studies of vowel articulation in elementary school-aged children, other than studies of biofeedback in speech therapy, e.g. [21]. This study presents data from a longitudinal investigation of vowel formant frequencies and subglottal resonances, along with lingual articulatory measurements for speech data from children in grade 1 through grade 4 (six to ten years of age).

2. Methods

Data were analysed for 4 male and 4 female native speakers of Midwestern American English, between grades 1 and 4. All children had Goldman and Fristoe Test of Articulation, 3rd Edition (GFTA-3) standardized scores within the normal range (> 85) and were identified as typically developing. In addition to speech utterances, age and gender were recorded. Ultrasound images were recorded concurrently with microphone and neck accelerometer signals. In this study, acoustic analysis was performed using the microphone and accelerometer utterances, and articulatory analysis was performed using the ultrasound images. Speech utterances included the following pairs of words "Apple"/"Vacuum" (/æ/), "Teeth"/"Zebra" (/i/), "Shoe"/"Zoo" (/u/) and, "Watch"/"Frog" (/ɑ/), i.e., two words for each corner vowel. The articulatory features were analyzed by extracting the midsagittal tongue contour from ultrasound images that were time-aligned with the corresonding microphone and accelerometer signals. The participants were recorded once each year for four years, beginning in 1st grade (ages 6;4-7;3), until 4th grade (ages 9;5-10;3).

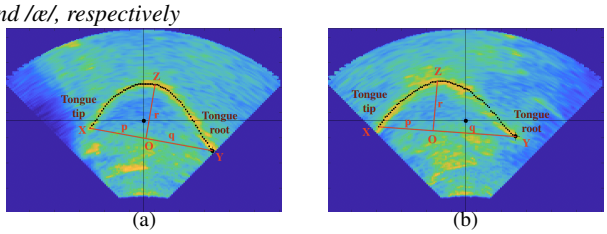
Speech utterances were elicited by a speech-language pathology student-clinician, using a picture naming task commonly used by speech-language pathologists (GFTA-3). Participants were seated in a chair in a double-walled sound booth. A SHURE KSM30 microphone on a stand positioned in front of the participant recorded the speech, while a K&K Sound HotSpot accelerometer held against the skin of the neck below the thyroid cartilage recorded subglottal acoustics. A Philips x6-1 3D/4D digital ultrasound transducer was held under each

10.21437/Interspeech.2023-2236

participant's chin to record the tongue movement with a Philips EPIQ 7G ultrasound system.

For each of the eight children analyzed in this study, the first two formant frequencies (F1, F2) for every corner vowel were measured manually from the microphone signal using Praat. Measurements were made every year between 1st and 4th grade. Therefore, there were 32 measurements of F1 and F2 per child, i.e., a total of 512 measurements. The first two subglottal resonances (Sg1, Sg2) were measured manually from the accelerometer signal for each child in each grade, yielding 32 measurements of Sg1 and Sg2 since SGRs are vowel-independent [22].

Figure 1: *Midsagittal profile of the tongue during the production of the vowel /u/ in word "Zoo" and /æ/ in "Vacuum" by a boy in 1st Grade. (a) and (b) are the Ultrasound images for /u/ and /æ/, respectively*



(a) (b)

Articulatory features were quantified using the Tongue Curvature Degree (TCD) and Tongue Curvature Position (TCP) measures [23]. These measures are relatively insensitive to variations in ultrasound probe placement, and have commonly been used (with other similar measures) to characterize the shape of the midsagittal tongue contour [24].

The midsagittal profiles from the ultrasound images of the tongue during the production of two vowels, /u/ and /æ/, are shown in Figs. 1a and 1b, respectively. We manually extract the midsagittal tongue contour, as highlighted in black dotted lines in Figs. 1a and 1b. Points X, Y, Z, and O are marked on the midsagittal contours, as shown in Fig. 1a and 1b. Here, X and Y are located at the tip and root of the tongue, respectively. Z is the point on the tongue contour that is at the maximum perpendicular distance from the line XY. We calculate the TCD and TCP values as follows:

$$TCD = \frac{len(ZO)}{len(XO)} = \frac{r}{p} \qquad TCP = \frac{len(YO)}{len(XO)} = \frac{q}{p}$$

We make similar measurements for all children for all four vowels. Hence, we have 32 measurements (4 vowels $*$ 2 words/vowel $*4$ grades $= 32$) of TCD and TCP per child.

