

Chapter 1 Results: vocalic development

Meg Cychosz

10 June 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.

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 f_0 , 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.

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 categories 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

¹Note that this cleaning procedure was not conducted on the mid-vowels [e] and [o] due to data scarcity. Some mid-vowel

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

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)	2080.49 (932)	3959.64 (423)	NA (NA)	205
5	491.6 (187)	2258.95 (1009)	3887.07 (482)	NA (NA)	288
6	483.44 (209)	2152.89 (1114)	3711.44 (415)	NA (NA)	313
7	484.25 (143)	2338.16 (1053)	3811.39 (396)	NA (NA)	488
8	500.69 (217)	2229.63 (1073)	3660.5 (390)	NA (NA)	293
9	471.67 (179)	1916.46 (1094)	3668.35 (556)	NA (NA)	193
10	485.78 (159)	1948.64 (1022)	3501.85 (421)	NA (NA)	303
adult	460.91 (123)	1603.28 (931)	2952.27 (289)	4060.2 (375)	328

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	145-1290	783-3767	2607-4528	NA-NA
7	273-1163	794-3734	2501-4606	NA-NA
8	245-1216	808-3635	2526-4422	NA-NA
9	248-1219	847-3672	2237-4597	NA-NA
10	249-1076	796-3541	2630-4447	NA-NA
adult	262-1091	701-2929	2231-3668	2832-4726

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.

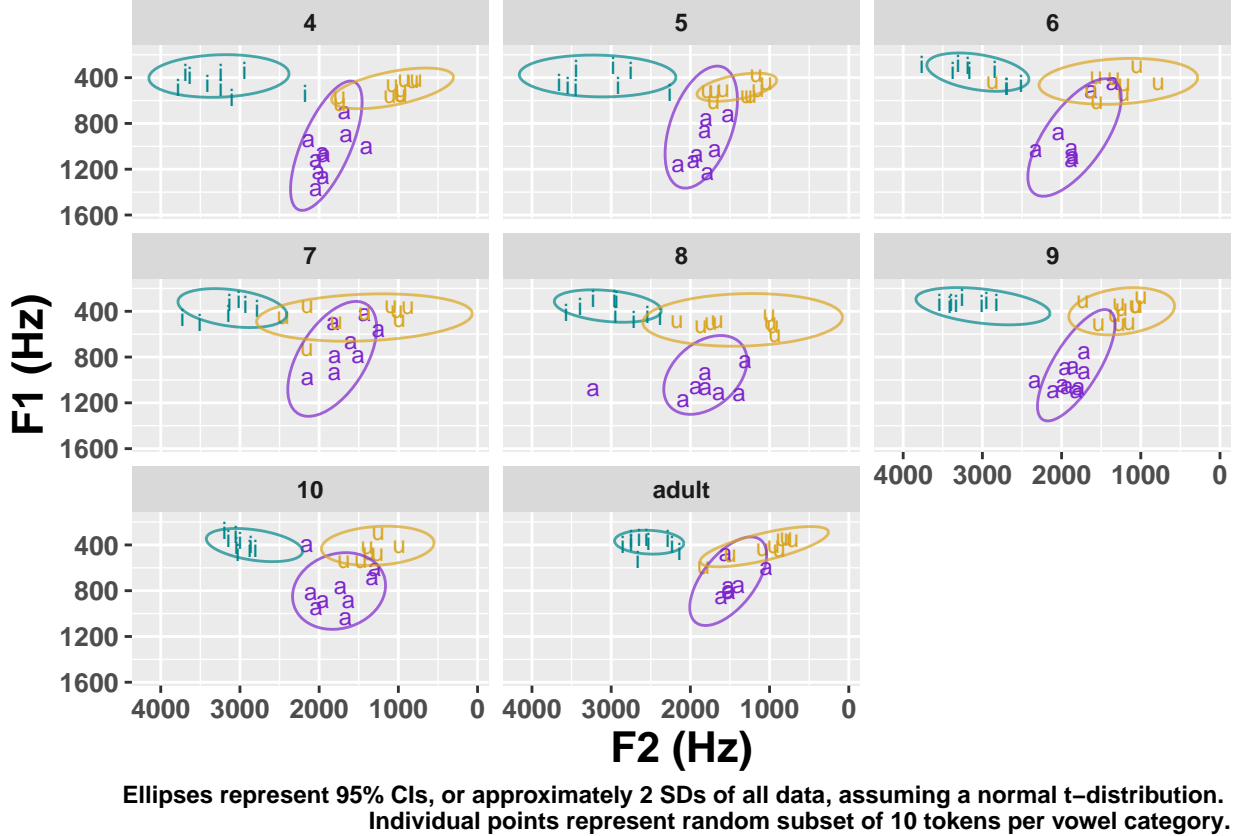


Figure 1: Vowel category development by age, in Hertz: adults and children

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)

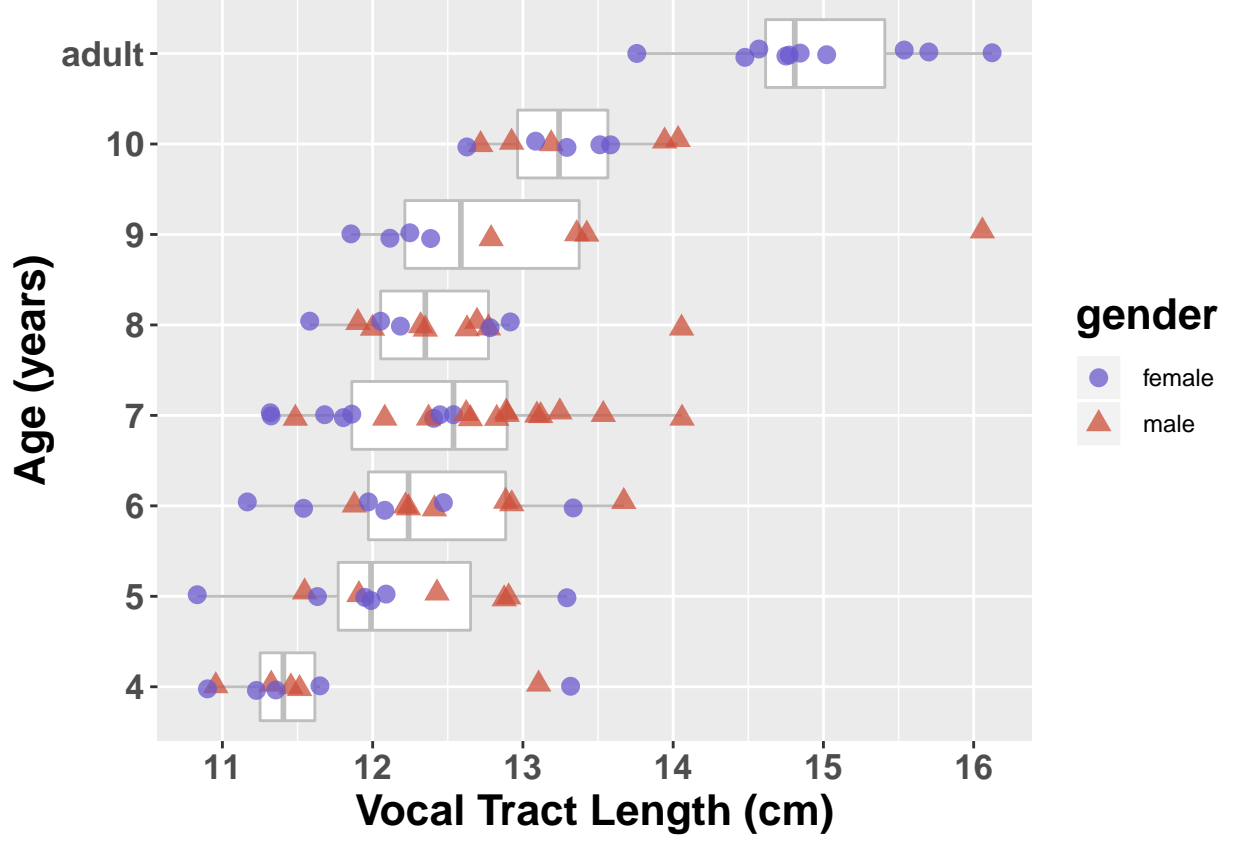


Figure 2: Average vocal tract length by age: adults and children

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.

