Chapter 1 Results: vocalic development

Meg Cychosz

18 May 2020

1 Results

The primary research question in this study asks if children's vowel variation decreases over the course of development in a language with a three-vowel contrast. The results for the first experiment begin with descriptive statistics of the participants' formant patterns by age, prior to speaker normalization, for the three phonemic vowels /a, i, u/. Then, the formant measures for these phonemic vowels are normalized using one of the scaling techniques - ΔF - described in the methods section. Using these normalized data, a series of models are fit to predict the degree of participants' vowel dispersion and determine if vowel variability decreases over the course of development, as predicted in previous work (Eguchi & Hirsch, 1962; Lee et al., 1999; Ménard et al., 2007). The second half of the results section is devoted to comparing the uniform (ΔF) and non-uniform (Lobanov) formant frequency scaling techniques.

All analyses were conducted in the RStudio computing environment (version: 1.2.5019; {rstudio}. Data visualizations were created with {ggplot2} {ggplot2}. Modeling was conducted using the {lme4} {lme4}, {lmerTest} {lmertest}, and {glmmTMB} {glmmtmb} packages and summaries were presented with {papaja} {papaja} and {Stargazer} {stargazer}. Tests of residual normality were conducted using the {normtest} package {normtest}. The significance of potential model parameters was determined using a combination of log-likelihood comparisons between models, AIC estimations, and p-values procured from model summaries. In all models, continuous predictors were mean-centered to facilitate model interpretation.

```r
# we need a quick estimation of the vowels removed for the 40 children analyzed in chapter 4
ch4_children <- c('c81','c65','c72','c76', 'c56', 'c29', 'c39', 'c59', 'c42', 'c49', 'c55','c33','c26',

vls2_ch4 <- vls2[vls2$spkr %in% ch4_children,] # get the vowel data

vls_removed_ch4 <- vls2_ch4 %>%
  filter(non_na_count_f1<4 | non_na_count_f2<4) %>% # all categories with less than 4 measurements
  select(Phone, spkr) %>%
  distinct() %>%
  nrow #17 tokens removed

# how many phones*age removed?
vls_removed_tbl_ch4 <- vls2_ch4 %>% # how many categories removed? (by phone and age)
  filter(non_na_count_f1<4 | non_na_count_f2<4) %>%
  select(spkr, Age, Phone) %>%
  distinct() %>%
  group_by(Age, Phone) %>%
  count() %>%
  spread(key = "Phone", value = "n")

# now create the data to output
write.csv(vls_removed_tbl_ch4, '/Users/Meg/Box Sync/Dissertation/Experiment_3/results/vls_removed_tbl_cl
```

1.1 Vowel category dispersion

1.1.1 Descriptive statistics of unnormalized data

The first objective in this study is to test how intra-subject and intra-age group vowel variability changes over the course of development from age 4;0 to 10;11. Because we intended to compute vowel variability on an individual speaker and group-level basis, a precautionary data cleaning step was taken before proceeding with the analyses. Any speaker vowel categories that had less than four F1 observations or four F2 observations were removed from analysis (e.g. less than four observations of the F1 of [i] from a given speaker). Since children's voices may be more prone to formant tracking errors due to their higher f0, and thus data removal, we wanted to ensure that any differences between higher voices and lower voices was not due to a data scarcity or abundance in any particular age group. This cleaning procedure helped to standardize the measurements across ages. The reason for differing amounts of tokens per vowel category between speakers was due to the data cleaning procedures and occasional wind interference in the recording, as explained in the methods section.

Table 1: Number of vowel sets removed by age and phone to standardize measurements across age groups. No categories were removed from the adult speakers.

Age	a	i	u
4	NA	1	7
5	NA	NA	4
6	2	1	5
7	4	2	8
8	5	NA	4
9	NA	1	1
10	NA	NA	3

The removal of vowel categories with less than four observations resulted in the removal of 48 vowel categories (see Table 1 for distribution by age group and vowel). All of the adults had at least four clean F1 and F2 measurements for each of their vowel categories so no adult data were removed.

To further ensure that we were accurately comparing between age groups - since younger children might be more likely to have data removed for tracking reasons than the adults - we also selected a random subset of 10 observations for those speaker vowel catgories with more than 10 observations. In this way, no individual speaker contributed more than 10 or less than 4 data points for a given vowel. Unless noted otherwise, all analyses were conducted on these vowel categories that contained 4-10 observations.[1]

Table 2: Median absolute deviation formant measurements in Hertz for children and adults

Age	F1 (MAD)	F2 (MAD)	F3 (MAD)	F4 (MAD)	n
4	551.38 (302)	2086.4 (922)	3952.37 (424)	NA (NA)	206
5	494.73 (194)	2244.96 (1028)	3896.36 (475)	NA (NA)	284
6	483.44 (205)	2119.71 (1131)	3706.18 (402)	NA (NA)	315
7	485.63 (155)	2243.72 (1104)	3810.44 (398)	NA (NA)	495
8	494.64 (208)	2323.6 (1037)	3673.99 (391)	NA (NA)	295
9	459.92 (165)	1892.8 (1094)	3616.57 (550)	NA (NA)	193
10	480.96 (158)	1943.71 (1024)	3503.2 (401)	NA (NA)	305
adult	460.91 (121)	1604.87 (927)	2953.89 (300)	4053.8 (372)	325

[1]Note that this cleaning procedure was not conducted on the mid-vowels [e] and [o] due to data scarcity. Some mid-vowel categories had less than four observations from a given speaker. As a result, while the descriptive statistics report on mid-vowel data, statistical analyses were conducted exclusively on the peripheral vowels.

Table 3: Range (min-max) of formant measurements in Hertz for children and adults

Age	F1 range	F2 range	F3 range	F4 range
4	266-1474	774-3782	2938-4659	NA-NA
5	284-1451	797-3767	2343-4652	NA-NA
6	145-1264	783-3722	2607-4528	NA-NA
7	268-1203	794-3790	2501-4606	NA-NA
8	245-1377	808-3579	2526-4422	NA-NA
9	248-1182	847-3672	2237-4597	NA-NA
10	249-1144	796-3541	2630-4447	NA-NA
adult	243-1091	701-2929	2397-3606	2832-4726

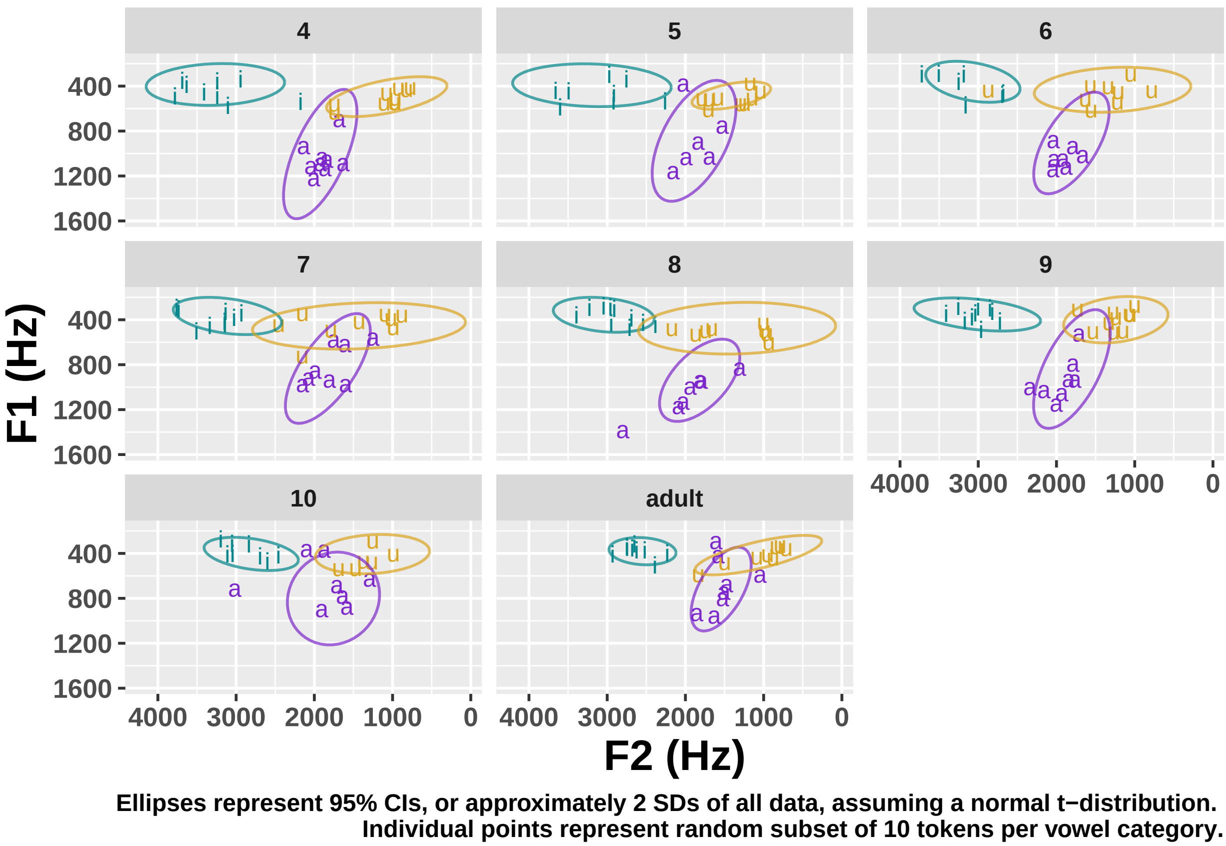


Figure 1: Vowel category development by age, in Hertz: adults and children

Summary statistics of acoustic vowel measurements by age (in Hz) are presented in Tables 2 and 3. The F2-F1 space with the three phonemic vowels, by age, is displayed in Figure 1. (See appendices for individual vowel plots by participant and a plot containing median formant values for phonemic and allophonic vowel categories.) We report on the median and median absolute deviation (MAD) of each formant, instead of the mean and standard deviation (SD), to provide a non-parametric estimate of the variation. The MAD of a distribution is calculated by first computing the median values of the distribution, subtracting this median from each point in the distribution, and finally computing the median of the computed absolute differences. As such, the MAD is relatively less susceptible to the effect of outliers than other measures of dispersion such as SD.

Table 2 demonstrates that overall, as anticipated, the median F1, F2, and F3 values, in Hertz, decrease with age as the vocal tract lengthens. Within the children, the median F2 appears to increase slightly in the 7;0 and 8;0 groups, likely due to the concentration of [u] at higher F2 frequencies. The adult women still exhibit a much lower median F2 than the any of the child age groups. Also as anticipated, the median formant value increases from F1 to F3/F4 across all participants (again F4 was not tracked for the children).

Between-speaker variability, quantified as the MAD, also decreases with age from roughly 300Hz for F1 in the 4;0 group to less than 130 Hz (for F1) in the adults. Notably, the variability even decreases between the older children: for example, for F1-F3, the MAD decreases between the 9;0 group and the 10;0 group, and again decreases between the 10;0 group and the adults. This pattern of variability reduction with age was not always apparent in the higher formants, however. The higher variability in F1 could be due to harmonic spacing in the lower frequencies. In fact, all age groups from 6;0-9;0 were more variable than the 4;0 and 5;0 groups along the F2 dimension. This is somewhat surprising since young children typically master jaw control (correlated with F1) earlier than horizontal lingual control (correlated with F2). But again the higher median F2 variability simply could be reflecting a shifted vowel space. For F3, variability does not appear to decrease notably by age until the 10;0 and adult group.

These results appear to confirm previous work on vowel development in English (Lee et al., 1999) and French (Ménard et al., 2007): younger children are more variable than adults. However, I wanted to ensure that the acoustic variability in the children's speech was due to articulatory instability and differences between acoustic-articulatory mappings in children, and not to the higher formant frequency ranges that the children speak in. In other words, in unnormalized data, child speech could simply *appear* to be more variable because a given amount of acoustic and/or articulatory slop at higher frequencies would result in less auditory perturbation than the same acoustic slop at lower frequencies.

1.1.2 Descriptive statistics of ΔF-normalized data

To ensure that variability in the children's vowel production was not simply due to the frequency ranges of the children's voices, the vowel data were normalized and the variation of each vowel category was computed. The vowels were normalized with the ΔF formant frequency scaling measure reported in the methods. As the calculation of ΔF requires estimation of vocal tract length, the following section begins with a description of the distribution of vocal tract lengths computed from F1-F3 for the children and F1-F4 for the adults, and the ratio between formants (ΔF) by age in this population. (Vocal tract length and formant ratios were computed on all vowels /a, i, e, o, u/). Then, descriptive statistics of formant measurements resulting from the ΔF normalization are presented and within-category dispersion of the phonemic vowels is again evaluated over the course of development.