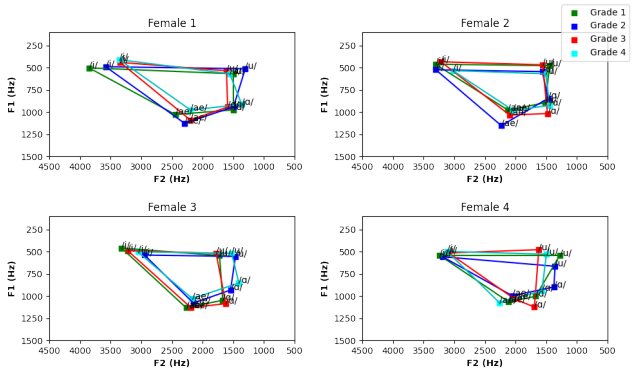
3. Results and Discussion

3.1. Acoustic Analysis

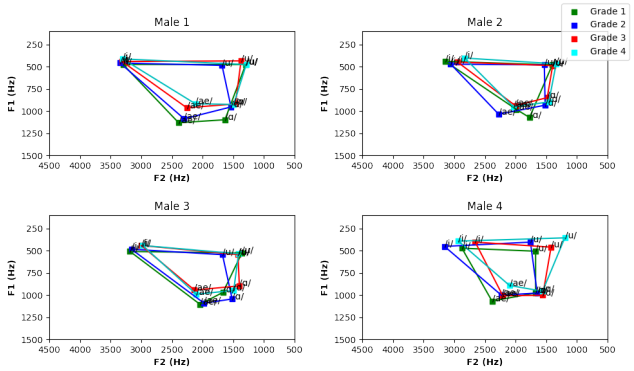
Given in Figs. 2a and 2b are the vowel plots for females and males, respectively. Each subplot represents a child. We use different colors in the subplot to represent vowel plots for different grades. F1 and F2 form the y-axis and x-axis, respectively.

As mentioned in Section 2, each vowel is analyzed in two different words. Hence, in a given grade, we have two values each of formant frequencies F1 and F2 for each vowel analyzed. We have taken the average F1 and F2 values to plot the per child, grade-wise vowel quadrilateral in Figs. 2a and 2b. In Fig. 2c, we plot an average gender-specific F1-F2 plot by averaging F1 and F2 values per gender in a given grade. Hence, Fig. 2c represents a cross-sectional analysis of formant frequency development. For most children, with every passing grade, the vowel plot appears to move up towards the right. This indicates a gradual reduction in formant frequency values. However, the longitudinal analysis reveals that few children, e.g. "Female 3" and "Male 4", slightly deviate from this behavior.

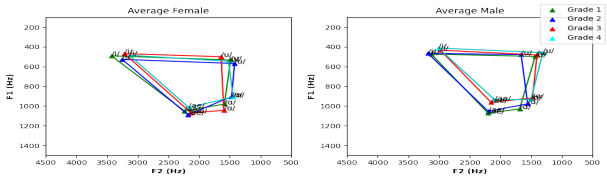
Figure 2: *Vowel plots highlighting the changes in the first two formant frequencies (F1 and F2) in children over the four grades. F1 and F2 formant frequencies form the y-axis and x-axis, respectively. Averaged F1-F2 plot is given in subplot (c).*



(a) *Grade-wise F1-F2 plots for four vowels for 4 females*



(b) *Grade-wise F1-F2 plots for four vowels for 4 males*



(c) *Grade-wise averaged F1-F2 plots for four vowels per gender*

The gender-wise vowel plots for each grade are given in Fig. 3. Here, for a given grade, we average the formant frequency values for all children of one gender to get one set of values for a vowel. Using these average formant frequency values, we plot separate vowel plots for each gender per grade. Each subplot represents a grade. The averaged vowel plots for the two genders are shown in different colors.