1.1.3 Measuring within-category variability

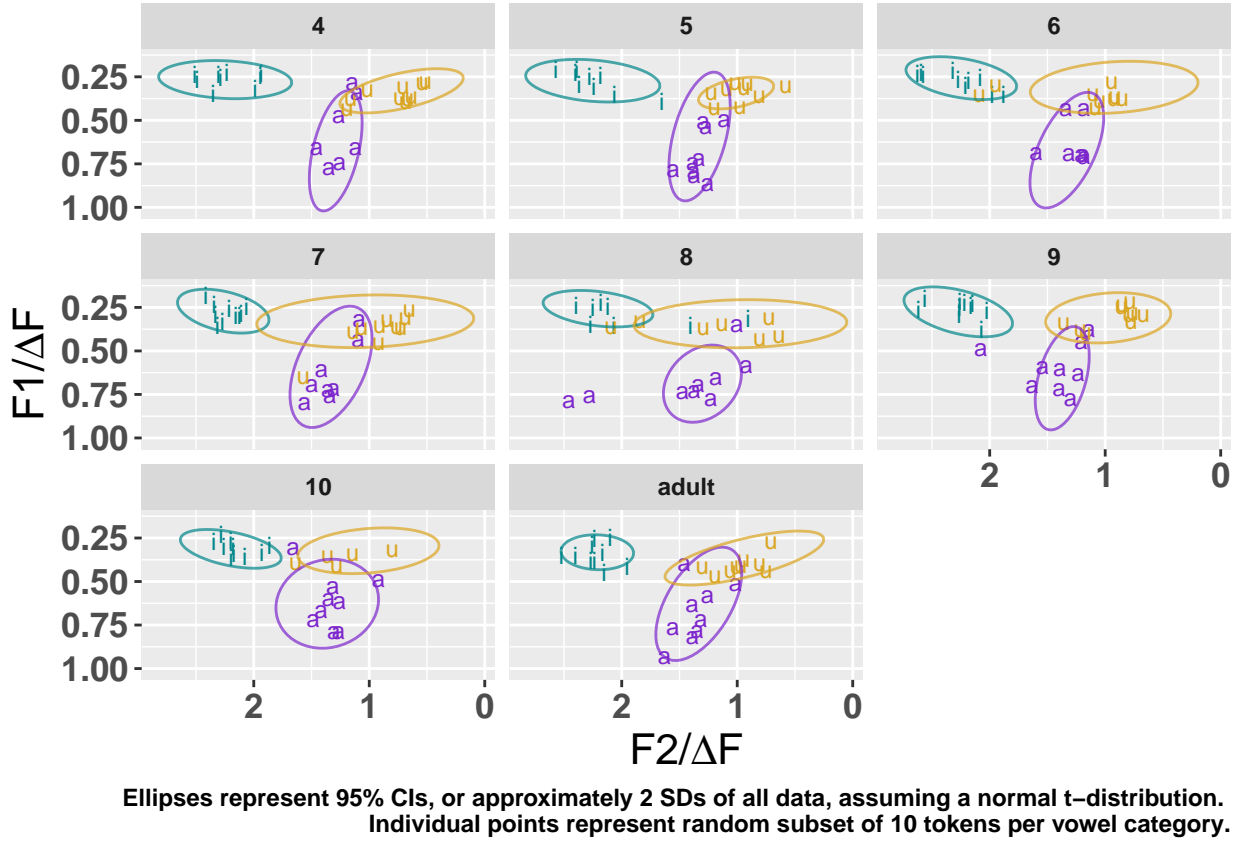


Figure 3: DeltaF-normalized vowels by age: adults and children

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 12 - 14 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, the root mean square (RMS) of each vowel category was computed, which resulted in a single coefficient per vowel category.² This RMS coefficient reflects both the *mean* value of each vowel category and its *variability* along the F1 and F2 dimensions. To calculate the RMS, the following steps were taken.

1. First, the mean value of F1 and F2, for each age group's vowel categories, was measured (e.g. the

²Previous work has used the Coefficient of Variation (CoV) to measure vowel category dispersion in children (e.g. Lee et al. (1999)). 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). One disadvantage of the CoV is that, unlike the RMS technique, separate coefficients must be computed for each acoustic dimension so the result is the CoV of F1, CoV of F2, etc.. Nevertheless, to facilitate comparison of this study with previous work, the CoV for each vowel category was also computed (for F1 and F2). In brief, the dispersion results measured via the CoV replicated the dispersion results measured via the RMS coefficient: the only reliable differences by age when measuring dispersion via CoV were for the F2 of [i]. Full dispersion results using the CoV, including descriptive statistics and statistical modeling for the CoV for each formant, are included in the appendices.

mean F1 and F2 of [a] for the five-year-olds). This step estimated the position of the vowel category in acoustic space.

2. Next, the difference between each individual formant measurement and its vowel category mean was computed and the difference was squared. This was done for F1 and F2. The resulting F1 and F2 coefficients were summed.
3. Finally, the square root of the sum was taken.

These steps were repeated for each vowel production. The mean value of all productions of a given vowel from each age group was then computed (e.g. five-year-olds' [a] productions). This mean value reflected the dispersion of the category in space. The result was a single variability coefficient for each vowel category, for each child age group, that factored in both formants.

Table 5: Vowel category dispersion by age group and phone

Age	[a] mean (SD) range	[i]	[u]
4	0.17 (0.11) 0.02 - 0.57	0.22 (0.2) 0.03 - 0.95	0.19 (0.13) 0.03 - 0.44
5	0.19 (0.1) 0.02 - 0.46	0.25 (0.25) 0.03 - 1.36	0.13 (0.08) 0.03 - 0.42
6	0.19 (0.11) 0 - 0.53	0.2 (0.17) 0 - 0.95	0.3 (0.25) 0.05 - 1.08
7	0.2 (0.1) 0.04 - 0.53	0.16 (0.12) 0.02 - 0.67	0.38 (0.28) 0 - 1.38
8	0.19 (0.21) 0.01 - 1.15	0.19 (0.2) 0.01 - 1.28	0.36 (0.22) 0.1 - 1.07
9	0.16 (0.13) 0.02 - 0.71	0.19 (0.15) 0.02 - 0.65	0.22 (0.14) 0.03 - 0.73
10	0.21 (0.15) 0.01 - 0.86	0.17 (0.12) 0.01 - 0.49	0.25 (0.27) 0.03 - 1.51
adult	0.2 (0.12) 0.02 - 0.59	0.13 (0.07) 0.01 - 0.32	0.29 (0.2) 0.03 - 1.1

Note:

4-10 observations per speaker

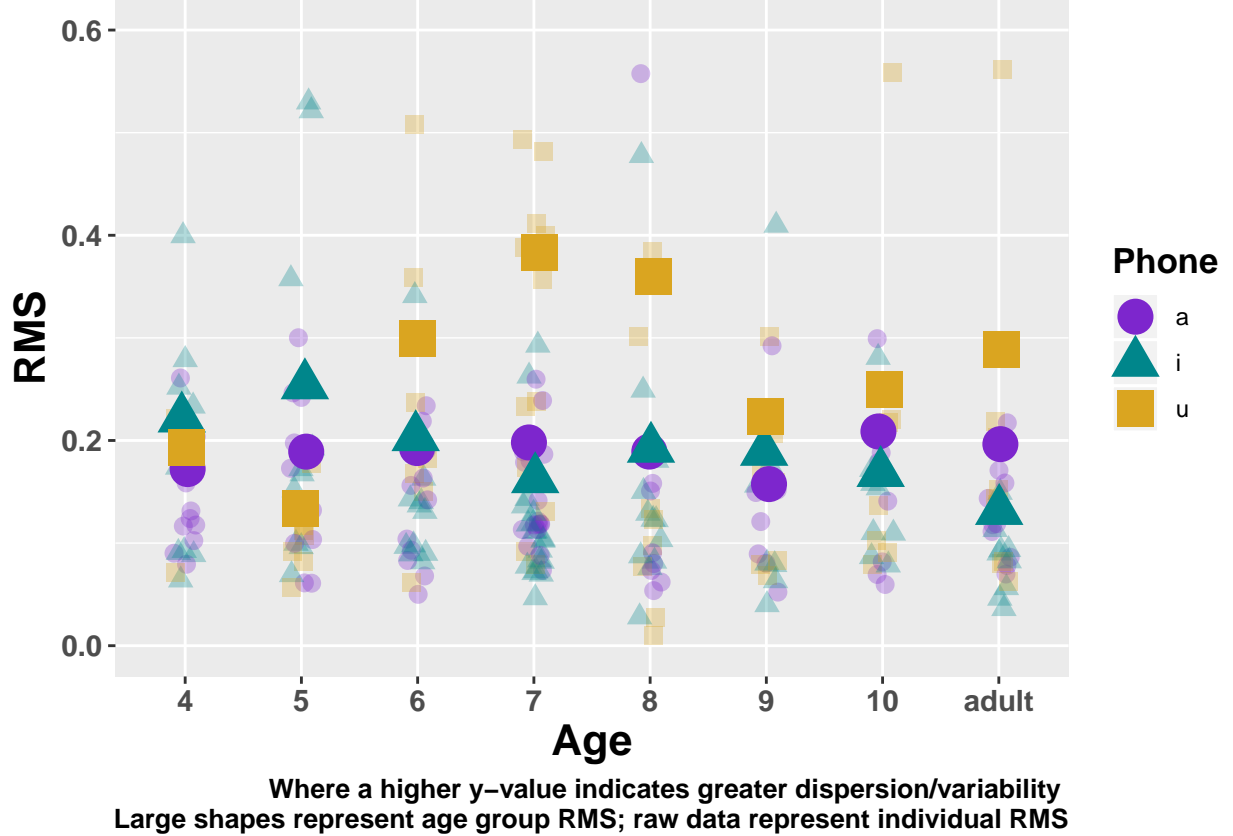


Figure 4: Group-level RMS by age and phone

The average RMS by phone and age group is listed in Table 5 and Figure 4. Recall that these measurements were made over vowel categories containing 4-10 observations per speaker.³ The results demonstrate that the RMS 1) varies by phone and 2) does not always decrease linearly with age. For [a], the RMS does not appear to decrease by age at all - the average RMS for [a] for the adults was 0.18, similar to the values for almost all the other age groups. The RMS for [u] is highest for the 7;0 and 8;0 groups, again not showing clear evidence of a linear decrease in variability. The RMS for [i] showed more change with age and was smallest in the adults (.13) and highest in the 4;0 and 5;0 groups (0.22 and 0.25, respectively).

The RMS of each vowel category was additionally computed on a by-speaker basis. This resulted in a single RMS coefficient for each vowel from each speaker. Computing the RMS on an individual speaker level generated a datapoint for each speaker’s three vowel categories which were used to fit a series of models predicting vowel category dispersion.

