Corner vowels in males and females ages 4 to 20 years: Fundamental and F1–F4 formant frequencies

Houri K. Vorperian[a)] and Raymond D. Kent
Vocal Tract Development Laboratory, Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, Wisconsin 53705, USA

Yen Lee and Daniel M. Bolt
Department of Educational Psychology, University of Wisconsin-Madison, 1086 Educational, Sciences Building, 1025 West Johnson Street, Madison, Wisconsin 53706, USA

(Received 21 June 2019; revised 6 September 2019; accepted 8 October 2019; published online 12 November 2019)

The purpose of this study was to determine the developmental trajectory of the four corner vowels' fundamental frequency (f_o) and the first four formant frequencies (F1–F4), and to assess when speaker-sex differences emerge. Five words per vowel, two of which were produced twice, were analyzed for f_o and estimates of the first four formants frequencies from 190 (97 female, 93 male) typically developing speakers ages 4–20 years old. Findings revealed developmental trajectories with decreasing values of f_o and formant frequencies. Sex differences in f_o emerged at age 7. The decrease of f_o was larger in males than females with a marked drop during puberty. Sex differences in formant frequencies appeared at the earliest age under study and varied with vowel and formant. Generally, the higher formants (F3-F4) were sensitive to sex differences. Inter- and intra-speaker variability declined with age but had somewhat different patterns, likely reflective of maturing motor control that interacts with the changing anatomy. This study reports a source of developmental normative data on f_o and the first four formants in both sexes. The different developmental patterns in the first four formants and vowel-formant interactions in sex differences likely point to anatomic factors, although speech-learning phenomena cannot be discounted.
 https://doi.org/10.1121/1.5131271

[EJ] Pages: 3255–3274

I. INTRODUCTION

The acoustic properties of speech vary with the age and sex of speakers, and these age-sex differences need to be considered for purposes such as determining anatomic-articulatory-acoustic relationships through the lifespan, designing automatic speech recognition for diverse speaker groups, setting parameter values for synthesized speech of children and adults, and interpreting clinical data from individuals with speech disorders such as cerebral palsy, Down syndrome, and hearing impairment. In their classic study on vowels, Peterson and Barney (1952) convincingly showed the considerable dispersion in the formant frequencies of men, women, and children, presumably reflecting differences in vocal tract anatomy related to sex and age. Developmental trajectories were more clearly defined in subsequent studies that reported data on fundamental frequency (f_o) and vowel formant frequencies in children of both sexes and of various ages (see reviews by Vorperian and Kent, 2007, and Kent and Vorperian, 2018). The developmental data that are most extensive in covering the childhood years are those of Eguchi and Hirsh (1969), Perry *et al*. (2001), Assmann *et al*. (2008), and Lee *et al.* (1999), with data being most abundant for the first two formants, F1 and F2, less so for F3 (third formant), and least of all for F4 (fourth formant). Although F1 and F2 often suffice to establish the phonetic identity of vowels, the higher formants F3 and F4 enrich the speech production acoustic signal and have been correlated to important features of the vocal tract anatomy. F3 is associated with the most anterior region of the front cavity (Fant and Pauli, 1974), and F4 is associated with laryngeal descent/elevation (Sundberg and Nordström, 1976), as well as the pharyngeal and hypopharyngeal cavities (Lin *et al*., 1989; Takemoto *et al*., 2006). Data on the first four formants may help to determine anatomic-acoustic relationships for typical and atypical vowel development in both sexes. In addition, F3 and F4 have been shown to be important for specifying the acoustics of liquid sounds, both rhotics (Hagiwara, 1995) and laterals (Ladefoged and Maddieson, 1996); normalizing both rhotic and non-rhotic vowels (Disner, 1980; Hillenbrand and Gayvert, 1993); explaining the speaker's formant (a closeness of F3 and F4; Bele, 2006; Leino *et al*., 2011) and the singer's formant (a clustering of F3– F5; Sundberg, 1974); determining the acoustic correlates of differences in maxillary arch dimensions (Hamdan *et al*., 2018); and studying the consequences of clinical procedures such as tonsillectomy (Švancara *et al*., 2006), orthodontic treatment (Kulak Kayikci *et al*., 2012), and supracricoid laryngectomy (Buzaneli *et al*., 2018).

Comparing the formant data from published studies is complicated by differences in methodology, especially differences in speech samples, speaker dialect, and formant

[a)]Electronic mail: vorperian@waisman.wisc.edu

estimation procedures, and these factors may explain differences in results for the characteristics of the vowel quadrilateral (Kent and Vorperian, 2018). A common limitation of the developmental studies published to date is that they typically report data for only one word per vowel (e.g., /u/ in the word *boot*) and often do not control for influences such as children's familiarity with the test words, phonological neighborhood density of a given word, and coarticulatory effects. Consideration of all these factors in continuing research may lead to more valid data comparisons and more reliable estimates of derived indices, such as vowel space area (VSA) or other measures of vowel dispersion (Kent and Vorperian, 2018).