Table 4: Average vocal tract length and ratio between formant frequencies in Hertz (DeltaF) by age

Age	Vocal tract length (SD)	DeltaF (SD)
4	11.68 (0.84)	1461.73 (97.34)
5	12.13 (0.71)	1405.64 (82.37)
6	12.37 (0.7)	1378.49 (77.51)
7	12.49 (0.74)	1365.79 (80.52)
8	12.48 (0.62)	1365.18 (64.92)
9	13.03 (1.35)	1315.59 (119.91)
10	13.29 (0.48)	1280.58 (45.92)
adult	14.96 (0.68)	1138.81 (51.75)

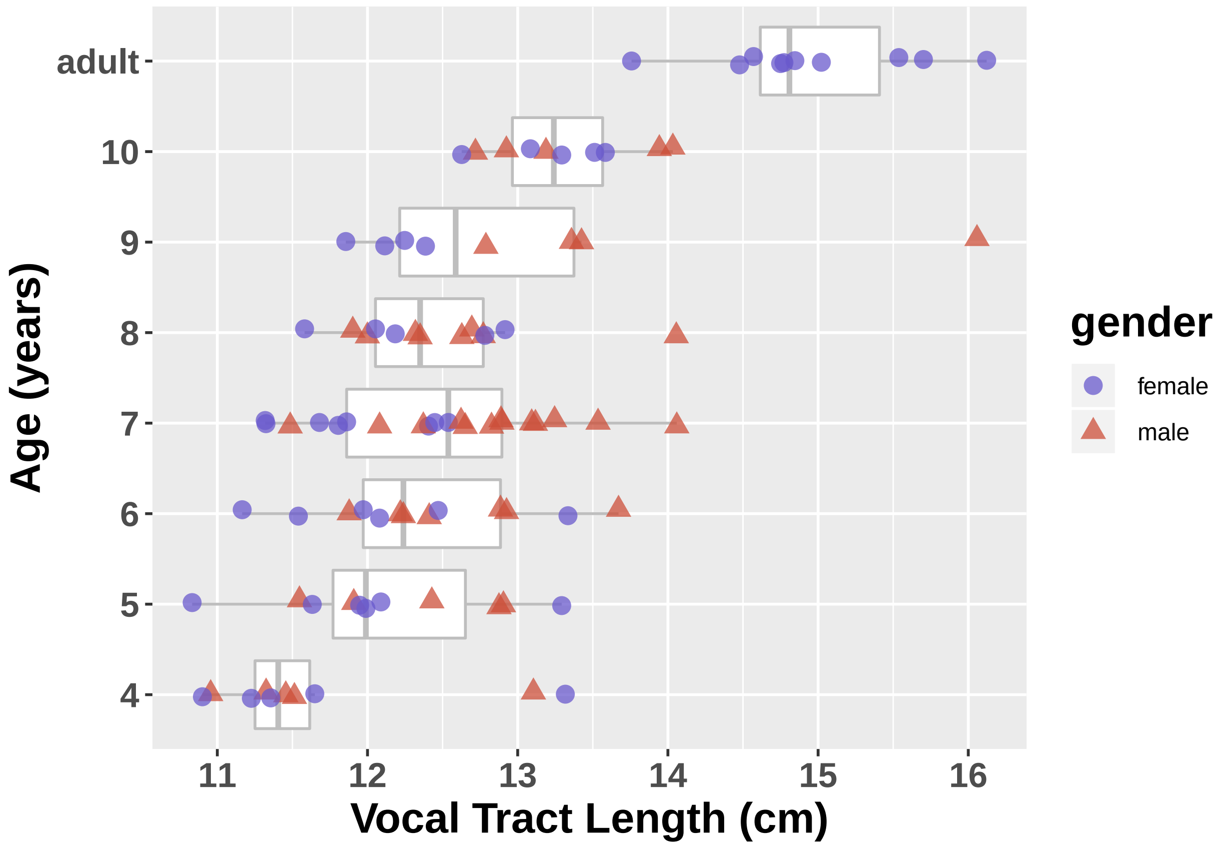


Figure 2: Average vocal tract length by age: adults and children

Table 4 summarizes vocal tract length and the ratio between formant frequencies (ΔF) by age (also see Figure 2). Unsurprisingly, the estimated average vocal tract length increases with age from roughly 12 cm in the four-year-olds to between 13 and 15 cm in the ten-year-olds and adults. Due to the lengthening of the vocal folds and vocal tract, the average ratio between formant frequencies (ΔF) also decreases with age as the average formant frequencies lower. The vocal tract lengths computed acoustically here resemble the measurements taken from magnetic resonance images of vocal tract development in North American children (Vorperian et al., 2005). These acoustically-derived vocal tract length measures are slightly longer than those measured from articulatory imaging; acoustically-derived measures overreport vocal tract lengths since the effect of the end of the tube is just outside of the lips for those measures.

Table 5: Group-level average CoV by formant and vowel for children and adults

	F1			F2		
Age	[a]	[i]	[u]	[a]	[i]	[u]
4	0.26	0.18	0.15	0.09	0.13	0.30
5	0.28	0.20	0.14	0.10	0.16	0.15
6	0.22	0.22	0.22	0.09	0.10	0.38
7	0.25	0.23	0.23	0.14	0.09	0.42
8	0.19	0.20	0.18	0.18	0.12	0.41
9	0.25	0.22	0.20	0.11	0.10	0.26
10	0.25	0.17	0.17	0.16	0.09	0.34
adult	0.28	0.16	0.22	0.12	0.07	0.34

1.1.3 Measuring within-category variability

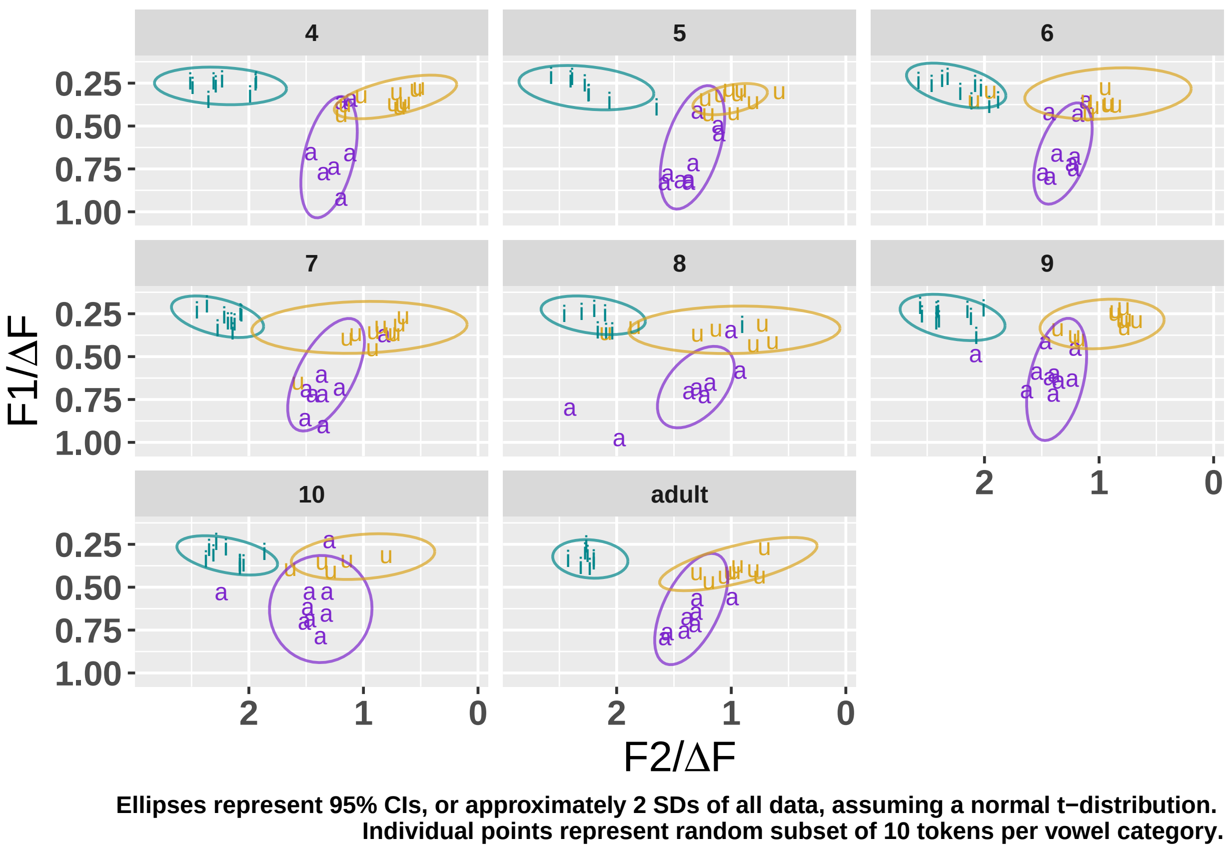


Figure 3: DeltaF-normalized vowels by age: adults and children

Table 6: Individual-level CoV by formant and vowel for children and adults:F1

Age	[a] mean (SD) range	[i]	[u]
4	0.19 (0.13) 0.04 - 0.48	0.11 (0.09) 0.01 - 0.23	0.13 (0.01) 0.12 - 0.14
5	0.25 (0.17) 0.05 - 0.52	0.15 (0.07) 0.04 - 0.26	0.1 (0.07) 0.01 - 0.22
6	0.17 (0.12) 0.03 - 0.38	0.15 (0.1) 0.04 - 0.32	0.19 (0.09) 0.03 - 0.31
7	0.22 (0.11) 0.04 - 0.46	0.14 (0.09) 0.04 - 0.42	0.17 (0.09) 0.04 - 0.39
8	0.12 (0.09) 0.04 - 0.28	0.09 (0.06) 0.01 - 0.2	0.05 (0.04) 0.01 - 0.12
9	0.21 (0.14) 0.04 - 0.47	0.11 (0.06) 0.04 - 0.17	0.15 (0.07) 0.05 - 0.23
10	0.16 (0.14) 0.03 - 0.43	0.11 (0.07) 0.02 - 0.21	0.08 (0.06) 0.02 - 0.16
adult	0.18 (0.15) 0.04 - 0.49	0.13 (0.05) 0.06 - 0.22	0.17 (0.1) 0.03 - 0.32

Table 7: Individual-level CoV by formant and vowel for children and adults:F2

Age	[a] mean (SD) range	[i]	[u]
4	0.07 (0.02) 0.05 - 0.09	0.1 (0.08) 0.03 - 0.25	0.26 (0.13) 0.11 - 0.37
5	0.08 (0.05) 0.04 - 0.21	0.13 (0.11) 0.03 - 0.34	0.13 (0.06) 0.07 - 0.25
6	0.07 (0.03) 0.03 - 0.13	0.07 (0.03) 0.04 - 0.12	0.27 (0.12) 0.06 - 0.43
7	0.11 (0.06) 0.03 - 0.24	0.06 (0.04) 0.01 - 0.2	0.32 (0.14) 0.08 - 0.53
8	0.12 (0.11) 0.04 - 0.39	0.1 (0.08) 0.02 - 0.34	0.16 (0.1) 0.02 - 0.32
9	0.09 (0.07) 0.04 - 0.23	0.08 (0.06) 0.03 - 0.22	0.18 (0.1) 0.07 - 0.33
10	0.12 (0.08) 0.03 - 0.26	0.08 (0.04) 0.05 - 0.14	0.26 (0.19) 0.08 - 0.52
adult	0.08 (0.06) 0.02 - 0.22	0.05 (0.02) 0.01 - 0.08	0.2 (0.14) 0.04 - 0.48