1.1.4 Fitting models to predict within-category variability

For the RMS model fitting, we fit generalized linear mixed effects models (GLMMs) because the outcome variable - the individual speaker level RMS - was necessarily non-negative and left-skewed. Gamma GLMMs were fit using a log linking function to appropriately model the skewed, non-Gaussian distribution of the residual. Three models were fit, one for each 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 parameter **Age Group** (4, 5, 6, 7, 8, 9, 10, adult) was added.

³RMS measurements computed over *all* data points, including those categories with less than 4 data points, are included in Table 15 in the appendices. Overall there were no large differences observed between the RMS values computed over 4-10 tokens per vowel category and those computed over the entire dataset. However, standardizing the number of tokens per category avoids skewing in the event of data loss.

Table 6: Model predicting RMS for [a]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-2.11	0.15	-14.19	0.00	-1.82,-2.4
10	0.19	0.21	0.88	0.38	0.6,-0.23
4	0.07	0.21	0.33	0.74	0.48,-0.34
5	0.14	0.21	0.67	0.50	0.54,-0.26
6	0.01	0.21	0.03	0.97	0.41,-0.4
7	0.12	0.19	0.64	0.52	0.49,-0.25
8	-0.10	0.22	-0.45	0.65	0.34,-0.54
9	0.00	0.22	-0.01	0.99	0.44,-0.44

Table 7: Model predicting RMS for [i]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-2.49	0.17	-15.06	0.00	-2.17,-2.82
10	0.54	0.23	2.31	0.02	1,0.08
4	0.62	0.24	2.58	0.01	1.09,0.15
5	0.79	0.23	3.46	0.00	1.24,0.34
6	0.45	0.22	2.03	0.04	0.89,0.01
7	0.24	0.20	1.18	0.24	0.64,-0.16
8	0.42	0.22	1.91	0.06	0.85,-0.01
9	0.27	0.26	1.06	0.29	0.78,-0.23

Table 8: Model predicting RMS for [u]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-1.77	0.23	-7.59	0.00	-1.31,-2.23
10	0.11	0.33	0.32	0.75	0.75,-0.53
4	-0.05	0.43	-0.10	0.92	0.81,-0.9
5	-0.45	0.34	-1.33	0.18	0.21,-1.12
6	0.31	0.33	0.94	0.35	0.95,-0.33
7	0.59	0.28	2.10	0.04	1.15,0.04
8	-0.27	0.33	-0.81	0.42	0.38,-0.91
9	-0.12	0.33	-0.38	0.71	0.52,-0.77

The parameter **Age Group** was not significant for the model predicting the RMS of [a].⁴ For the [u] model, only one child age group, the 7;0 group, differed significantly from the adults: the 7;0 had a significantly larger [u] category than the adults. The parameter **Age Group** did improve upon the model predicting the RMS of [i]. The model summaries for each phone model, with the **Age Group** parameter added, are presented in Tables 6-8.

As the beta coefficients in the [i] model summary demonstrates, there were some differences by age group in category dispersion for the [i] phoneme. Specifically, the adults had significantly tighter [i] categories than all children except for the 7;0 and 9;0 groups (the difference between the adults and 7;0 only approached significance). Furthermore, as the beta coefficients demonstrate, the children's RMS values did not necessarily decrease linearly with age - the 10;0 group, for example, actually had significantly more dispersed categories than the 9;0 group.

Overall, this analysis shows that there are not significant differences in category dispersion between adults and children for [a]. For [u], only the 7;0 group had a significantly larger category than the adults. There are significant differences in category dispersion for [i]: the adults have tighter categories than many of the child age groups. However, the expected trend of reduced dispersion with age is not observed.

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 results in similar formant measurements, 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 were normalized using two formant scaling techniques - one uniform (ΔF) and one non-uniform (Lobanov), as previously described. 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 12-14 in the appendices, and Figure 3 in the previous section, for the ΔF -normalized formant values.

⁴The baseline models for [a] and [i] did not converge, but the models with the addition of **Age Group** did. Consequently, potential differences between age groups were determined solely from model summary outputs for those two models.

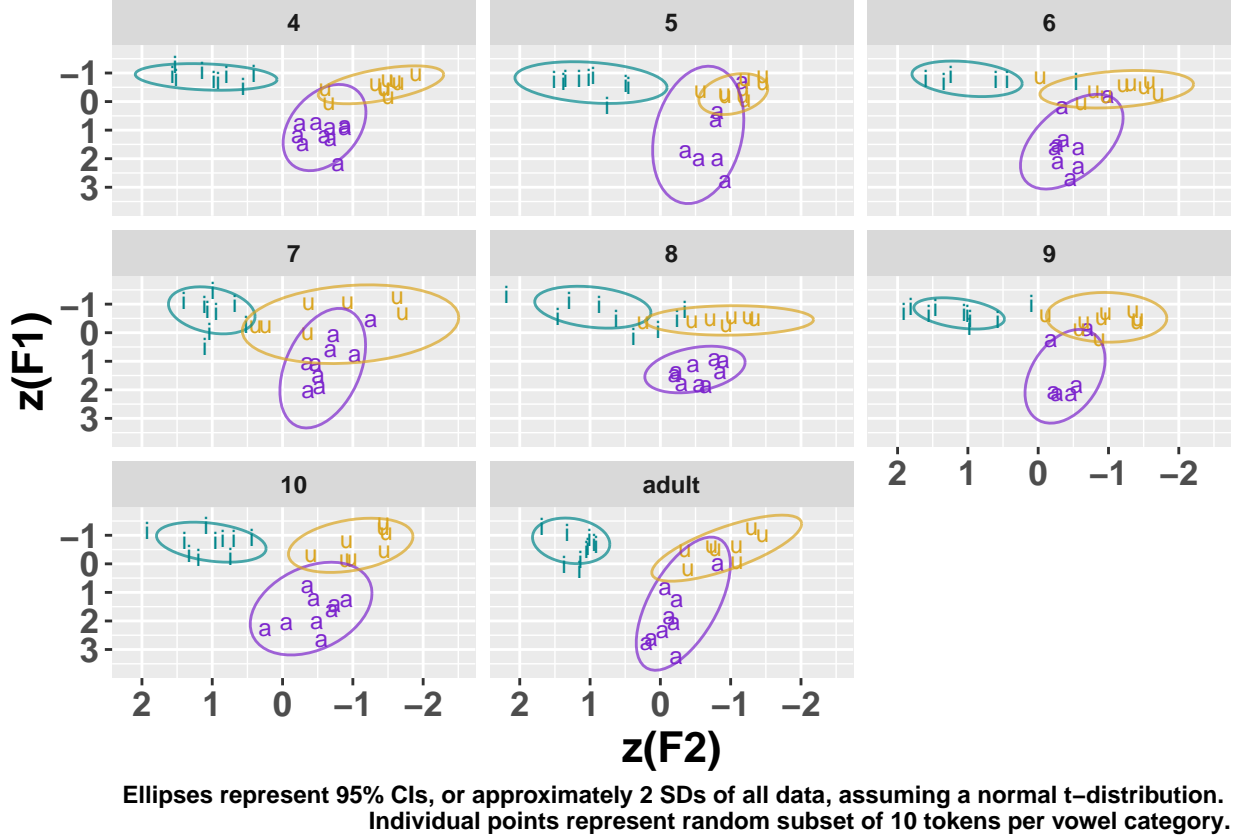


Figure 5: Lobanov-normalized vowels by age: adults and children

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) by computing the RMS coefficient, as described in the previous section. This measurement was made separately to the ΔF -normalized vowels and the Lobanov-normalized vowels.

The RMS measurement was again used to calculate the difference in dispersion between adults and children. As before, the mean value of each formant, for each vowel category and age group, was computed. Then, the difference between the mean value for the adults and a given child age group was measured. For example, the mean F1 of [a] was measured over all of the 5;0 children and the mean F1 of [a] over all of the adults and then the difference between these values was taken. As in the previous section, this step estimated the position of the vowel category in acoustic space. The resultant F1 and F2 measurements were both squared, and then summed. Finally, the square root of the sum was taken.