From Fig. 3, the averaged vowel plots for both genders overlap in grade 1. Hence, there aren't many gender-based differences in the acoustics in grade 1. However, from year 2 onwards, we see a slight upward shifts in the male vowel plots, indicating slightly lower formant frequency values for males compared to females. These results align with the observations made in [9]. Hence, when the acoustic parameters are averaged across children belonging to the same gender, we observe the emergence of gender-based differences from grade 2.

Plots illustrating the development of the first (Sg1) and second (Sg2) subglottal resonance (SGR) is given in Fig. 4. We

Figure 3: *Grade wise average vowel plots highlighting gender based differences*

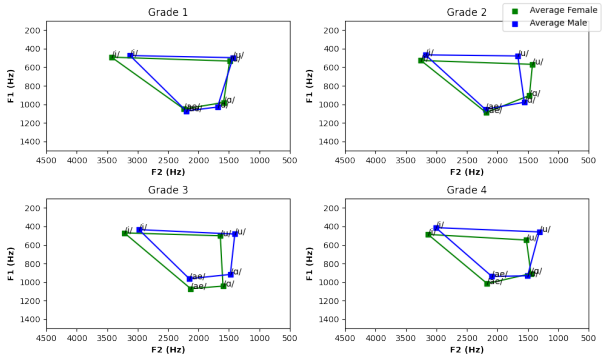
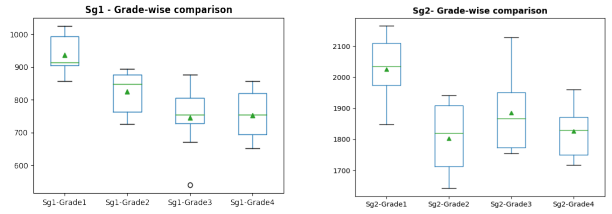


Figure 4: *Grade-wise Sg1 and Sg2 plots for all children in this study. The mean value is represented by a green triangle marker within each boxplot.*



(a) *Grade-wise Sg1 for all children* (b) *Grade-wise Sg2 for all children*

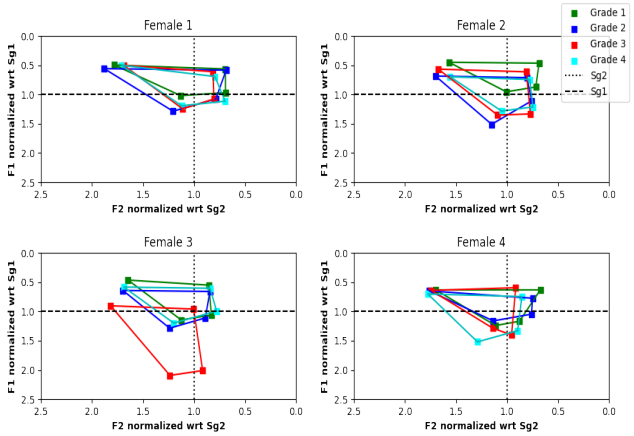
observe an overall reduction in the Sg1 and Sg2 values over the four grades. Plots with Sg1 and Sg2 values overlayed on the vowel plots for each child are shown in Fig. 5. In Fig. 5, for a given child, the F1 and F2 values have been normalized using Sg1 and Sg2 values respectively. Each subplot represents a child with dotted horizontal and vertical lines representing the normalized Sg1 and Sg2 values, respectively.

The effect of SGRs on child vowel spaces is similar to their effect on adult acoustics [25, 26, 19, 27]. We observe that, with age, the SGRs become slightly better boundaries separating the vowels, beginning mostly around grades 2 and 3.