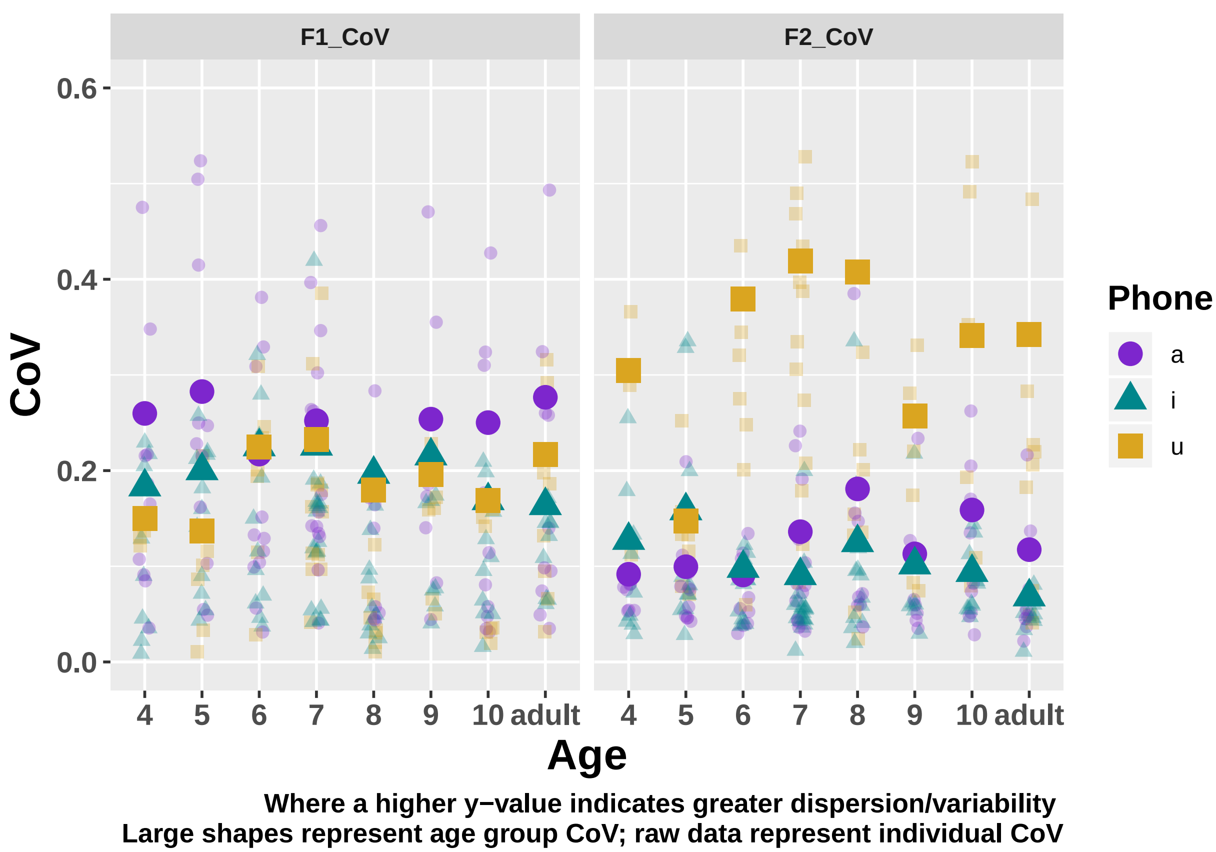


Figure 4: Group-level CoV by age and phone

The comparison of within-category variability across age groups was conducted on the ΔF-normalized formant values. Figure 3 plots the median ΔF-normalized formant values. (See Tables 13 - 15 in the appendices for descriptive statistics of ΔF-normalized formant values.) As the vowel plots demonstrate, speakers tend to have larger within-category dispersion for [a] and [u] than [i]. However, the pattern by age is less identifiable. With speaker-intrinsic information (ratio between formant frequencies on a by-speaker basis) factored out, which is what ΔF does, adults appear to have somewhat tighter, more compact acoustic vowel categories for some vowels, particularly [i], than even the eldest children. Other vowels show little difference by age, or, in the case of [u], do not seem to follow a strict linear pattern of decreased variation with age.

To better ascertain the developmental pattern of vowel variability, we computed the acoustic dispersion of each vowel category across the age groups. To do so, we followed previous work in computing the Coefficient of Variation (CoV) for F1 and F2 of each vowel category in the ΔF-normalized formants. The CoV is the ratio of the standard deviation of the mean to the mean of each phoneme category (Bradlow, 1995; Eguchi & Hirsch, 1965; Lee et al., 1999). Again, this was computed individually for F1 and F2 of each phoneme for each age group.

Recall that these measurements were made over vowel categories containing 4-10 observations per speaker. The group- and individual speaker-level CoVs computed over *all* data, including those categories with less than 4 data points, are included in the appendices (CoV by age group: Table 16 and CoV by individual speaker: Tables 17 and 18); overall we did not notice a large difference between the CoVs computed over 4-10 tokens per vowel category and those computed over the entire dataset. However, we felt that standardizing the number of tokens per category would result in the least bias across the age groups.

The average CoV for F1 and F2 by phone and age group is listed in Table 5 and plotted in Figure 4 (following Lee et al, (1999), we report the average CoV measurements and their SDs, not the median or MAD): the results demonstrate that the CoV for by phone and formant 1) varies by phone and 2) does not always decrease linearly with age. For the F1 of [a], the CoV does not appear to decrease by age - the average [a] CoV for the adults was 0.24 but 0.18 for the 8;0 group. The F1 CoV for [u] is highest for the 6;0-7;0 groups, again not showing clear evidence of a linear decrease in variability. The F2 CoV for [i] and [a] show limited change with age, though the CoV is smallest for [i] in the adults.

We additionally computed a CoV on a by-speaker basis. This resulted in a single coefficient for F1 and F2 from each speaker, for each vowel. The mean of these CoVs, SDs, and ranges are listed in Table 7. These individual speaker-level CoVs are also plotted as the raw data in Figure 4. Once we computed the CoV on an individual level, we then had a datapoint for each speaker's three vowel categories which we used to fit a series of models predicting the CoVs for F1 and F2.

1.1.4 Fitting models to predict within-category variability

For the CoV model fitting, we fit generalized linear mixed effects models (GLMMs) because the outcome variable - the individual speaker level CoV - was necessarily non-negative and right-skewed. Gamma GLMMs were fit using a log linking function to appropriately model the skewed, non-Gaussian distribution of the residual. A total of six models were fit, one for each combination of formant (F1 and F2) and phone (/a, i, u/). The model fitting procedure followed the same procedure for all models: first the baseline model, with just the random effect of **Speaker**, was fit. Then, the parameters of interest were added, with the primary parameter of interest (**Age**) added last. First, **Gender** was added, then **Age Group** (4, 5, 6, 7, 8, 9, 10, adult). The effect of **Gender** did not improve upon any of the baseline (with only **Speaker**) model fits. This suggests that individual speaker-level CoV does not reliably differ between any of the studied age groups in this population and the parameter **Gender** waas not further analyzed.

The parameter **Age Group** improved upon baseline model fits for two models: the model predicting F2 of [i] and the model predicting F1 of [u]. The model summary for the F2 of [i] is shown in Table 8 and the summary for the F1 of [u] is in Table 9. As the beta coefficients in the model summaries demonstrate, there were some differences by age group in category dispersion. However, these differences only approached significance reliably - meaning that almost all age groups significantly differed from the adults - in the F2 of [i] model. Furthermore, as the beta coefficients demonstrate, the children's CoV values did not decrease linearly with age, with older children necessarily demonstrating more adult-like dispersion.

Thus, although there are some trends apparent in the descriptive statistics of the individual-level and group-level CoVs, as well as in the vowel plots, overall the modeling suggests that category dispersion does

Table 8: Model predicting CoV for F2 of [i]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-3.09	0.19	-16.37	0.00	-2.72,-3.46
10	0.59	0.26	2.27	0.02	1.1,0.08
4	0.64	0.27	2.37	0.02	1.17,0.11
5	0.88	0.26	3.43	0.00	1.39,0.38
6	0.47	0.25	1.89	0.06	0.96,-0.02
7	0.20	0.23	0.90	0.37	0.65,-0.24
8	0.63	0.25	2.54	0.01	1.11,0.14
9	0.49	0.29	1.71	0.09	1.05,-0.07

Table 9: Model predicting CoV for F1 of [u]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-1.77	0.21	-8.61	0.00	-1.36,-2.17
10	-0.73	0.31	-2.37	0.02	-0.13,-1.34
4	-0.28	0.41	-0.68	0.50	0.53,-1.08
5	-0.54	0.31	-1.74	0.08	0.07,-1.15
6	0.10	0.31	0.34	0.74	0.71,-0.5
7	0.01	0.27	0.06	0.96	0.54,-0.51
8	-1.22	0.30	-4.09	0.00	-0.64,-1.81
9	-0.14	0.31	-0.46	0.65	0.47,-0.75

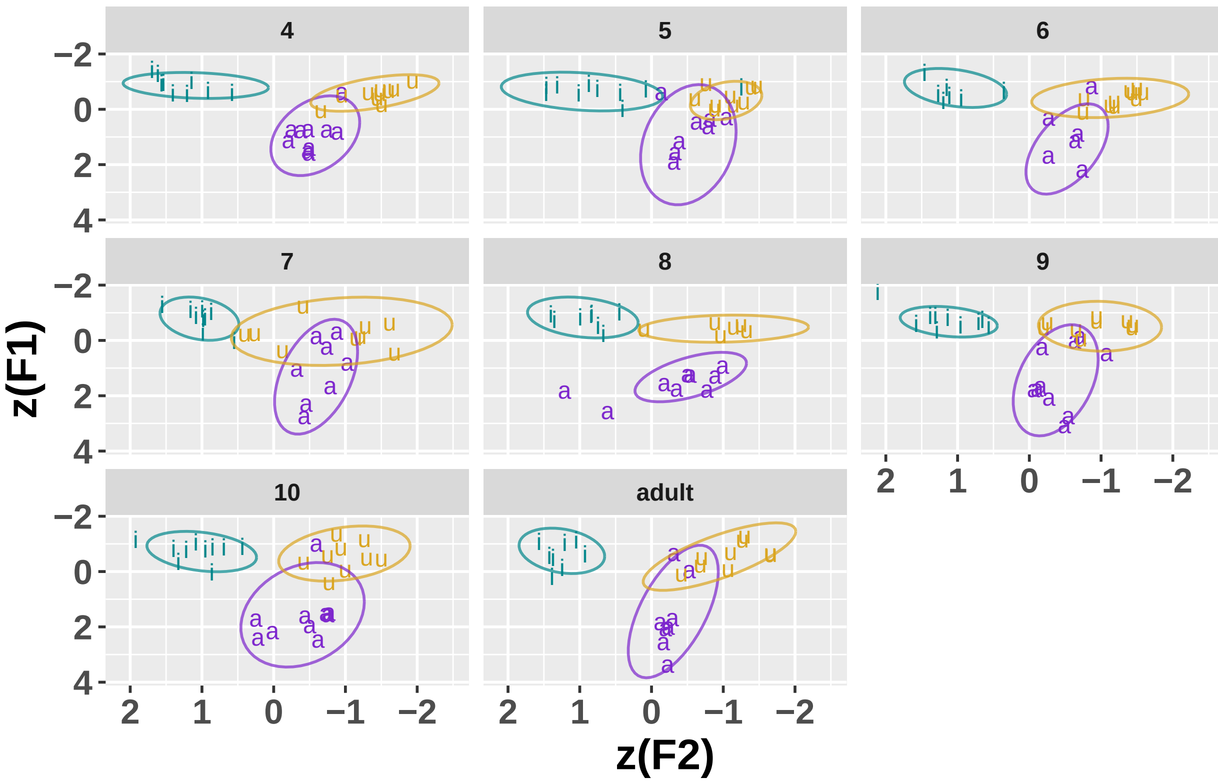
not differ significantly between the adults and the child age groups. Furthermore, when there are statistically reliable differences between adults and children for within-category dispersion, these differences do not diminish with age.

1.2 Compensation for vocal tract morphology

Having addressed the first objective of this study, the secondary research question asks if uniform formant frequency scaling adequately factors out anatomical differences between children and adults. If a comparison of uniform and non-uniform formant frequency scaling techniques shows similar results, this may demonstrate that children compensate articulatorily for their vocal tract morphology during vowel production. If, however, a comparison of the two scaling techniques shows large differences between child and adult within-category variability, this suggests that there are additional sources that explain the differences between adults and children, beyond phone identity and anatomical difference.

To evaluate this question, the vowel data are normalized using two formant scaling techniques - uniform (ΔF) and non-uniform (Lobanov), as described in the methods section. For this analysis, the allophonic vowels [e, o] are included, in addition to the phonemic vowels /a, i, u/. The difference in category variability between the scaling techniques is then compared: if there is additional, unexplained variability present in the Lobanov-normalized vowels - meaning there is additional variability after scaling the formants uniformly (accounting for vocal tract length) and factoring in phone identity - then these results would suggest that another factor, such as articulatory configuration, may differ between the speakers. Such a result would suggest the children may not always compensate for the ratio between their supraglottal cavities during vowel production.