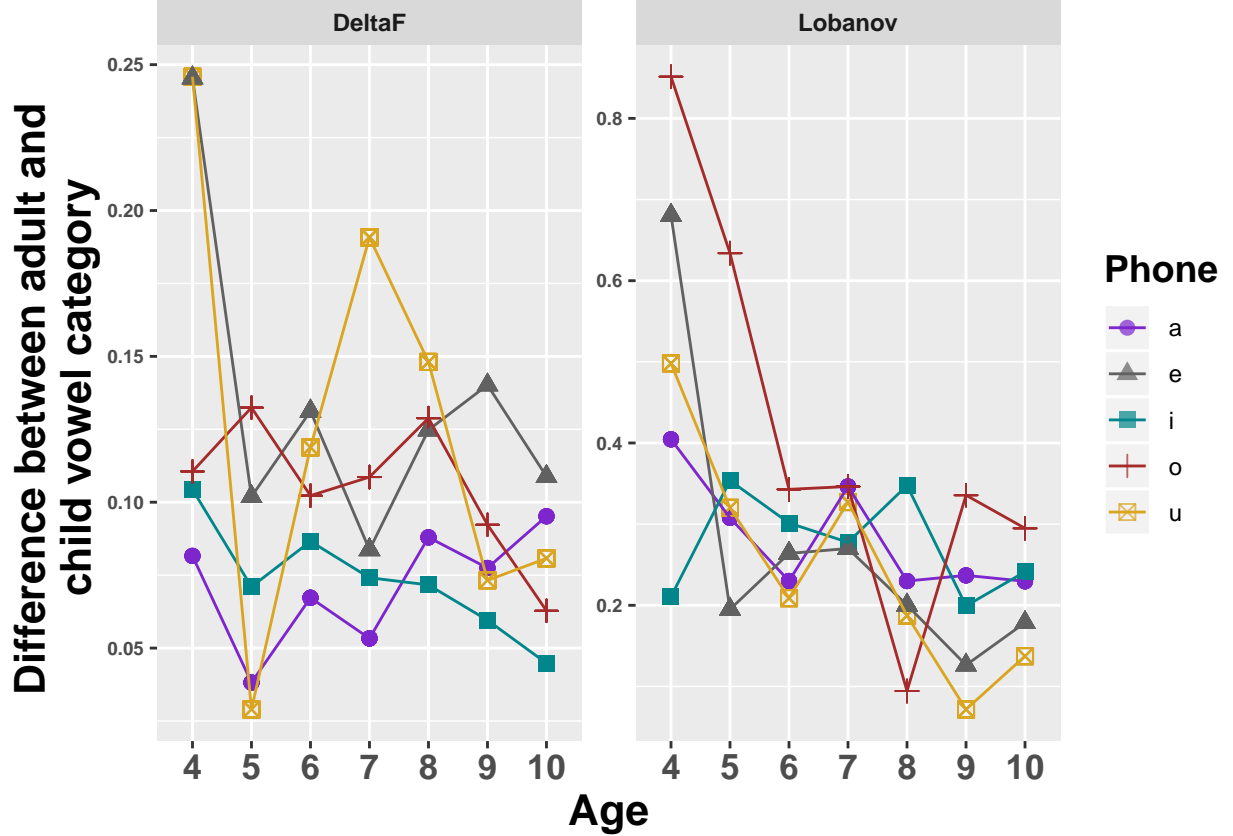


Figure 6: Difference between adult and child vowel categories: RMS method

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 take 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. However, Turner et al. (2009) concluded that this outstanding variability was due to measurement error in formant tracking in spectrograms and the use of LPC. The authors concluded that any remaining variability beyond that was likely to be 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 might 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.

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 adult-like acoustics). The rest of the children were classified 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 phone identity 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 would be 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 relative to the 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 entirely the issue of affiliation (Apostol et al., 2004).

Finally, we chose F1 to F2 and F2 to F3 ratios 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. To do that, the hypotheses concerning the formant ratios were evaluated on two data sets containing formant measurements from adult speakers. 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 because that ratio appears to be the most reliable indicator of cavity size.

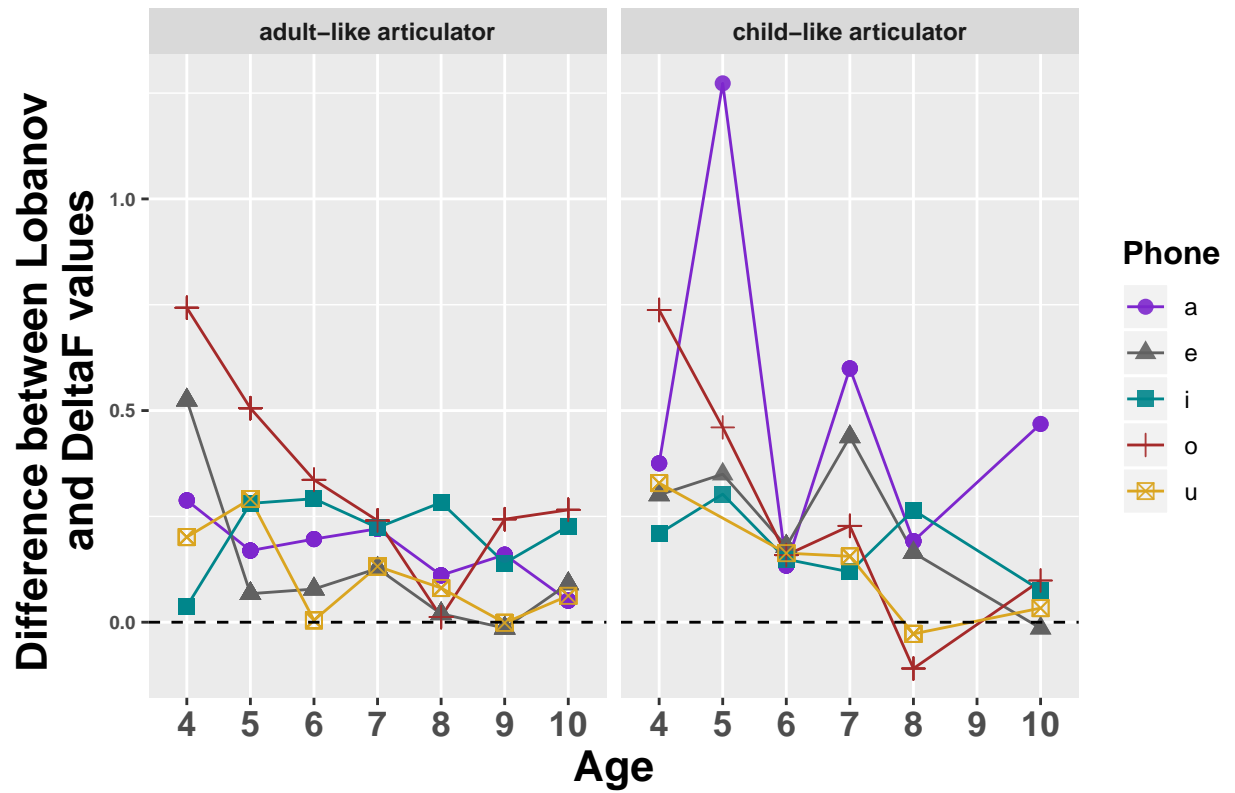
Having confirmed that the formant ratios do reflect differences in cavity size, we turned to the calculations of formant ratios for the children in the current study. To do this, we calculated the mean F1, F2, and F3 for [a] for each child. (Participant c64 only had one valid [a] token, so formant ratios 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 be maximally conservative in the classification of the children.

Table 9 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 9: 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



No children in the 9;0 group were classified as child-like articulators.
 The 5;0 child-like articulator did not have 4 formant observations for [u] so those data are not reported.

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

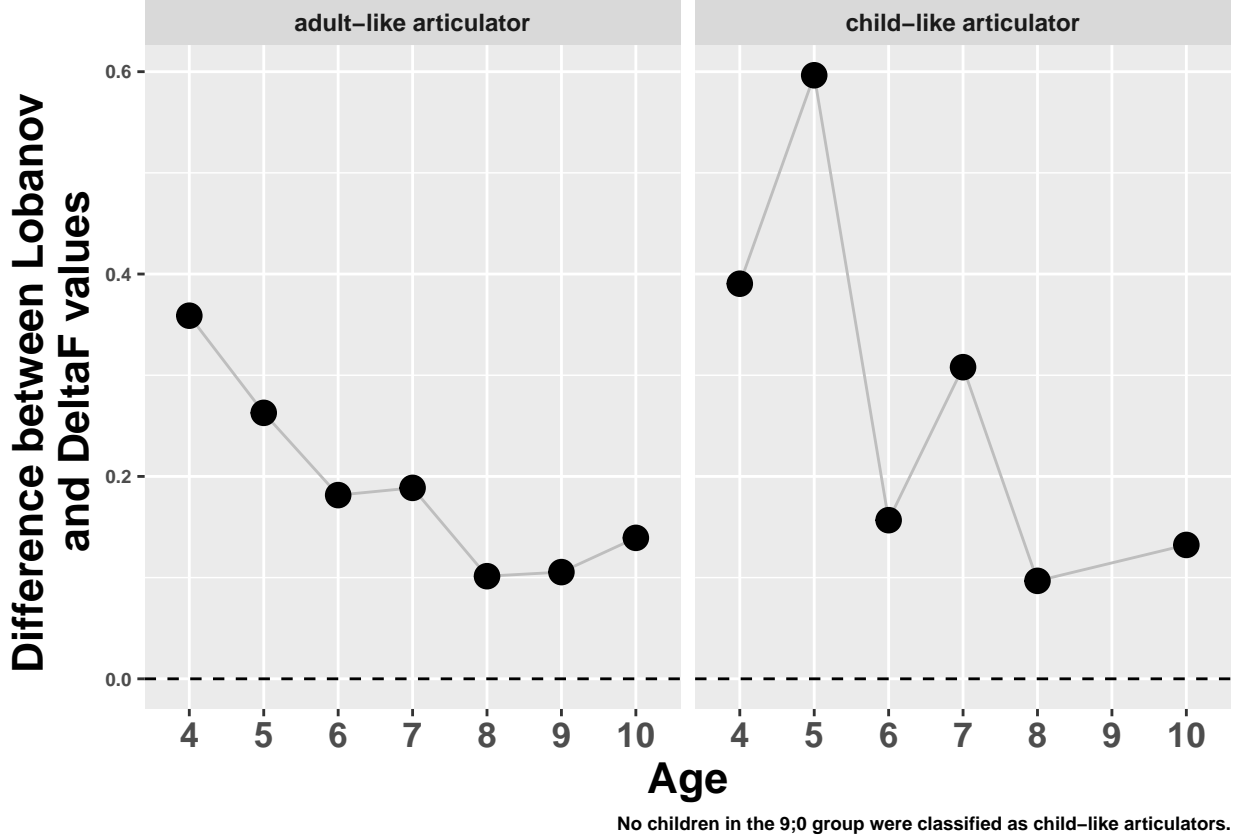


Figure 8: Average difference between adult and child vowel categories by scaling method and articulatory status.