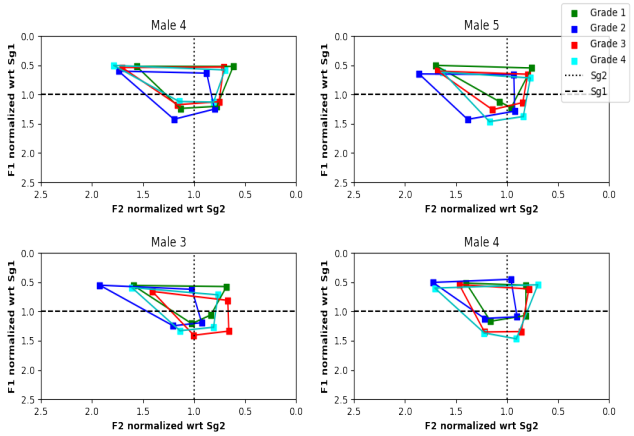
3.2. Articulatory Analysis

Following [23], we hypothesized that larger values of TCD might be associated with high vowels, which require raising the tongue body toward the hard or soft palate. We also hypothesized that larger values of TCP might be associated with front vowels, which require a fronting of the tongue body. Fig. 1 illustrates these associations for two vowels. For the low front vowel /æ/ the tongue curves nearer the tongue tip and the degree of curvature is smaller, while for the high back vowel /u/ the tongue curves nearer the tongue root and the degree of curvature is greater. These two hypotheses were tested by comparing the TCD measures in high vs. low vowels (Fig. 6a), and the TCP measures in front vs. back vowels (Fig. 6b). "TCD_High" represents TCD values for both high vowels /i/ and /u/ over all four years for all children. Hence, there were 128 measurements (2 high vowels $*2$ words per vowel $*8$ children $*4$ years) involved in "TCD_High" box plot. Similary, we plot "TCD_Low" in Fig. 6a and "TCP_Front" and "TCP_Back" in Fig. 6b. The mean values are represented by green triangle markers within each boxplot. The "p-value" obtained from 2-sample, 1-tailed t-tests are given in red font within Figs. 6a and 6b. The mean TCD for high vowels was found to be larger than the mean TCD for low vowels, and the mean TCP for front vowels was found to be larger than the mean TCP for back vowels, as hypothesized.

Figure 5: *Grade-wise vowel plots for all children in this study. F1 and F2 values are normalized with respect to the first and second subglottal resonances, respectively. Dotted horizontal and vertical lines in every subplot indicate normalized subglottal resonance frequencies. Normalized F1 and F2 values form the y-axis and x-axis, respectively for each subplot.*



(a) *Grade-wise SGR normalized F1-F2 plots for four females*



(b) *Grade-wise SGR normalized F1-F2 plots for four males*

Since TCD and TCP are correlated with vowel height and vowel frontness, we further hypothesized that the vowel articulatory space using the TCD and TCP features might look similar to the vowel acoustic space using F1 and F2. The vowel articulatory spaces are plotted in Fig. 7. TCD is plotted on the y-axis and TCP is plotted on the x-axis. Although the articulatory space in some instances (e.g. "Male 2" grade 2) bears a rough resemblance to the acoustic space, in most cases this resemblance is not apparent. It is noteworthy, however, that high vowels almost always have larger TCD values than low vowels. This indicates that, although TCD and TCP on the whole are correlated with vowel height and frontness, the vowel- and child-specific correlations are not straightforward, especially for TCP. This could be due to the fact that jaw movement (related to vowel height) and lip rounding (correlated with vowel backness) were ignored.

The fact that the relationship between vowel height and TCD was more straightforward than the relationship between vowel backness and TCP might be due to the fact that raising the tongue body (i.e. increasing TCD) creates a vocal tract Helmholtz resonator associated with a low target F1 frequency [28]. In contrast, the target F2 frequency of each vowel is not necessarily uniformly affiliated with either the "front (oral) cav-

Figure 6: *Box plots comparing articulatory parameters TCD and TCP. Within each subplot, in red is given the results of either 2 sample, 1-tailed t-test ($\alpha = 0.05$) or single factor ANOVA, comparing the parameters within the respective subplots. For example in subplot (a), the outcome of the t-test between High and Low vowels for the TCD parameter was $t = 1.651; p = 0.0001$. The mean value is represented by a green triangle marker within each boxplot.*