The vowel data were first normalized using the Lobanov scaling technique. Figure 5 plots the median Lobanov-normalized formant values (for phonemes) by phone and age group and Tables 19-21 in the appendices list the descriptive values for the phonemes /a, i, u/. See Tables 13-15 in the appendices, and Figure 3 in the previous section, for the ΔF-normalized formant values.



Ellipses represent 95% CIs, or approximately 2 SDs of all data, assuming a normal t−distribution.
Individual points represent random subset of 10 tokens per vowel category.

Figure 5: Lobanov-normalized vowels by age: adults and children

After scaling the vowels using the two scaling methods, the vowel variability between the children and adults for both sets of normalized vowels (Lobanov and ΔF) was measured. Again, for the Lobanov normalization technique separate normalization factors are calculated for F1 and F2 for each speaker. The ΔF technique, however, uses a single normalization coefficient for each speaker. We measured the variability of each vowel category for each age group (so three categories per age group) using a single coefficient that reflects both the *mean* value of each vowel category and the *variability* of each category along the F1 and F2 dimensions. To calculate this coefficient, we took the following steps.

1. First, we measured the difference between the mean value of each vowel category for the adults and a given child age group. For example, we measured the mean F1 of [a] over all of the 5;0 children and the mean F1 of [a] over all of the adults and then we took the difference between these values. This calculation was performed for the F1 and F2 measures for each vowel category. This step estimated the position of the vowel category in acoustic space.

2. For the next step, we took the root mean square of the difference between the vowel categories for F1 and F2 (calculated in step one). Finally, these root mean square calculations for F1 and F2 were summed. This step estimated the standard deviation of the vowel category (disperion of the category in space).

These steps resulted in a single coefficient for each vowel category, for each child age group, that factored in both formants. These steps were applied separately to the ΔF-normalized vowels and the Lobanov-normalized vowels.

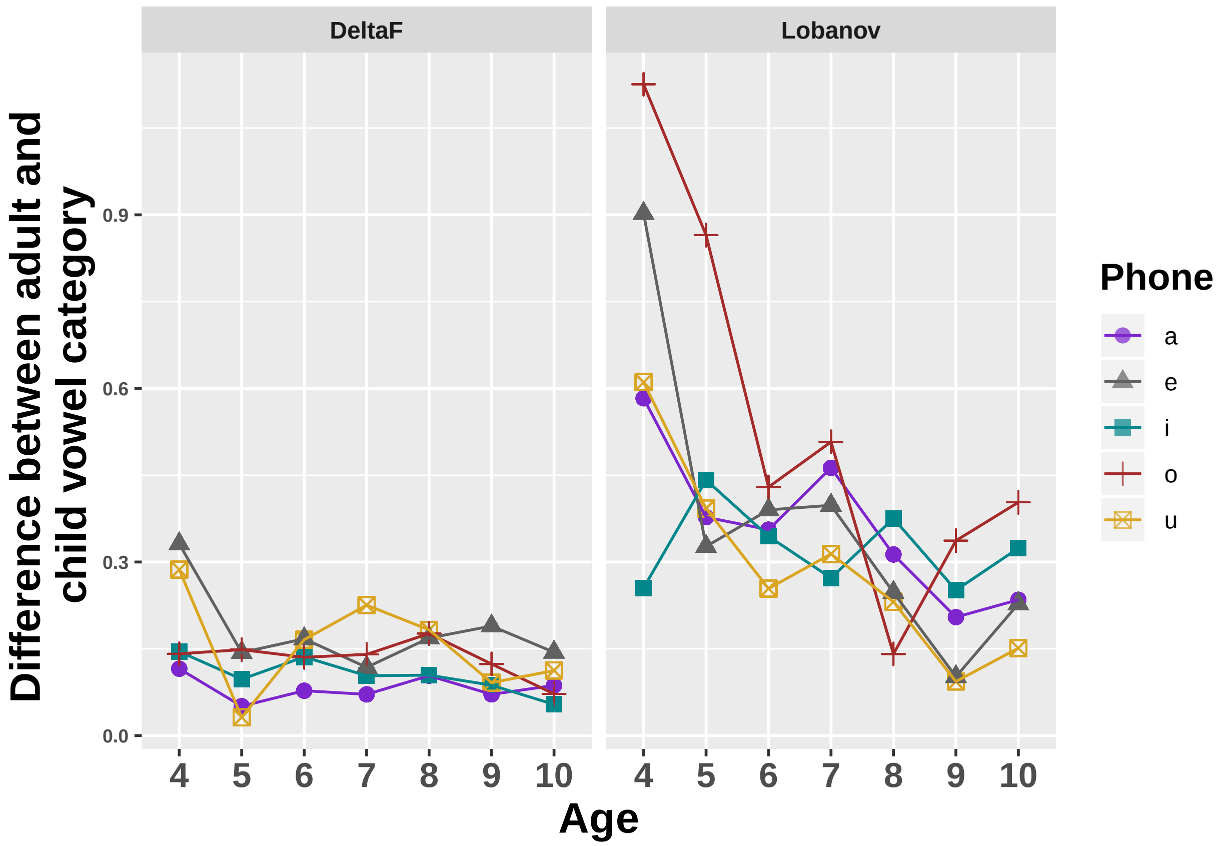


Figure 6: Difference between adult and child vowel categories: RMS method

Figure 6 displays the difference between the adult and child vowel categories for the two scaling techniques. The left panel of the figure shows the difference between adult and child ΔF-normalized categories, or normalization that incorporates just vocal tract length. The right panel shows the difference between adult and child Lobanov-normalized categories, which incorporate individual formants into the normalization procedure. The visual clearly demonstrates that there are larger differences between normalized adult and child vowel categories that taken into account individual formants than vocal tract alone. In other words, beyond vocal tract length and even phone identity, there is still additional variability between adults and children to be accounted for, as the larger values in the right panel demonstrate. These differences between Lobanov-normalized categories and ΔF-normalized categories will be quantified below. It is important to note that the differences between adult and child vowel categories did nevertheless differ by age and phone. For example, there was a large amount of variation between the adult and 4;0 [e] category, but less between the adult and 10;0 [o] category.

Thus far, though we employed a different methodology, these results corroborate Turner et al. (2009): there is additional, unexplained variability in formant frequencies even after removing the effects of phone identity and vocal tract length. Turner et al. (2009) concluded that this outstanding variability was due to measurement error resultant from formant tracking in spectrograms and employing LPC. The authors concluded that any remaining variability beyond that was likely statistically meaningless.

On the basis of this observed difference between the two scaling techniques, we hypothesized that the increased variability in the children's productions *could* be due to their articulatory configurations, and not simple measurement error. Specifically, if some children were not compensating for the ratios between their front and back cavities by adjusting their lingual positioning to approximate adult-like formant frequency ratios between cavities, then we would see this exact difference between the uniform and non-uniform scaling techniques. However, this preliminary conclusion requires additional evidence.

We took several steps to evaluate the idea that the outstanding variability might reflect children's lack of compensation for their vocal tract morphologies - at least some of the time.

1. We classified the children into two groups on the basis of their ratios between back-affiliated and front-affiliated formants. We did so using a technique that did not require us to make assumptions about which formants originated in which cavity (to be explained in detail in the following section). This split was designed to classify the children into more "adult-like" articulators (those children who have either approximated adult-like cavity ratios or who adjust their lingual articulations to approximate the adult-like acoustics) and the rest of the children into more "child-like" articulators (those children who may not compensate articulatorily as much for their vocal tract morphologies).

2. We fit a series of statistical models to the children's un-normalized formant frequency data. In doing so, we were able to evaluate in a stepwise manner the effect of adding parameters that are known to influence formant frequency values, such as the identity of a phone or vocal tract length, as well as parameters that we hypothesized *might* influence formant values, such as the ratio between front- and back-cavity affiliated formants.

The following two sections outline the results of these analyses.

1.2.1 Classification into adult-like and child-like articulators

The first way that we evaluated if children were compensating for their vocal tract morphology was to divide the children into two groups: adult-like articulators and child-like articulators. This classification was made on the basis of the children's ratios between the first and second formants for [a] and ratios between the second and third formants for [a]. We opted for these criteria for the classification for several reasons. One, if children do not alter their articulatory strategies to compensate for their relatively longer oral cavity, then they are expected to show different ratios between these formants when compared to adults. Specifically, the longer oral cavity relative to pharyngeal - because the pharyngeal cavity grows disproportionately fast in childhood for boys and girls - results in a heightened F2 compared to an adult model (when assuming adult-like formant-cavity affiliations). Thus, those formants deriving from the back cavity in children would be lower than formants affiliated to a front cavity.

The decision was made to calculate formant frequency *ratios* because doing so does not require knowledge or assumptions of formant-cavity affiliations. Formant-cavity affiliations may be problematic assumptions to draw acoustically because they can become highly unpredictable in the event of some articulatory modifications (e.g. undershoot). Formant-cavity affiliations can also potentially vary due to idiosyncratic vocal tract characteristics (e.g. a heightened palate). Formant ratios, such as the measures proposed here, merely reflect the relationship between the cavity sizes, independent of specific affiliations, allowing us to skirt the issue of cavity-formant affiliation (Apostol et al., 2004).

Finally, we chose the ratios between F1 to F2 and F2 to F3 for [a] in particular because it has been predicted that the ratio between F1 and F2 for [a] is larger in children than adults and that the ratio between F2 and F3 for [a] is smaller in children than adults (Ménard et al., 2007). We additionally opted to calculate formant ratios for [a], instead of [i] or [u], as Helmholtz resonators are the source of (some of) the formants of [i] and [u]. Since Helmholtz resonators reflect constriction length and area, in addition to cavity length and volume, these were less ideal options to reflect the ratio between cavity sizes in the children.

Before splitting the children into groups on the basis of these formant ratios, we performed an additional check on our assumptions concerning the ratios between back- and front-affiliated formants. Like the child vocal tract, though to a lesser extent, the female adult vocal tract has a smaller pharyngeal cavity relative to oral. Consequently, to validate the formant ratio approach, we calculated the mean F1, F2, and F3 for [a] over all of the adult speakers in the classic Petersen & Barney (1952) vowel dataset and also the Hillenbrand et al. (1995) vowel dataset using the phonTools R library (Barreda, 2015). The Petersen & Barney dataset contained n=28 adult women and n=33 adult men and the Hillenbrand et al. dataset contained n=48 adult women and n=45 adult men. Overall, we confirmed the relationship between F1 and F2 for [a]: in the Petersen & Barney dataset, women had a higher average F1:F2 ratio (female mean=0.70, sd=0.06) than men (male mean=0.66, sd=0.04). This was also the case for F1:F2 in the Hillenbrand dataset: F1:F2 for women (mean=0.61, sd=0.07) and F1:F2 for men (mean=0.58, sd=0.06).

The ratio between F2 and F3 in the datasets was less straightforward. As anticipated, women had a smaller F2:F3 ratio in the Petersen & Barney dataset, though this difference was slight (female: mean=0.44, sd=0.05, male: mean=0.45, sd=0.06). However, women had a larger average F2:F3 ratio in the Hillenbrand data (female: mean=0.54, sd=0.05, male: mean=0.52, sd=0.06). As a result of this, although we factor both F1:F2 and F2:F3 into the calculation of the median split by children, we rely only on the F1:F2 ratio for our statistical modeling procedure presented in the following section.

Having confirmed that the formant ratios express differences in cavity size, we turned to the calculations of formant ratios for the children in the current study. To do this, we calculated the mean F1, F2, and F3 for [a] for each child. (Participant c64 only had one valid [a] token, so formant ratios for that participant were not computed and c64 is not included in the following analyses.) Then, we calculated the median F1:F2 ratio and F2:F3 ratio for [a] over all the children. Those children who had both a higher F1:F2 ratio *and* lower F2:F3 ratio for [a] were classified as "child-like" articulators. Children who did not satisfy both criteria were classified as "adult-like" articulators. So a child who only had a higher F1:F2 ratio or only a lower F2:F3 ratio was not classified as child-like. This method allowed us to maximally conservative in the classification of the children.