1.2.2 Fitting models to un-normalized formant frequency data

As a final step to determine what accounts for the outstanding variation in the formant data, I built statistical models to predict the values of unnormalized 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 formant frequency variance. 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**. 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 Tract Length**, and **F1:F2 ratio for [a]**. The final model summary for F1 is listed in Table 10 and the model summary for F2 is in Table 11. This result for the models suggests that there is additional variance in

Table 10: 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 11: 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

the formant frequency measures beyond phone identity and vocal tract length - something that Turner et al. (2009) likewise concluded. 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 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 some school-aged children 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

2.1 Descriptive statistics of Delta-F-normalized formants

Table 12: DeltaF-normalized formant frequencies by age group: [a]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	0.70	0.11	1.28	0.10	2.63	0.26	NA	NA
5	0.56	0.20	1.31	0.12	2.70	0.31	NA	NA
6	0.69	0.12	1.33	0.13	2.72	0.30	NA	NA
7	0.61	0.16	1.34	0.16	2.67	0.28	NA	NA
8	0.69	0.11	1.29	0.11	2.58	0.38	NA	NA
9	0.64	0.14	1.38	0.09	2.59	0.29	NA	NA
10	0.61	0.11	1.35	0.19	2.60	0.34	NA	NA
adult	0.61	0.13	1.34	0.13	2.44	0.22	3.42	0.21

Table 13: 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	0.05	2.24	0.22	2.81	0.23	NA	NA
6	0.25	0.06	2.25	0.18	2.79	0.19	NA	NA
7	0.26	0.06	2.27	0.18	2.83	0.21	NA	NA
8	0.25	0.04	2.19	0.16	2.77	0.26	NA	NA
9	0.26	0.06	2.22	0.22	2.79	0.29	NA	NA
10	0.31	0.05	2.21	0.23	2.79	0.27	NA	NA
adult	0.33	0.05	2.20	0.18	2.71	0.13	3.68	0.25

Table 14: 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	0.06	1.04	0.22	2.92	0.39	NA	NA
adult	0.37	0.07	0.91	0.33	2.54	0.23	3.52	0.38

Table 15: Vowel category dispersion by age group and phone: all data

Age	[a] mean (SD) range	[i]	[u]
4	0.17 (0.11) 0.02 - 0.57	0.22 (0.2) 0.03 - 0.94	0.15 (0.12) 0.03 - 0.42
5	0.2 (0.11) 0.01 - 0.56	0.23 (0.21) 0.03 - 1.36	0.13 (0.09) 0.02 - 0.42
6	0.19 (0.11) 0.02 - 0.55	0.19 (0.16) 0.01 - 0.97	0.32 (0.26) 0.04 - 1.08
7	0.19 (0.09) 0.01 - 0.54	0.18 (0.12) 0.02 - 0.67	0.38 (0.27) 0.01 - 1.37
8	0.21 (0.23) 0.01 - 1.12	0.16 (0.17) 0.01 - 1.28	0.36 (0.23) 0.1 - 1.07
9	0.16 (0.11) 0.02 - 0.72	0.17 (0.13) 0.02 - 0.64	0.22 (0.15) 0.02 - 0.75
10	0.19 (0.13) 0 - 0.89	0.16 (0.11) 0.02 - 0.5	0.21 (0.14) 0.02 - 0.65
adult	0.18 (0.12) 0.01 - 0.63	0.12 (0.07) 0.01 - 0.3	0.29 (0.21) 0.03 - 1.11

Table 16: Average CoV by formant and vowel for children and adults

Age	F1			F2		
	[a]	[i]	[u]	[a]	[i]	[u]
4	0.26	0.18	0.15	0.09	0.13	0.30
5	0.30	0.19	0.14	0.10	0.16	0.15
6	0.23	0.22	0.22	0.12	0.12	0.38
7	0.26	0.23	0.23	0.13	0.09	0.42
8	0.17	0.19	0.18	0.18	0.13	0.41
9	0.22	0.22	0.20	0.11	0.10	0.26
10	0.22	0.17	0.17	0.16	0.09	0.34
adult	0.27	0.15	0.22	0.13	0.07	0.34

2.2 Dispersion measured via CoV

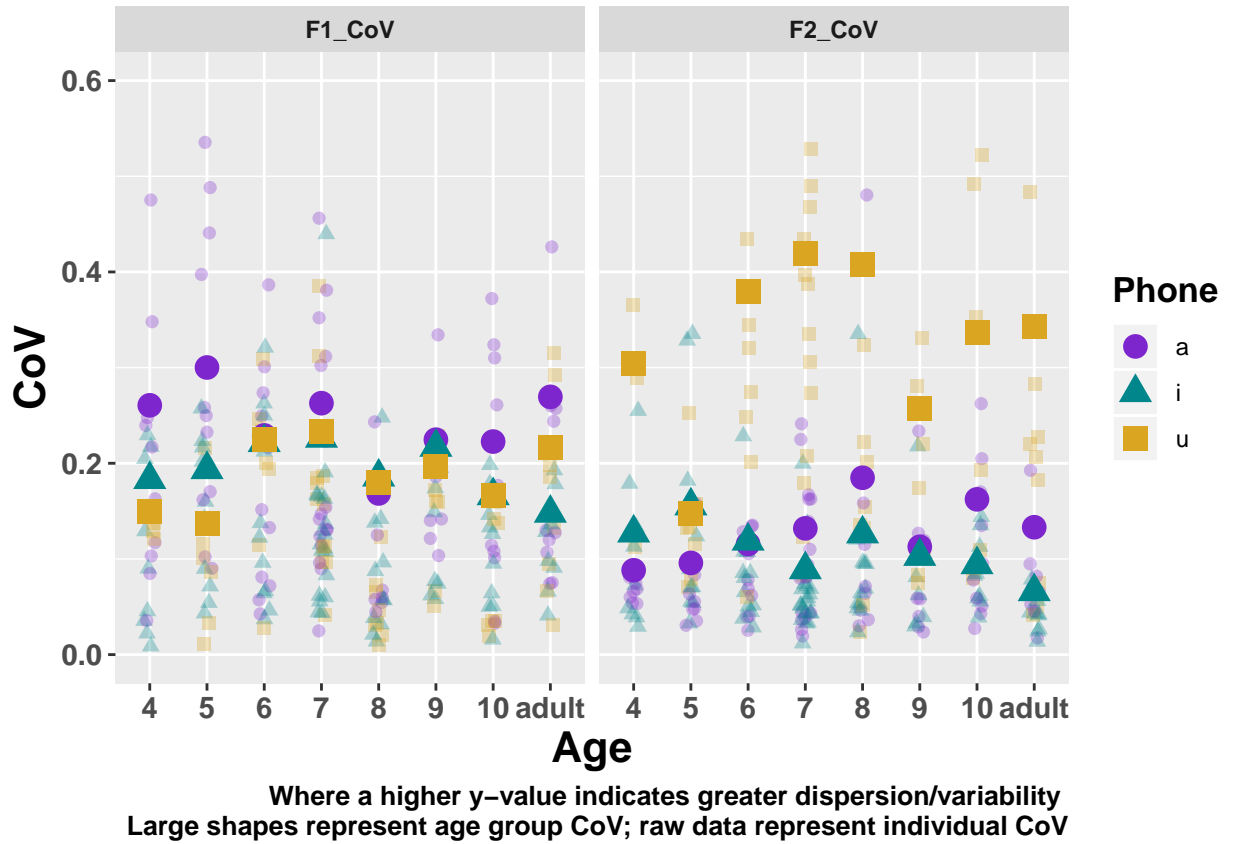


Figure 9: Group-level RMS by age and phone

Table 17: Model predicting CoV for F2 of [i]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-3.23	0.19	-17.22	0.00	-2.87,-3.6
10	0.71	0.27	2.68	0.01	1.23,0.19
4	0.69	0.27	2.54	0.01	1.23,0.16
5	0.93	0.26	3.60	0.00	1.44,0.43
6	0.57	0.25	2.24	0.03	1.07,0.07
7	0.31	0.23	1.33	0.18	0.76,-0.15
8	0.69	0.25	2.76	0.01	1.18,0.2
9	0.49	0.29	1.66	0.10	1.06,-0.09

Table 18: Model predicting CoV for F1 of [u]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-1.77	0.20	-8.63	0.00	-1.37,-2.17
10	-0.76	0.31	-2.46	0.01	-0.15,-1.37
4	-0.28	0.41	-0.68	0.50	0.52,-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	-4.10	0.00	-0.64,-1.81
9	-0.14	0.31	-0.46	0.64	0.46,-0.75

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

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	1.00	0.48	-0.64	0.28	-0.37	1.04	NA	NA
5	1.05	1.22	-0.56	0.35	-0.16	0.92	NA	NA
6	1.54	0.60	-0.48	0.32	0.10	1.20	NA	NA
7	1.30	1.03	-0.58	0.25	-0.48	1.08	NA	NA
8	1.29	0.33	-0.48	0.29	-0.37	1.17	NA	NA
9	1.56	0.77	-0.42	0.29	-0.52	0.78	NA	NA
10	1.55	0.67	-0.44	0.46	-0.48	0.96	NA	NA
adult	1.37	0.90	-0.28	0.31	-0.69	1.04	-0.41	0.84

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

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	-0.88	0.18	1.14	0.40	0.33	1.15	NA	NA
5	-0.64	0.26	0.98	0.43	0.24	1.01	NA	NA
6	-0.82	0.22	0.99	0.37	0.21	0.71	NA	NA
7	-0.81	0.32	1.04	0.24	0.39	0.91	NA	NA
8	-0.94	0.28	0.97	0.34	0.41	0.95	NA	NA
9	-0.68	0.18	1.05	0.22	0.50	0.76	NA	NA
10	-0.78	0.30	0.97	0.37	0.02	1.08	NA	NA
adult	-0.79	0.42	1.30	0.28	0.52	0.77	0.63	0.77