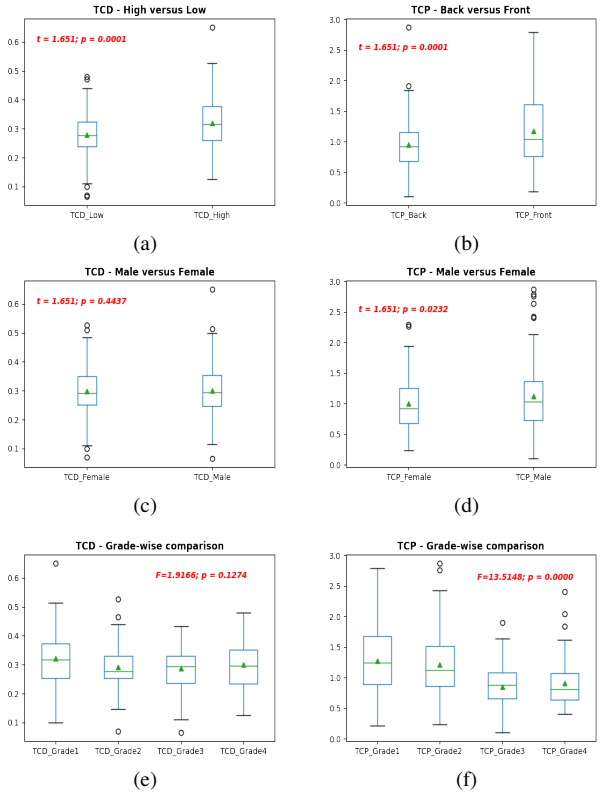
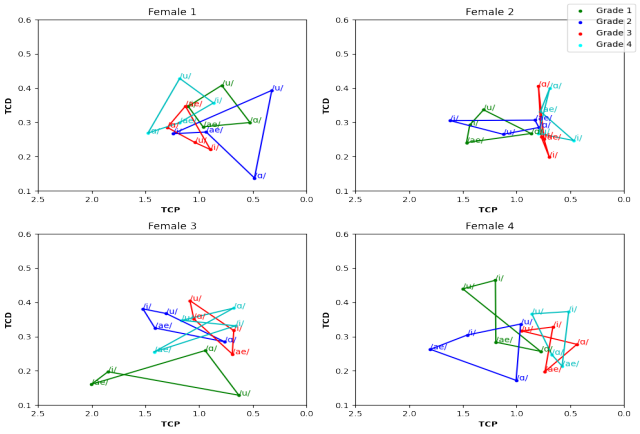
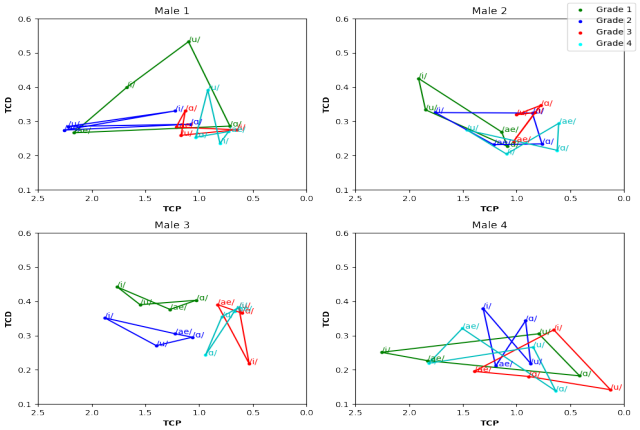


Figure 7: *Grade-wise TCD versus TCP plots. TCD is plotted along the y-axis and TCP is plotted along the x-axis for high/low vowels and Mean TCP for front/back vowels.*



(a) *Grade-wise mean TCD/TCP values for four females*



(b) *Grade-wise TCD/TCP values for four males*

ity" or the "back (pharyngeal) cavity" [29]. Specifically, the F2 of /u/ is a second Helmholtz resonance, the F2 of /ɑ/ could be affiliated with either the front or the back cavity depending on their relative lengths, and the F2 of both /i/ and /æ/ is most likely affiliated with the front cavity. This complex relationship between F2 and vocal tract configuration may explain the absence of a strong resemblance between the vowel acoustic spaces and the corresponding articulatory spaces defined by TCD and TCP.