Table 10 displays the number of children by age group who were classified as adult-like articulators and child-like articulators. As anticipated, in the older age groups 9;0 and 10;0, there are fewer children who were classified as child-like articulators. Still, the fact that several children were classified as child-like articulators, even a few in the eldest age groups, demonstrates that there may be great variability in when children learn to approximate adult-like acoustic patterning and adult-like cavity sizes. This makes sense as children grow at different rates, have growth spurts at different times, and have different experience practicing language. These factors are further considered in the discussion section.

Figures 7 and 8 display the results of the comparison by median split. On the left, the adult-like articulators show a linear decrease in difference between Lobanov and ΔF vowel category size. This linear decrease with age is anticipated because as the adult-like children age, there should be fewer and fewer differences between adults and children outside of vocal tract length. This is presumably because as children age, they are more likely to have mastered vocal tract morphology compensation. (A linear decrease is not exactly present for [i].)

Table 10: Number of children classified as child-like and adult-like articulators, by age

Age	adult-like articulator	child-like articulator	total
4	6	4	10
5	10	1	11
6	6	7	13
7	16	4	20
8	8	5	13
9	8	NA	8
10	8	2	10

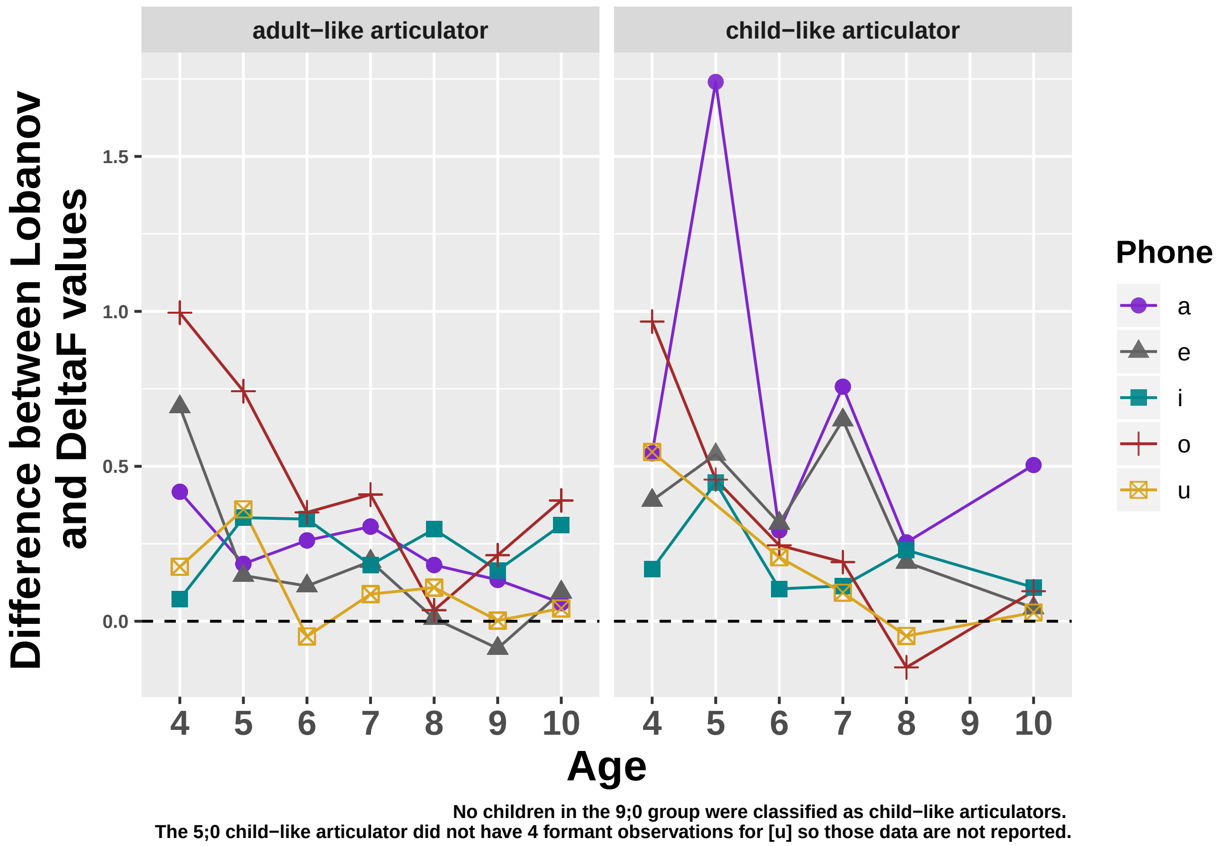


Figure 7: Difference between adult and child vowel categories by scaling method and articulatory status

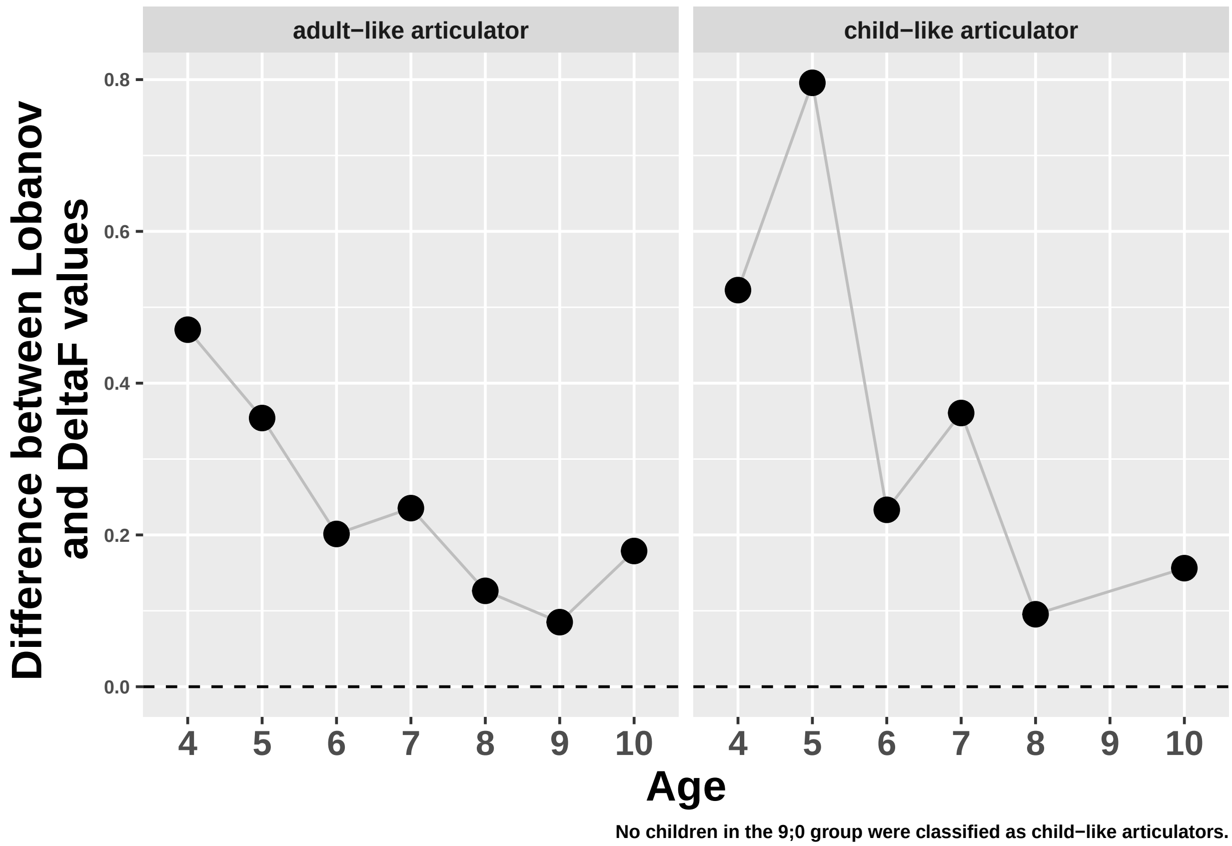


Figure 8: Average difference between adult and child vowel categories by scaling method and articulatory status.

1.2.2 Fitting models to un-normalized formant frequency data

As a final step in the analysis, we built statistical models to predict the values of un-normalized formant frequency data from the child speakers. The goal of this modeling was to determine the predictors of formant frequencies in the children. The constructed models permitted control of factors known to influence formant frequencies, such as phone identity and vocal tract length, so that we could determine if the addition of cavity size (ratio of F1:F2 for [a]) explained any remaining variance in the frequencies. Thus the dependent variable in the modeling was *un-normalized* formant frequency data (separate models for F1 and F2; details below) because the intent was to factor out anatomical differences via the addition of a **Vocal Tract Length** parameter in the models. Additionally, for this modeling, we only incorporated the peripheral, phonemic vowels /a, i, u/ because we did not have sufficient data for the allophonic vowels to include them in the modeling.

Two generalized linear mixed effects models (GLMMs) were fit to predict the children's un-normalized formant frequency data. One model was fit to predict the F1 measures and another model was fit to predict the F2 measures. The choice to fit GLMMs was made due to the non-negative and right-skewed nature of the F1 and F2 outcome variables. Continuous variables were mean-centered and standardized to facilitate model interpretation. Both models were fit according to the following procedure: first, a baseline model with only the random intercepts for individual **Speaker** was fit. Then, model parameters were added in the following order: **Phone** (/a, i, u/), **Vocal Tract Length**, **F1:F2 ratio for** [**a**], and **Gender**. Parameter significance was determined on the basis of log likelihood tests and AIC values as well as p-values procured from the model summaries. Once again the models were only fit to predict formant variability within the child speakers.

For both the F1 and F2 models, **Phone**, **Vocal Tract Length**, and **F1:F2 ratio for** [**a**] improved upon baseline model fits. **Gender** did not improve upon the F1 or F2 models containing **Phone**, **Vocal**

Table 11: Model predicting F1 frequencies

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	861.63	5.71	150.78	0.00	872.83,850.43
Phone:[i]	-479.88	7.76	-61.83	0.00	-464.67,-495.1
Phone:[u]	-403.58	9.87	-40.88	0.00	-384.23,-422.93
Vocal Tract Length	-47.40	5.16	-9.18	0.00	-37.28,-57.52
F1:F2 ratio for [a]	579.18	60.26	9.61	0.00	697.29,461.07

Table 12: Model predicting F2 frequencies

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	1,818.95	13.32	136.57	0.00	1845.06,1792.85
Phone:[i]	1,220.24	16.17	75.47	0.00	1251.93,1188.55
Phone:[u]	-400.71	21.21	-18.89	0.00	-359.14,-442.28
Vocal Tract Length	-174.68	12.94	-13.50	0.00	-149.32,-200.05
F1:F2 ratio for [a]	-537.91	150.20	-3.58	0.00	-243.52,-832.29

Tract Length, and **F1:F2 ratio for [a]**. The final model summary for F1 is listed in Table 11 and the model summary for F2 is in Table 12. This result for the models suggests that there is additional variance in the formant frequency measures beyond phone identity and vocal tract length - a finding that Turner et al. (2009) likewise ascertained. However, the addition of the **F1:F2 ratio for [a]** parameter in a model already controlling for extant variability (i.e. in the error term) suggests that the outstanding variance in the formant frequency measures can likewise be accounted for with the size of the children's cavities.

Overall, these results confirm previous research that vocal tract length and phone identity explain large amounts of between-speaker acoustic variation. However, via a step-by-step elimination of these known variables, and statistical model building, we have demonstrated that there may be some school-aged children who do not yet approximate adult-like articulatory configurations, as evidenced in the ratios of front and back cavity-affiliated formants. This lack of compensation for vocal tract morphology, which operates *somewhat* independently of vocal tract length, may explain some of the variability known to characterize child speech.