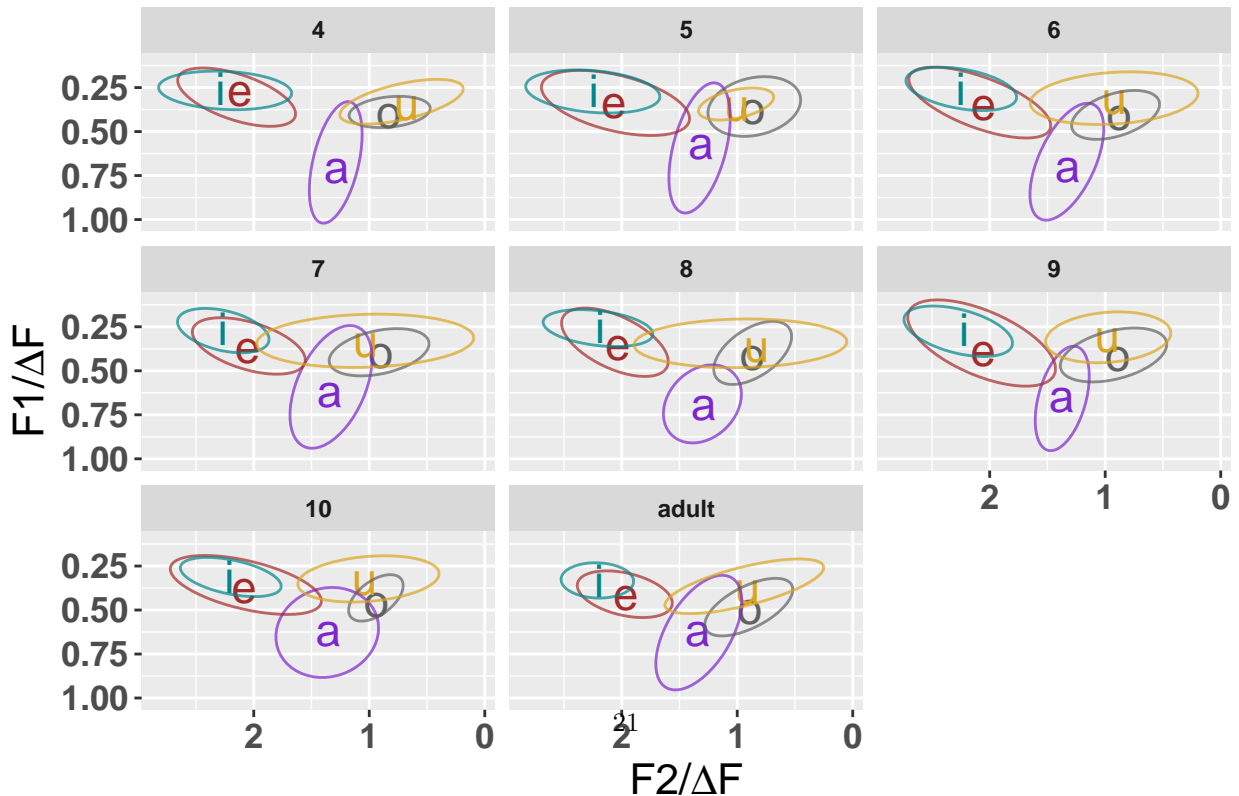
Table 21: Lobanov-normalized formant frequencies by age group: [u]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	-0.55	0.27	-1.47	0.21	-0.79	0.67	NA	NA
5	-0.24	0.35	-1.05	0.21	-0.84	1.37	NA	NA
6	-0.42	0.30	-1.15	0.35	-0.51	1.34	NA	NA
7	-0.29	0.62	-1.09	0.53	-0.14	1.24	NA	NA
8	-0.46	0.23	-1.03	0.41	-0.20	1.14	NA	NA
9	-0.57	0.34	-0.93	0.36	0.17	1.21	NA	NA
10	-0.58	0.42	-0.90	0.30	0.67	1.25	NA	NA
adult	-0.48	0.54	-0.95	0.51	-0.29	0.86	0.07	1.24

3 Models predicting dispersion measured via CoV

3.1 Descriptive statistics of Lobanov-normalized formants

3.2 Vowel plots



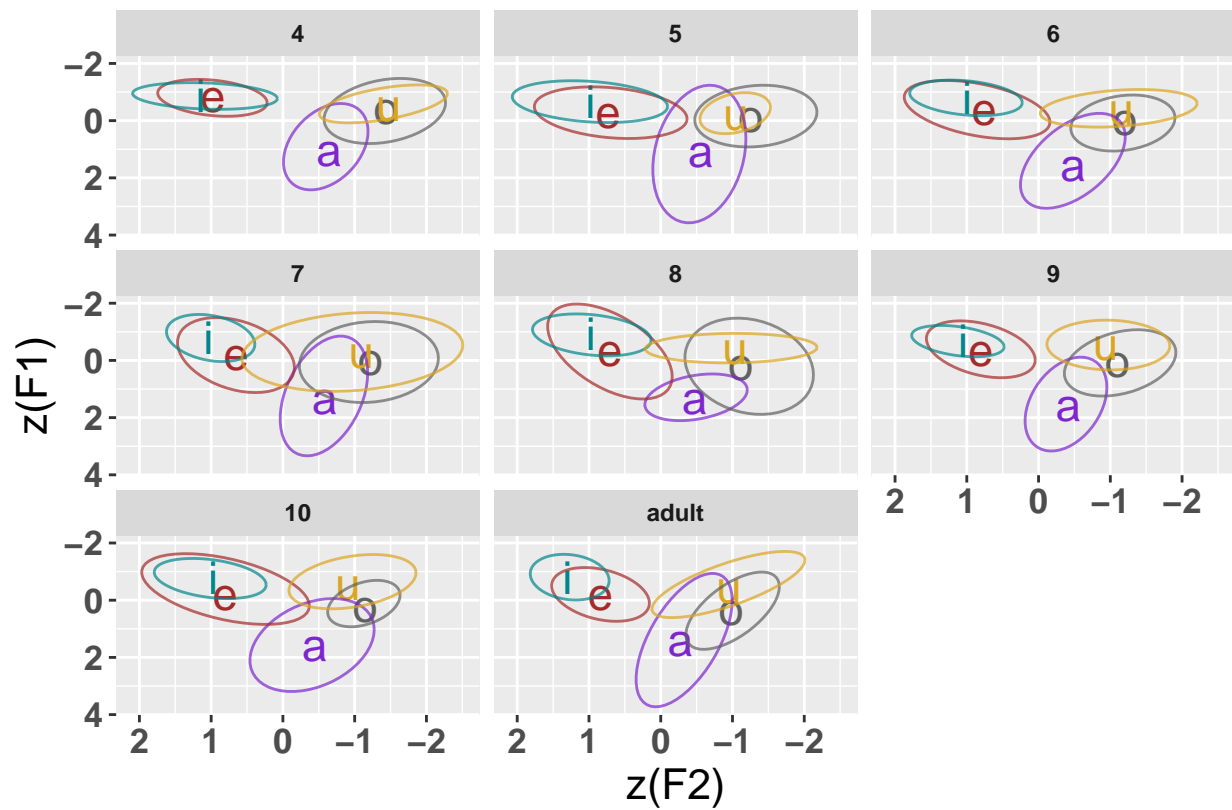
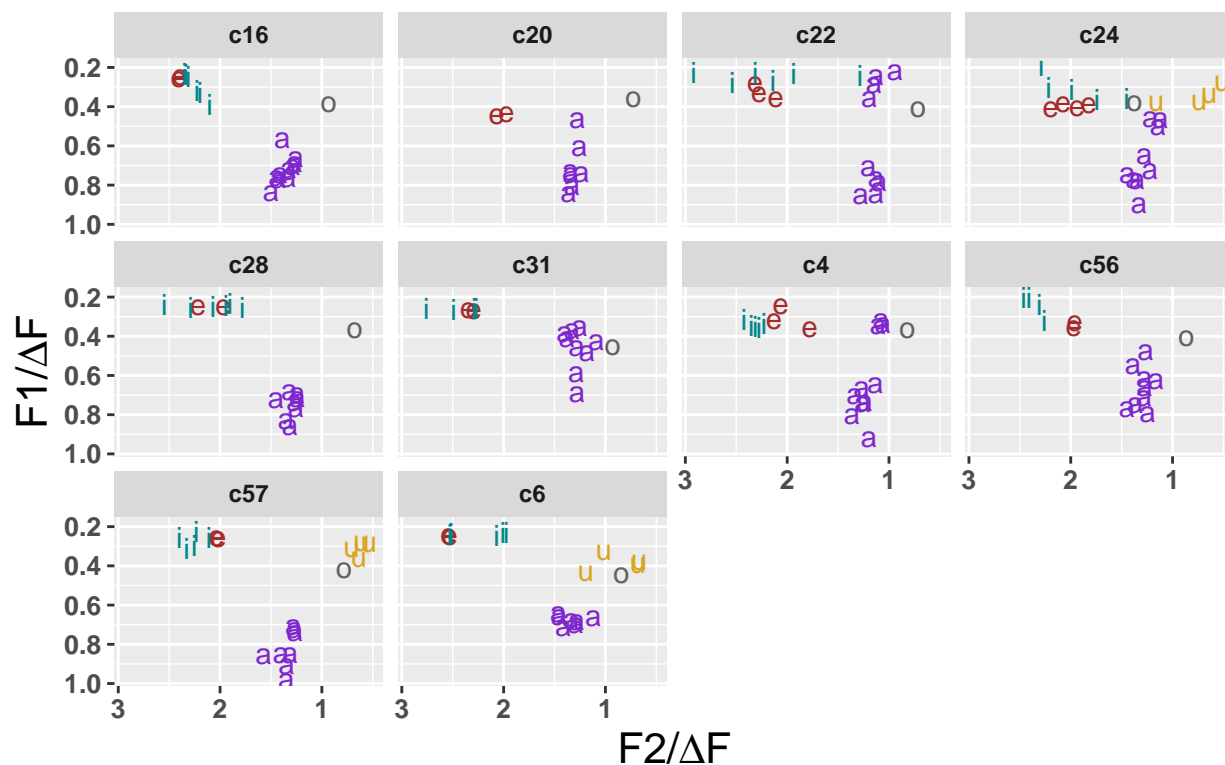
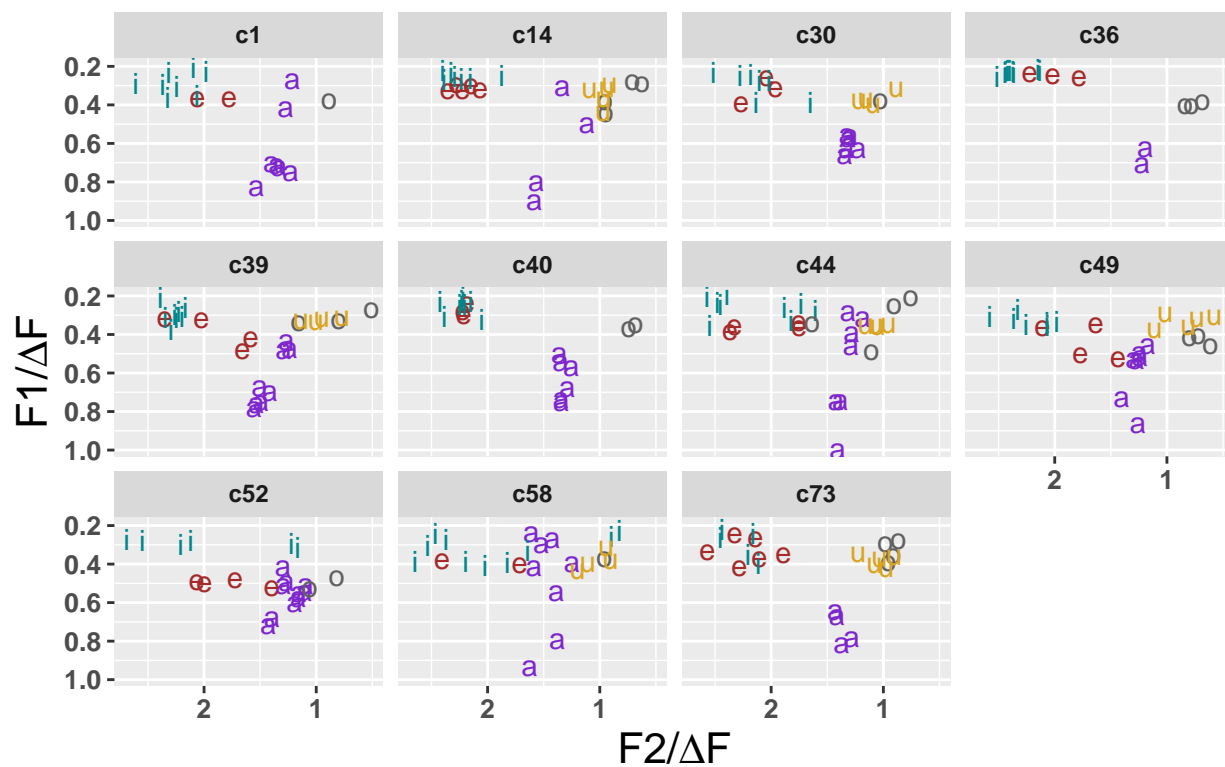