Male versus female TCD and TCP box plots are given in Figs. 6c and 6d respectively. In Fig. 6c, "TCD_Male" represents TCD values for all vowels /æ/, /i/, /u/, and /ɑ/ over the four years for all male children (128 measurements). Similary, we plot "TCD_Female" in Fig. 6c and "TCP_Male" and "TCP_Female" in Fig. 6d. The mean value is represented by a green triangle marker within each boxplot. The p-value obtained from 2-sample, 1-tailed t-tests comparing TCD and TCP of males and females are given in red font within Figs. 6c and 6d. There was no significant gender-based difference in TCD or TCP.

In Fig. 6e and 6f, we plotted TCD/TCP values for each grade. For example "TCD_Grade1" boxplot in Fig. 6e was plotted using 64 TCD grade 1 measurements (4 vowels $*2$ words per vowel $*8$ children). Boxplots for grades 2, 3, and 4 were similarly plotted. The p-values obtained from one-way ANOVAs examining a main effect of grade on TCD and TCP are given in red font within Figs. 6e and 6f. The significant main effect for TCP suggests that children alter the tongue shapes used in the production of vowels across the elementary school years (cf. [30]). The decreasing TCP values may indicate increasing differentiation between tongue body and tongue blade gestures, with tongue blade gestures reserved for the articulation of consonants.

4. Conclusions

We present the first longitudinal analysis of both acoustic and articulatory feature development in children between grades 1 and 4. Acoustic analyses of four corner vowels reveal that among the children in our study, gender differences emerge around grade 2. SGRs become better boundaries separating the front and back vowels around grade 2 or 3. Articulatory analysis revealed that the development of TCD and TCP parameters is generally unique to a child. With a few exceptions, our results show that, on average, there is a correlation between TCD - vowel height and TCP - vowel frontness. We did not see gender-based differences in TCD and TCP parameters. The development of TCD and TCP parameters is child- and vowel-specific, which cannot be properly observed in cross-sectional studies. Hence, children in our study use varied articulatory configurations to achieve similar acoustic targets. Future directions include analyzing data from more children and more words, with systematic variation of context, in addition to using MRI to obtain vocal tract area functions.

5. References

[1] G. Yeung, A. Afshan, K. E. Ozgun, C. Kaewtip, S. M. Lulich, and A. Alwan, "Predicting clinical evaluations of children's speech with limited data using exemplar word template references." in *SLaTE*, 2017, pp. 161–166.

[2] S. Shahnawazuddin, W. Ahmad, N. Adiga, and A. Kumar, "In-domain and out-of-domain data augmentation to improve children's speaker verification system in limited data scenario," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 7554–7558.

[3] R. Fan and A. Alwan, "DRAFT: A Novel Framework to Reduce Domain Shifting in Self-supervised Learning and Its Application to Children's ASR," in *Proc. Interspeech 2022*, 2022, pp. 4900–4904.

[4] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *The Journal of the acoustical society of America*, vol. 24, no. 2, pp. 175–184, 1952.

[5] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of american english vowels," *The Journal of the Acoustical society of America*, vol. 97, no. 5, pp. 3099–3111, 1995.

[6] S. Lee, A. Potamianos, and S. Narayanan, "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," *The Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1455–1468, 1999.

[7] G. Yeung, S. M. Lulich, J. Guo, M. S. Sommers, and A. Alwan, "Subglottal resonances of american english speaking children," *The Journal of the Acoustical Society of America*, vol. 144, no. 6, pp. 3437–3449, 2018.

[8] S. Lee, A. Potamianos, and S. Narayanan, "Developmental acoustic study of american english diphthongs," *The Journal of the Acoustical Society of America*, vol. 136, no. 4, pp. 1880–1894, 2014.