2 Appendices

Table 13: DeltaF-normalized formant frequencies by age group: [a]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	0.70	0.13	1.29	0.11	2.63	0.28	NA	NA
5	0.62	0.19	1.31	0.13	2.69	0.27	NA	NA
6	0.68	0.11	1.31	0.11	2.63	0.25	NA	NA
7	0.62	0.15	1.33	0.17	2.65	0.33	NA	NA
8	0.69	0.11	1.28	0.11	2.60	0.38	NA	NA
9	0.62	0.16	1.36	0.11	2.50	0.20	NA	NA
10	0.63	0.12	1.34	0.19	2.64	0.37	NA	NA
adult	0.61	0.13	1.35	0.08	2.43	0.22	3.4	0.17

Table 14: DeltaF-normalized formant frequencies by age group: [i]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	0.25	0.04	2.29	0.20	2.81	0.27	NA	NA
5	0.27	0.06	2.25	0.22	2.83	0.24	NA	NA
6	0.26	0.06	2.25	0.18	2.80	0.19	NA	NA
7	0.26	0.05	2.27	0.18	2.82	0.20	NA	NA
8	0.25	0.04	2.20	0.14	2.75	0.27	NA	NA
9	0.27	0.06	2.26	0.20	2.85	0.27	NA	NA
10	0.32	0.05	2.20	0.21	2.76	0.25	NA	NA
adult	0.33	0.05	2.24	0.15	2.71	0.14	3.68	0.27

Table 15: DeltaF-normalized formant frequencies by age group: [u]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	0.34	0.05	0.68	0.09	2.50	0.49	NA	NA
5	0.35	0.06	1.00	0.13	2.65	0.43	NA	NA
6	0.31	0.06	0.92	0.27	2.63	0.23	NA	NA
7	0.33	0.05	1.03	0.41	2.73	0.27	NA	NA
8	0.36	0.04	0.83	0.24	2.63	0.29	NA	NA
9	0.30	0.08	0.98	0.27	2.81	0.37	NA	NA
10	0.32	0.07	1.03	0.23	2.93	0.37	NA	NA
adult	0.37	0.07	0.91	0.33	2.54	0.23	3.52	0.38

Table 16: Group-level average CoV by formant and vowel for children and adults: all data

	F1			F2		
Age	[a]	[i]	[u]	[a]	[i]	[u]
4	0.26	0.18	0.13	0.09	0.13	0.27
5	0.28	0.21	0.11	0.10	0.15	0.15
6	0.22	0.23	0.23	0.10	0.09	0.40
7	0.25	0.23	0.25	0.12	0.09	0.41
8	0.19	0.20	0.18	0.19	0.12	0.42
9	0.26	0.22	0.20	0.11	0.10	0.26
10	0.26	0.16	0.18	0.16	0.09	0.25
adult	0.27	0.15	0.22	0.12	0.07	0.35

Table 17: Individual-level CoV by formant and vowel for children and adults: F1

Age	[a] mean (SD) range	[i]	[u]
4	0.19 (0.14) 0.04 - 0.5	0.11 (0.09) 0.01 - 0.23	0.13 (0.02) 0.1 - 0.14
5	0.23 (0.17) 0.03 - 0.52	0.15 (0.08) 0.04 - 0.26	0.08 (0.04) 0.01 - 0.13
6	0.17 (0.13) 0.03 - 0.41	0.15 (0.1) 0.04 - 0.33	0.19 (0.08) 0.05 - 0.31
7	0.2 (0.11) 0.02 - 0.46	0.13 (0.09) 0.03 - 0.42	0.18 (0.09) 0.04 - 0.39
8	0.12 (0.09) 0.04 - 0.3	0.09 (0.07) 0.01 - 0.2	0.05 (0.04) 0.01 - 0.12
9	0.21 (0.14) 0.05 - 0.47	0.1 (0.05) 0.04 - 0.17	0.15 (0.07) 0.05 - 0.23
10	0.17 (0.14) 0.03 - 0.43	0.11 (0.07) 0.01 - 0.23	0.07 (0.06) 0.01 - 0.16
adult	0.18 (0.14) 0.05 - 0.48	0.12 (0.05) 0.06 - 0.22	0.13 (0.08) 0.03 - 0.29

Table 18: Individual-level CoV by formant and vowel for children and adults: F2

Age	[a] mean (SD) range	[i]	[u]
4	0.07 (0.02) 0.04 - 0.09	0.11 (0.08) 0.03 - 0.25	0.23 (0.14) 0.08 - 0.37
5	0.08 (0.05) 0.01 - 0.21	0.13 (0.1) 0.03 - 0.33	0.14 (0.06) 0.08 - 0.25
6	0.07 (0.04) 0.03 - 0.13	0.07 (0.03) 0.04 - 0.13	0.26 (0.09) 0.09 - 0.35
7	0.1 (0.06) 0.03 - 0.26	0.06 (0.03) 0.01 - 0.17	0.31 (0.13) 0.08 - 0.51
8	0.12 (0.11) 0.04 - 0.39	0.08 (0.08) 0.02 - 0.32	0.16 (0.11) 0.02 - 0.36
9	0.09 (0.07) 0.03 - 0.23	0.08 (0.07) 0.03 - 0.22	0.19 (0.11) 0.08 - 0.33
10	0.11 (0.08) 0.02 - 0.26	0.08 (0.03) 0.05 - 0.14	0.18 (0.12) 0.08 - 0.43
adult	0.08 (0.06) 0.02 - 0.22	0.04 (0.02) 0.01 - 0.07	0.18 (0.14) 0.04 - 0.48

Table 19: Lobanov-normalized formant frequencies by age group: [a]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	1.06	0.48	-0.62	0.28	-0.38	1.04	NA	NA
5	1.37	1.15	-0.53	0.33	-0.12	0.92	NA	NA
6	1.58	0.55	-0.50	0.25	-0.38	0.82	NA	NA
7	1.29	1.09	-0.61	0.27	-0.53	1.05	NA	NA
8	1.30	0.33	-0.53	0.35	-0.13	1.37	NA	NA
9	1.63	0.66	-0.42	0.31	-0.68	0.76	NA	NA
10	1.68	0.62	-0.44	0.46	-0.42	1.07	NA	NA
adult	1.50	0.95	-0.33	0.24	-0.76	0.99	-0.58	0.74

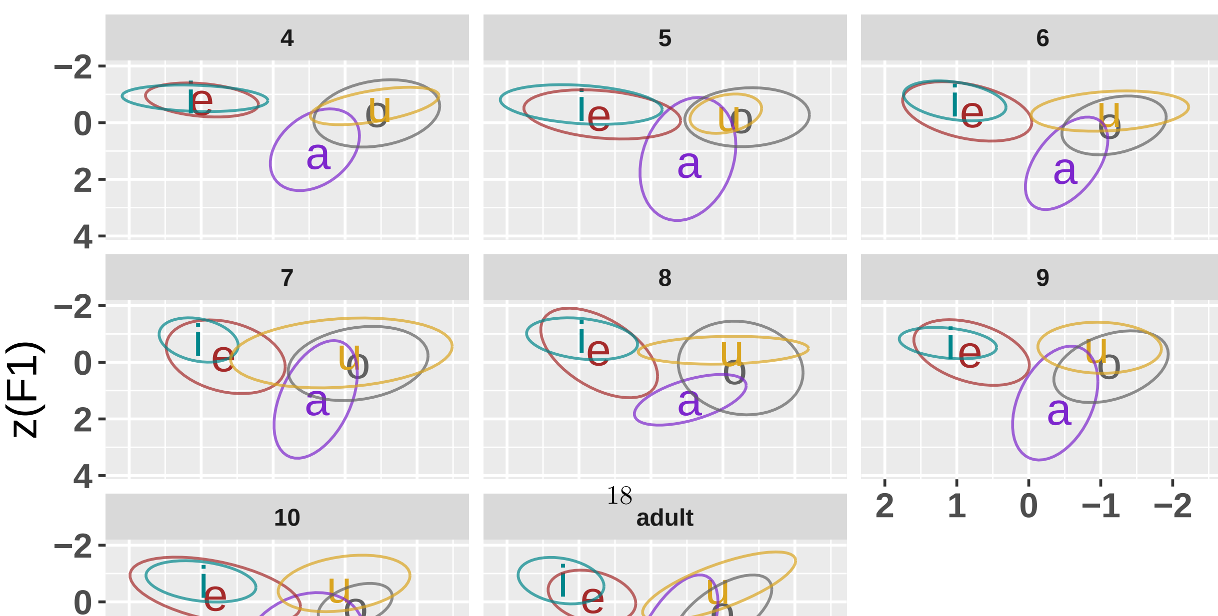
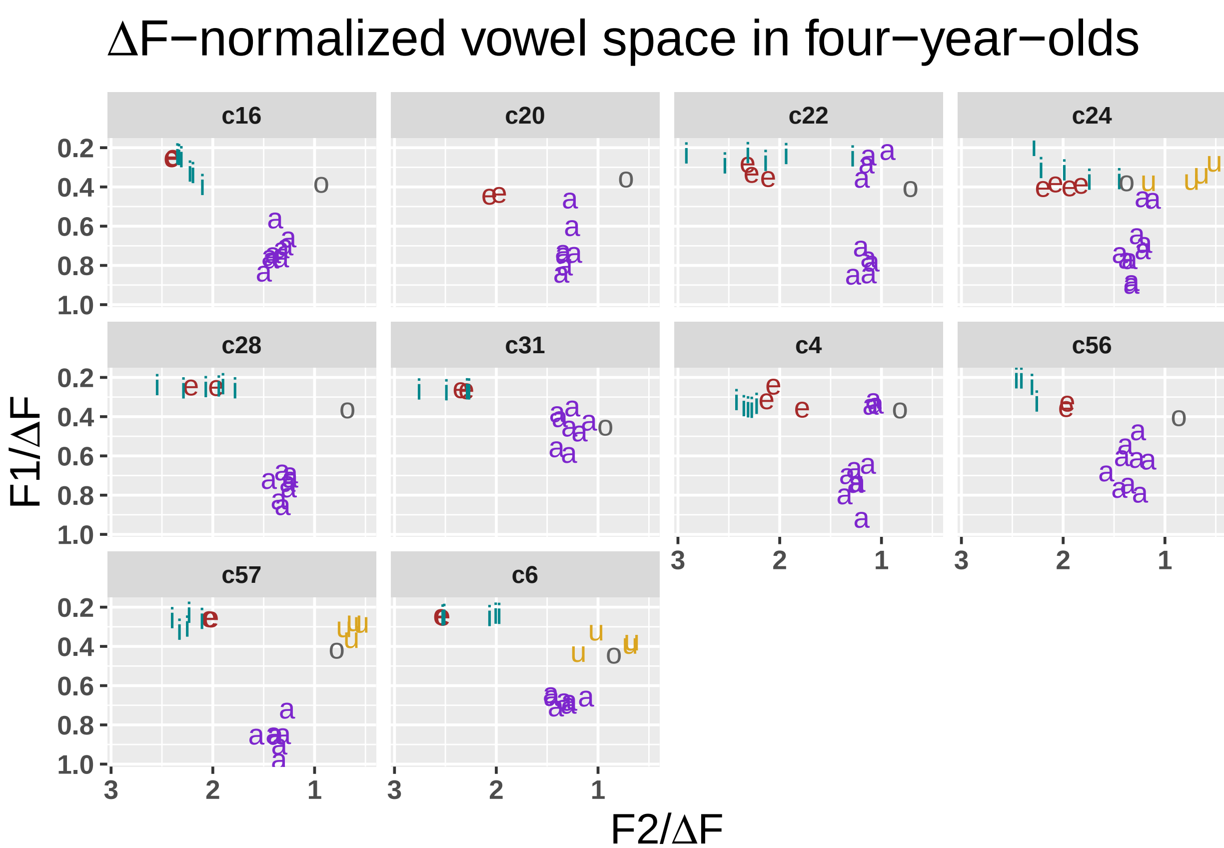