Figure 11: (`#fig:visualize_med_lobanov_vowels`)Lobanov-normalized vowels by age: adults and children

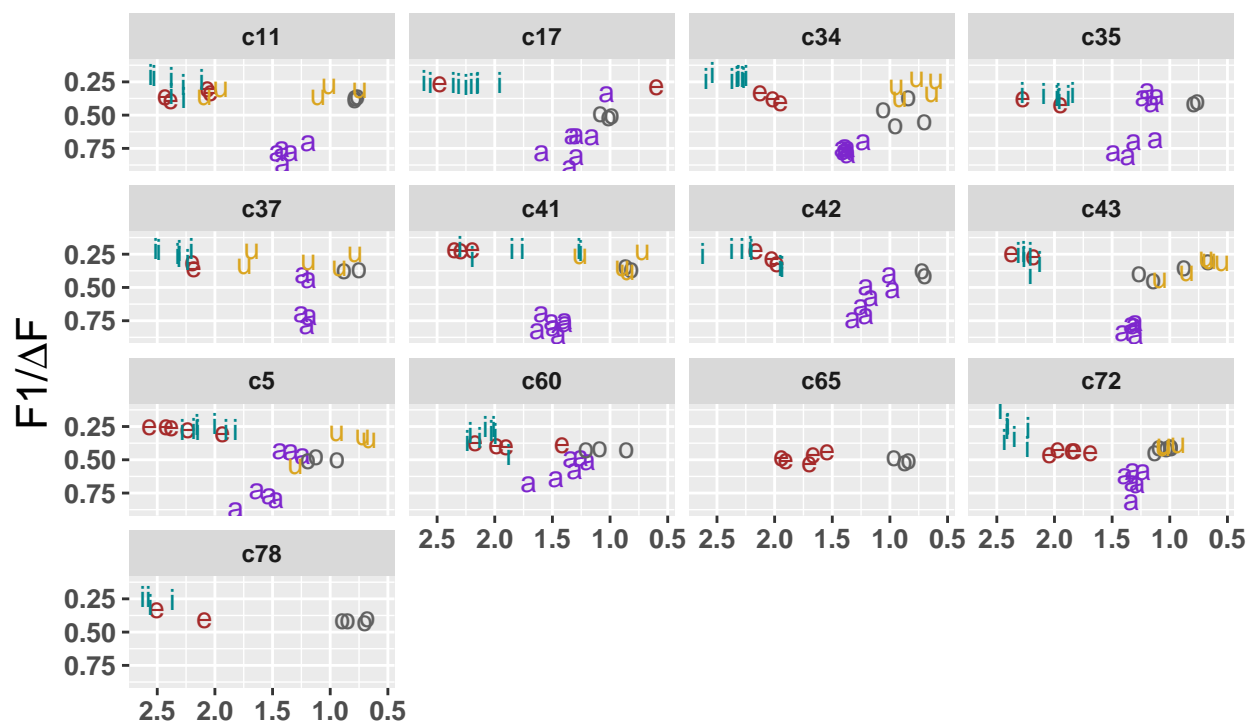
ΔF -normalized vowel space in four-year-olds



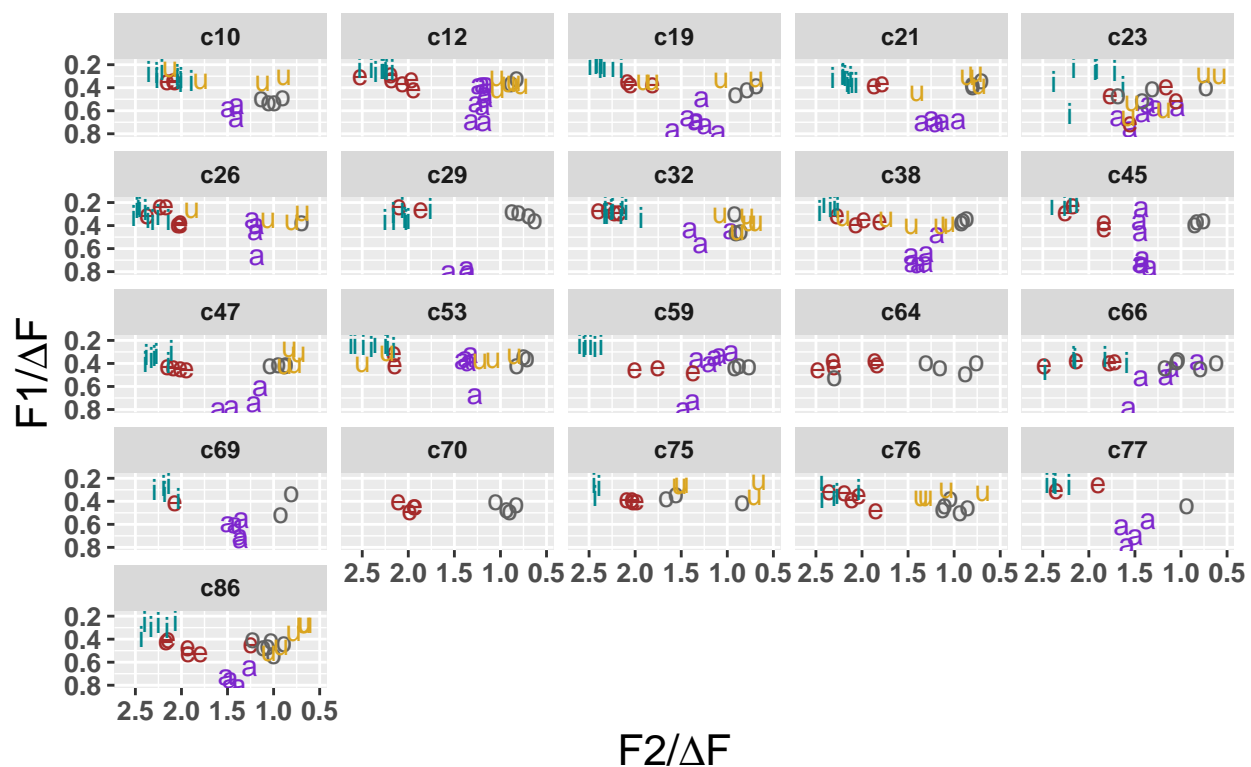
ΔF -normalized vowel space in five-year-olds