[9] H. K. Vorperian and R. D. Kent, "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," 2007.

[10] M. E. Kohn and C. Farrington, "Evaluating acoustic speaker normalization algorithms: Evidence from longitudinal child data," *The Journal of the Acoustical Society of America*, vol. 131, no. 3, pp. 2237–2248, 2012.

[11] S. M. Lulich and S. D. Charles, "Development of formant frequency distributions in american english-speaking elementary school-aged children: A longitudinal study," in *Proceedings of Meetings on Acoustics 179ASA*, vol. 42, no. 1. Acoustical Society of America, 2020, p. 060016.

[12] M. Diekhoff and S. M. Lulich, "Anatomical measures of the vocal tract in children ages 5 and 6," *The Journal of the Acoustical Society of America*, vol. 152, no. 4, pp. A59–A59, 2022.

[13] H. Hollien, R. Green, and K. Massey, "Longitudinal research on adolescent voice change in males," *The Journal of the Acoustical Society of America*, vol. 96, no. 5, pp. 2646–2654, 1994.

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

[15] W. T. Fitch and J. Giedd, "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1511–1522, 1999.

[16] Y. Jung *et al.*, "Acoustic articulatory evidence for quantal vowel categories: the features [low] and [back]," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[17] K. N. Stevens, "On the quantal nature of speech," *Journal of phonetics*, vol. 17, no. 1, pp. 3–45, 1989.

[18] K. N. Stevens and S. J. Keyser, "Quantal theory, enhancement and overlap," *Journal of phonetics*, vol. 38, no. 1, pp. 10–19, 2010.

[19] S. M. Lulich, "Subglottal resonances and distinctive features," *Journal of Phonetics*, vol. 38, no. 1, pp. 20–32, 2010.

[20] S. M. Lulich, H. Arsikere, J. R. Morton, G. K. Leung, A. Alwan, and M. S. Sommers, "Analysis and automatic estimation of children's subglottal resonances," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[21] T. M. Byun, E. R. Hitchcock, and M. T. Swartz, "Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 6, pp. 2116–2130, 2014.

[22] S. M. Lulich, J. R. Morton, H. Arsikere, M. S. Sommers, G. K. Leung, and A. Alwan, "Subglottal resonances of adult male and female native speakers of american english," *The Journal of the Acoustical Society of America*, vol. 132, no. 4, pp. 2592–2602, 2012.

[23] L. Ménard, J. Aubin, M. Thibeault, and G. Richard, "Measuring tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model," *Folia Phoniatrica et Logopaedica*, vol. 64, no. 2, pp. 64–72, 2012.

[24] N. Zharkova, F. E. Gibbon, and A. Lee, "Using ultrasound tongue imaging to identify covert contrasts in children's speech," *Clinical linguistics & phonetics*, vol. 31, no. 1, pp. 21–34, 2017.

[25] T. G. Csapó, Z. Bárkányi, T. E. Gráczi, T. Bőhm, and S. M. Lulich, "Relation of formants and subglottal resonances in hungarian vowels," in *Tenth Annual Conference of the International Speech Communication Association*, 2009.

[26] X. Chi and M. Sonderegger, "Subglottal coupling and its influence on vowel formants," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1735–1745, 2007.

[27] G. Dogil, S. M. Lulich, A. Madsack, and W. Wokurek, "Crossing the quantal boundaries of features: subglottal resonances and swabian diphthongs," *Tones and Features: Phonetic and Phonological Perspectives. De Gruyter Mouton*, pp. 137–148, 2011.

[28] K. Stevens, "Acoustic phonetics, cambridge," 1998.

[29] U. G. Goldstein, "An articulatory model for the vocal tracts of growing children," Ph.D. dissertation, Massachusetts Institute of Technology, 1980.

[30] L. Ménard, "Acoustic variability and adaptive articulatory strategies during vocal tract growth revealed by the rounding contrast in french," in *15th International Congress of Phonetic Sciences*, vol. 3, 2003, pp. 3169–3172.