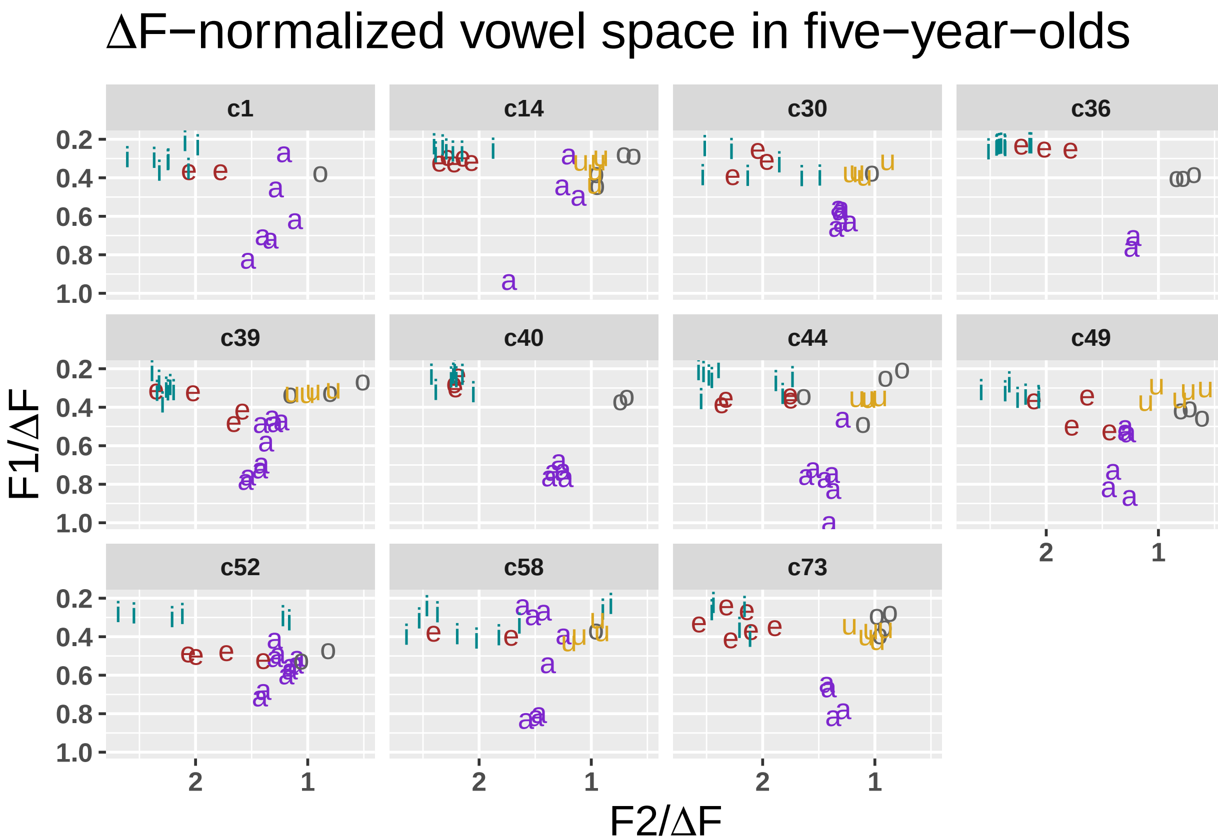
Table 20: Lobanov-normalized formant frequencies by age group: [i]

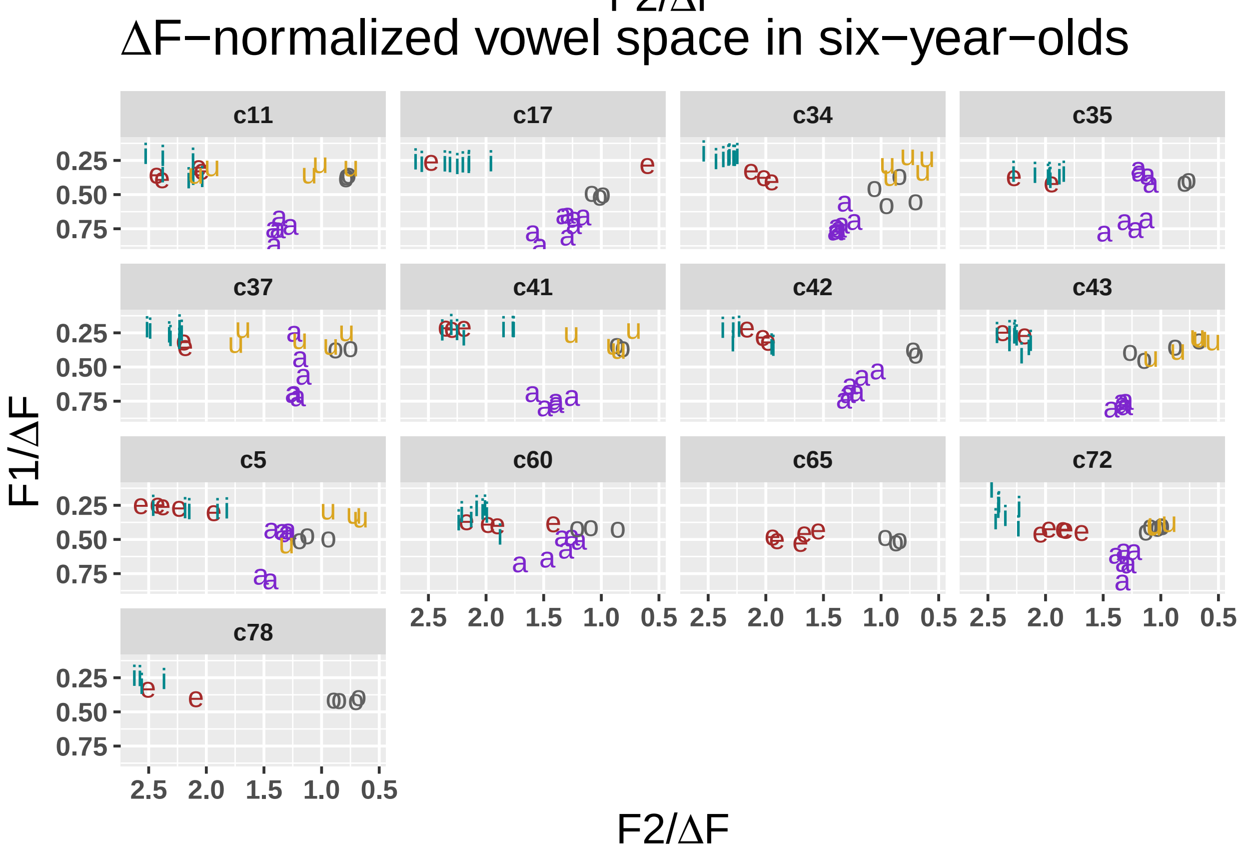
Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	-0.90	0.20	1.14	0.40	0.38	1.06	NA	NA
5	-0.63	0.26	0.96	0.52	0.38	1.10	NA	NA
6	-0.79	0.26	1.03	0.30	0.26	0.65	NA	NA
7	-0.80	0.34	1.04	0.21	0.41	0.91	NA	NA
8	-0.90	0.25	0.96	0.29	0.41	0.90	NA	NA
9	-0.68	0.22	1.09	0.25	0.51	0.73	NA	NA
10	-0.72	0.30	0.97	0.38	-0.03	1.01	NA	NA
adult	-0.76	0.37	1.22	0.28	0.52	0.86	0.67	0.68

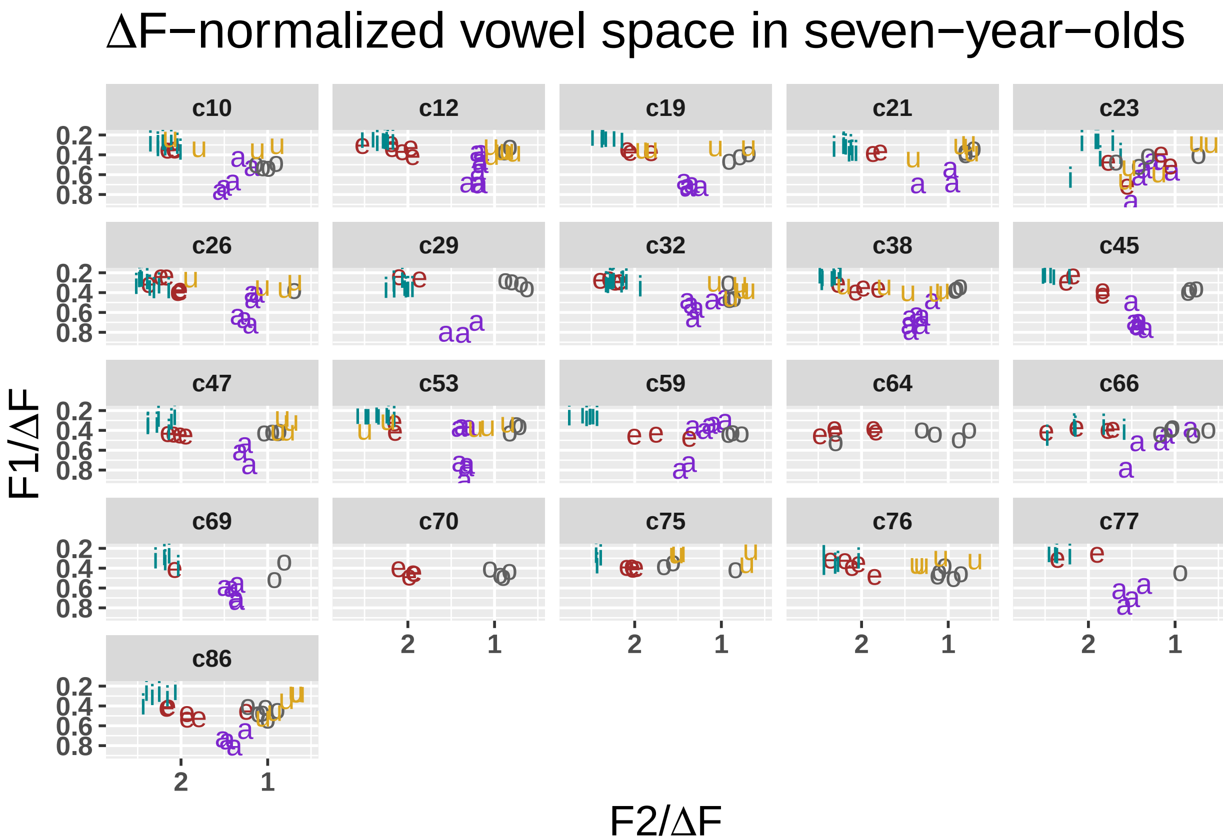
Table 21: Lobanov-normalized formant frequencies by age group: [u]

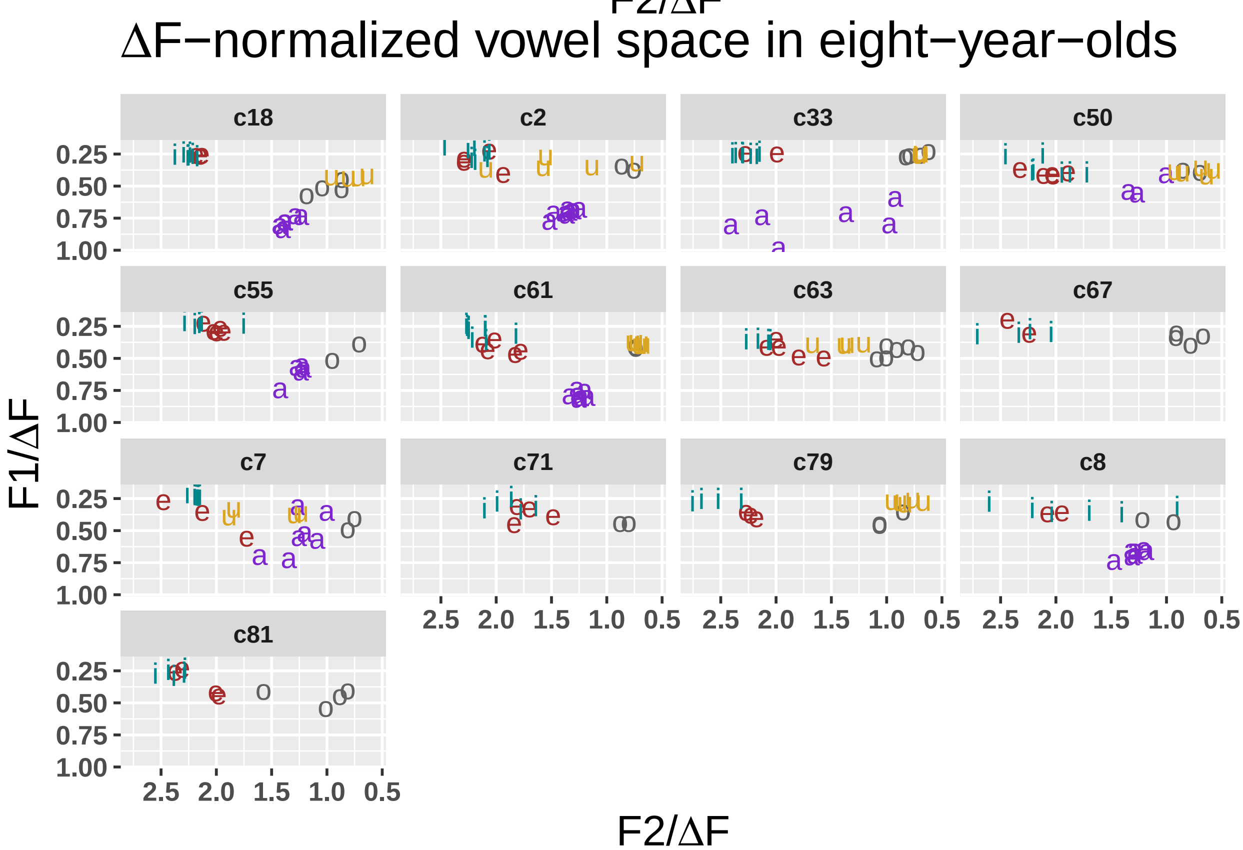
Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	-0.60	0.22	-1.48	0.21	-0.80	0.77	NA	NA
5	-0.28	0.31	-1.08	0.23	-0.83	1.29	NA	NA
6	-0.42	0.32	-1.12	0.45	-0.35	1.42	NA	NA
7	-0.32	0.49	-1.06	0.56	-0.09	1.17	NA	NA
8	-0.44	0.21	-1.12	0.34	-0.20	1.16	NA	NA
9	-0.55	0.31	-0.94	0.36	0.24	1.16	NA	NA
10	-0.55	0.50	-0.92	0.31	0.85	1.03	NA	NA
adult	-0.50	0.59	-0.93	0.54	-0.31	0.95	0.08	1.25