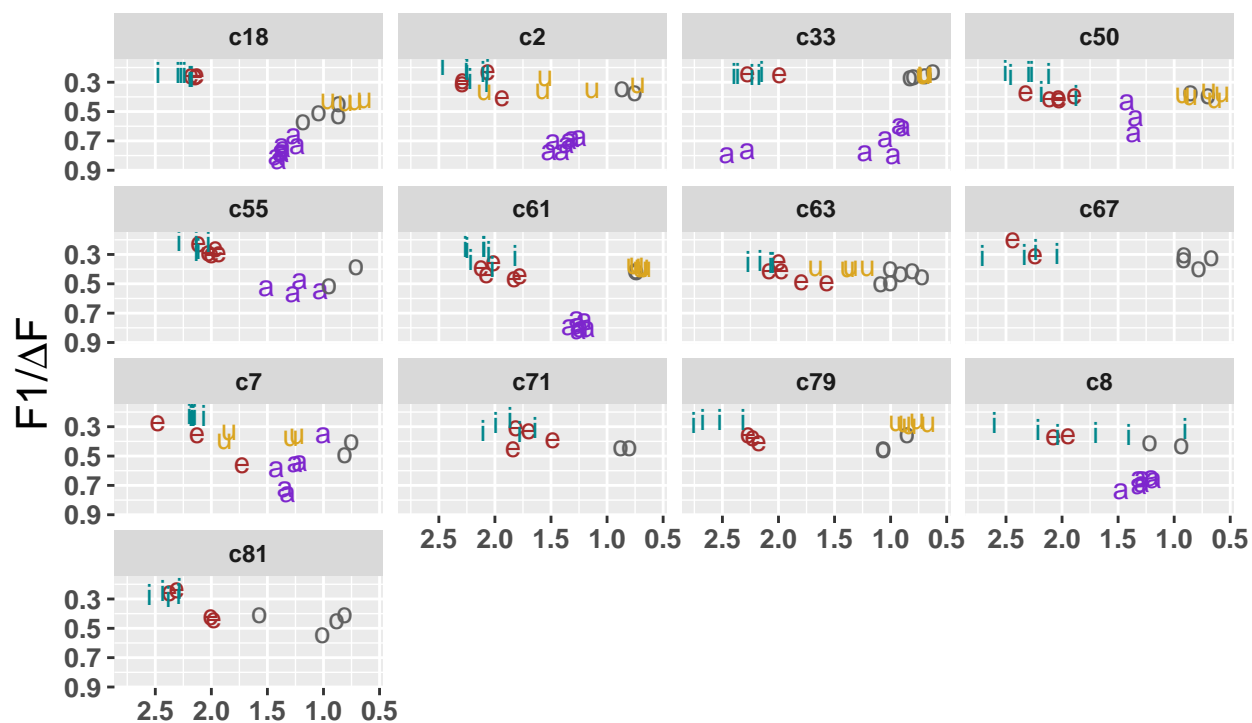
ΔF -normalized vowel space in six-year-olds



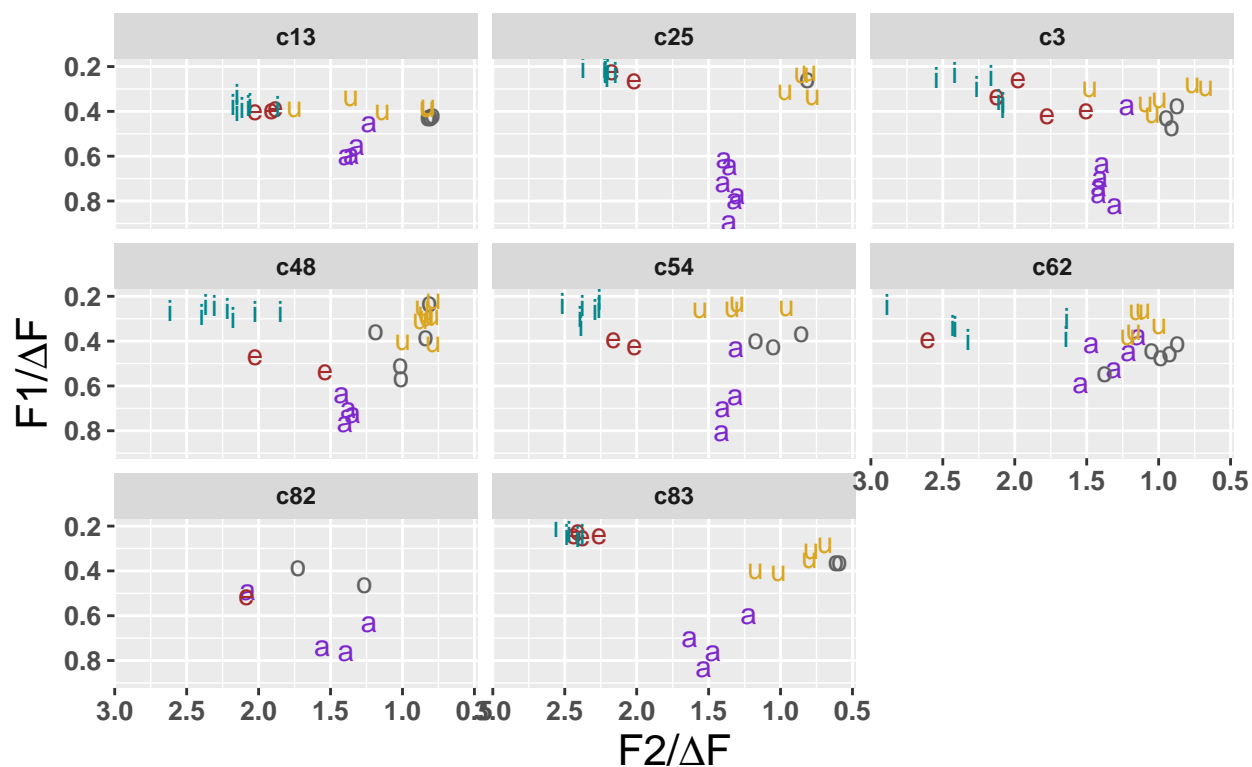
ΔF -normalized vowel space in seven-year-olds



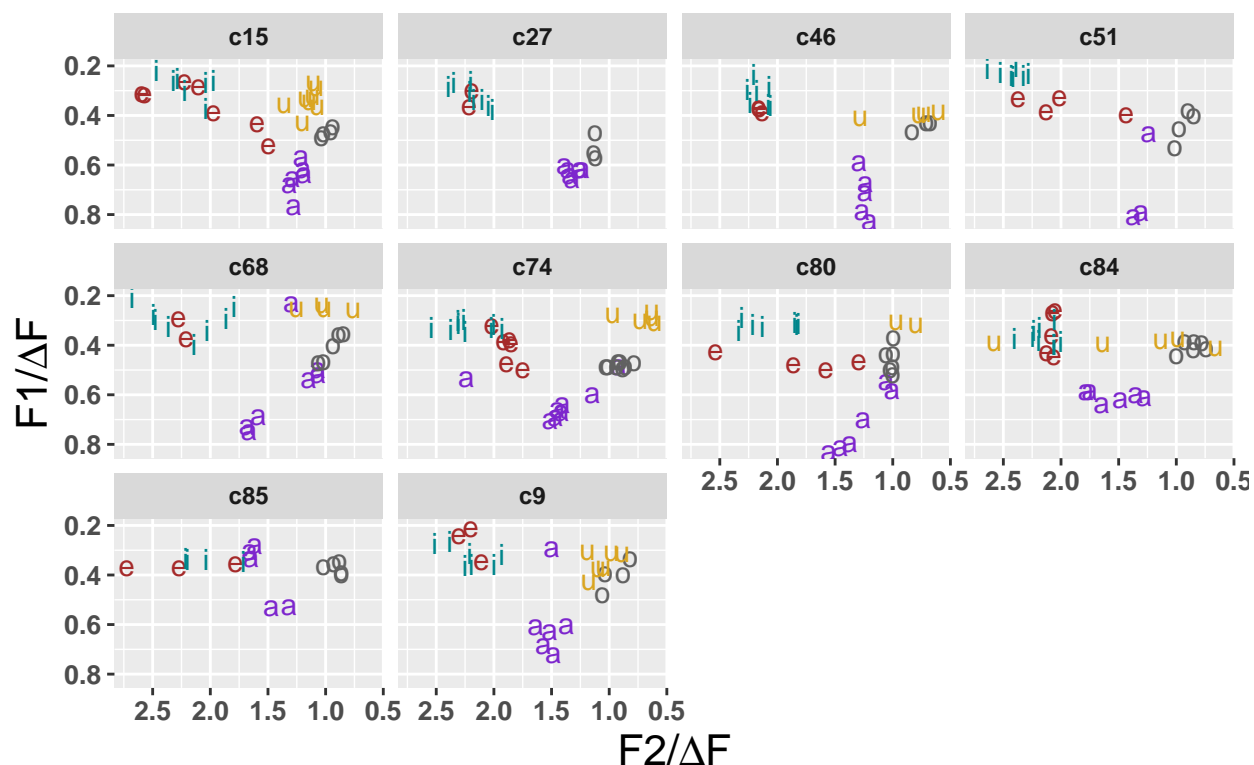
ΔF -normalized vowel space in eight-year-olds



ΔF -normalized vowel space in nine-year-olds



ΔF -normalized vowel space in ten-year-olds



ΔF -normalized vowel space in adults