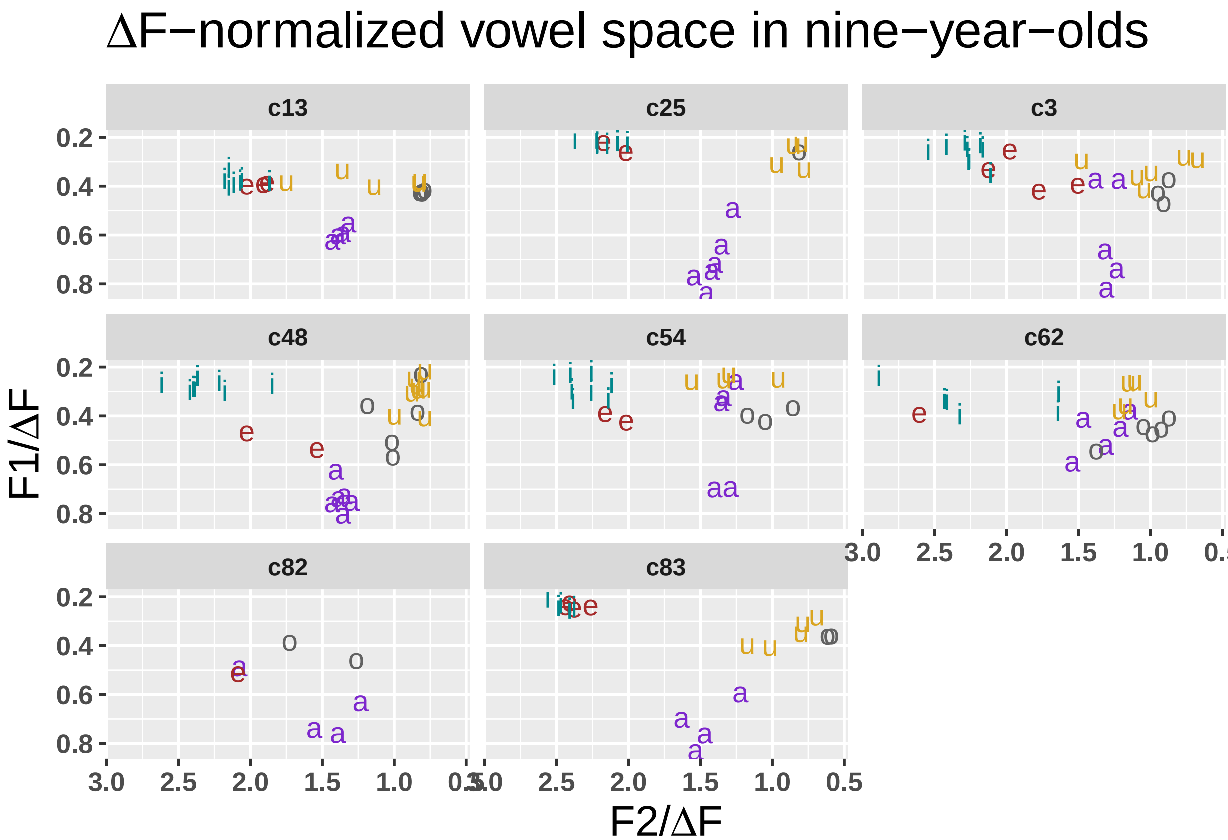


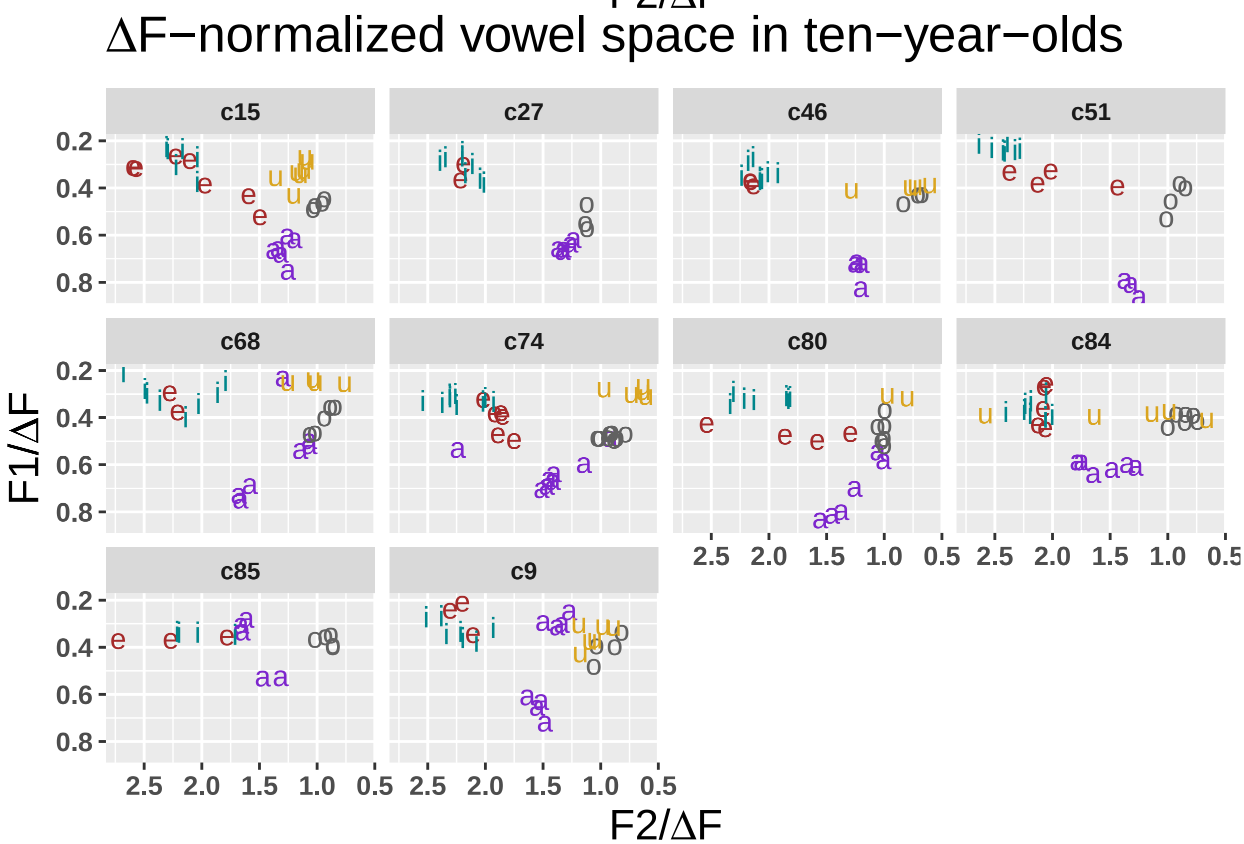

ΔF−normalized vowel space in five−year−olds
c1
c14
c30
c36
c39
c40
c44
c49
c52
c58
c73
F1/ΔF
F2/ΔF

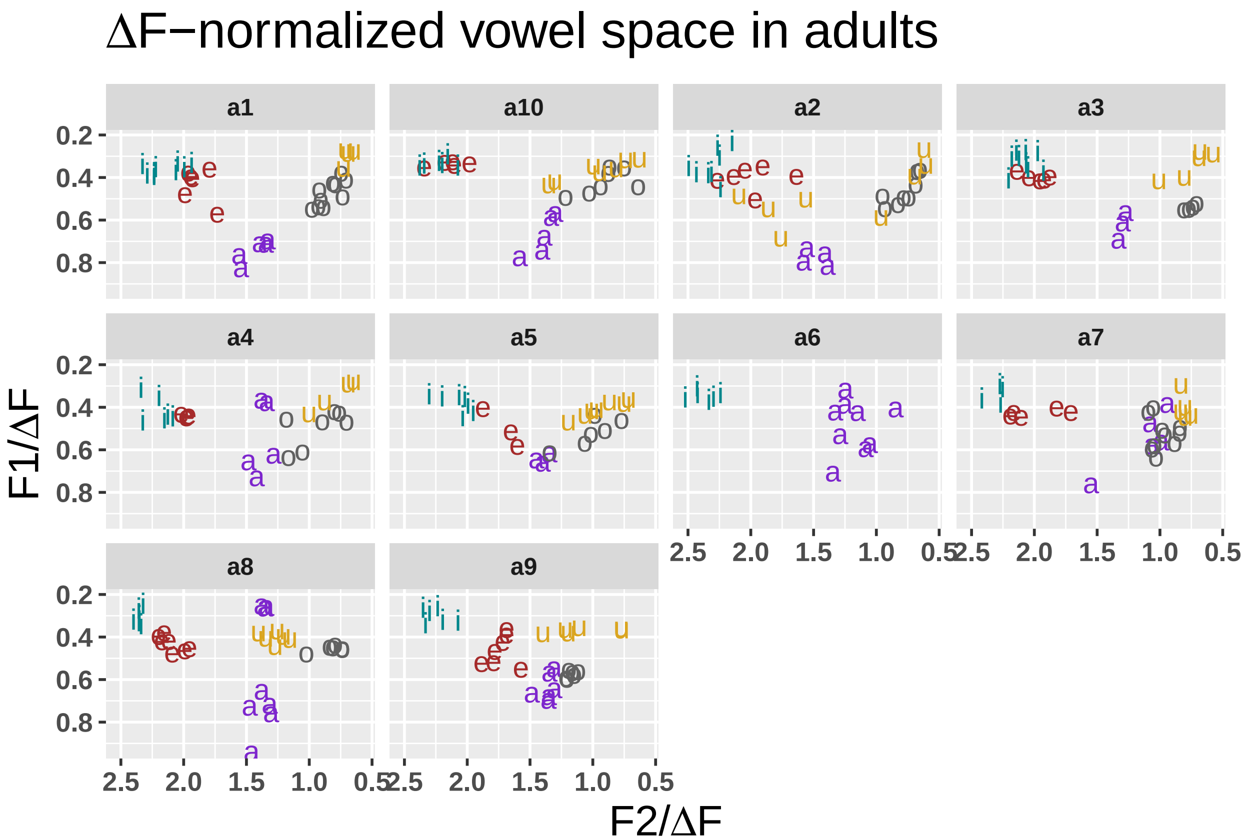

ΔF−normalized vowel space in six−year−olds
c11
c17
c34
c35
c37
c41
c42
c43
c5
c60
c65
c72
c78
F1/ΔF
F2/ΔF


ΔF−normalized vowel space in seven−year−olds
c10 c12 c19 c21 c23
c26 c29 c32 c38 c45
c47 c53 c59 c64 c66
c69 c70 c75 c76 c77
c86
F1/ΔF
F2/ΔF


ΔF−normalized vowel space in eight−year−olds
c18 c2 c33 c50
c55 c61 c63 c67
c7 c71 c79 c8
c81
F1/ΔF
F2/ΔF


ΔF−normalized vowel space in nine−year−olds
c13
c25
c3
c48
c54
c62
c82
c83
F1/ΔF
F2/ΔF


ΔF−normalized vowel space in ten−year−olds
c15
c27
c46
c51
c68
c74
c80
c84
c85
c9
F1/ΔF
F2/ΔF


ΔF−normalized vowel space in adults
a1
a10
a2
a3
a4
a5
a6
a7
a8
a9
F1/ΔF
F2/ΔF
0.2
0.4
0.6
0.8
2.5
2.0
1.5
1.0
0.5