The primary goal of the present study is to report formant frequency data for the corner vowels of the classic quadrilateral that can be used to construct developmental trajectories for the first four formants in male and female speakers over the age range of 4–20 years old. This goal is based on the need to establish the maximal acoustic and articulatory working space of vowel production in typically developing individuals, as defined by the corner vowels of the vowel quadrilateral, and to use such normative data to better address the understanding and study of anatomic-acoustic relations. Classic research indicates that vowels are mastered by the age of 4 (Donegan, 2013; however, Yang and Fox, 2013, present evidence of continuing maturation until at least 8 years of age) and anatomic maturation is generally assumed by the age of 20. Data obtained with the same methodology used in the present research were previously reported for adults aged 20–92 (Eichhorn *et al.*, 2018), so that the current work and its predecessor constitute a lifespan investigation of corner vowel acoustics in American English. In the procedure used here, each vowel is represented by five different monosyllabic words selected according to criteria suited to developmental research, including word familiarity and phonological neighborhood density (Munson and Solomon, 2004). Another goal of the present study is to assess the variability in formant frequencies as a function of speaker age. Studies have shown that variability in acoustic and physiologic measures of speech production declines with maturation (Kent, 1976; Lee *et al.*, 1999; Smith and Goffman, 1998; Walsh and Smith, 2002), but it is not clear if intra- and inter-speaker variabilities decrease monotonically or have a more complicated developmental pattern related to periods of accelerated growth of the vocal tract with re-adaptation of speech motor-control to the changing anatomy (Vorperian, 2000).

The following three research questions motivated this research:

(1) What is the trajectory of developmental change for f_o and all four formants across the corner vowels? Based on previous studies, we expected to observe a general decline of f_o with age and a conspicuous fall in male speakers at the onset of puberty. For the formants, we expected to observe a progressive decrease in the formant frequencies of all vowels, but also expected that developmental effects across different individuals would not be uniform across age, vowel-type, formants, or speaker-sex. Decreasing formant frequencies with age within each sex are presumed to reflect growth of the vocal tract, but we hypothesize that the detailed pattern of age-related changes varies with vowel and formant because of nonuniform growth in different regions of the vocal tract.

(2) At what age will speaker-sex differences be evident in f_o and formant frequencies? Based on previous studies, we expected that differences in f_o would emerge at around 12 years of age, but that speaker-sex differences in formant frequencies would be evident by about 4 years of age, the youngest age under study, and these differences would accelerate at the age of puberty.

(3) What is the pattern of inter- (between) and intra- (within) speaker variability across development for the corner vowels in f_o and formant frequencies? Earlier studies generally show reduced variability with age, presumably reflecting maturation of speech motor control. We expected greater inter-speaker and intra-speaker variability in the age periods of 4–6 years and during puberty, periods that are influenced by factors such as speech motor learning and rapid anatomic changes.

II. METHODS

A. Participants, acoustic stimuli, and data collection

Speech recordings were made from 190 (97 female and 93 male) typically developing participants ages 4–20 years old. Participants were judged to have the regional dialect that is representative of the general geographic region from which they were recruited. The age of 4 was the youngest age recruited for this study because the larger research protocol of which this acoustic study was a component required participation in tasks that are not easily performed by children younger than 4. The speech stimuli, also used in Wild *et al.* (2018) and Eichhorn *et al.* (2018), consisted of the following five different monosyllabic American English words for each of the four corner vowels: /i/ (*bead*$_2$, *bee*, *eat*$_2$, *sheep*, and *feet*), /u/ (*boo*, *boot*$_2$, *zoo*, *hoot*$_2$, and *shoe*), /æ/ (*bath*, *bat*$_2$, *cat*, *hat*$_2$, and *sad*), and /ɑ/ (*dot*, *hop*, *pot*$_2$, *top*, and *hot*$_2$). Two of the five words, marked with subscript "2," were presented twice to assess intra-speaker variability. Since the stimuli were selected with the intent of studying speech production in both typically developing and atypically developing children (e.g., children with Down syndrome, as in the study by Wild *et al.*, 2018), the words were produced in isolation (i.e., a carrier phrase was not used to limit demands on the production task), and the selection of the words was based on the following factors: (1) Words should be familiar to younger participants and have high phonological neighborhood density, which reportedly maximizes F1-F2 vowel space (Munson and Solomon, 2004). Frequency of occurrence of words also can affect vowel production (Munson and Solomon, 2004), but this feature was not controlled given the difficulty of finding words that meet multiple criteria. (2) Preference was given to words with bilabial and alveolar consonants over words with sounds, such as velars, which can be difficult for children with motor speech disorders.

Recording was done in a quiet room using a Shure-SM48 microphone (Shure Inc., Niles, IL) mounted on a floor stand and adjusted to each participant's seated height at a 15 cm distance and 45 degree angle laterally from the mouth. The microphone was connected to a Marantz-PMD 660 digital audio recorder (Marantz Professional in Music Brands, Inc., Cumberland, RI) that digitizes at a rate of 48 kHz with 16-bit resolution on a SanDisk Ultra II flashcard (SanDisk Western Digital Corporation, San Jose, CA). To optimize recording level, the Marantz recorder gain was adjusted to 6–12 dB below the maximum level. The stimuli were presented visually (picture and orthographic word) and aurally (recordings from an adult male—with a f_o of 110 Hz, from the Midwest, i.e., same regional dialect as where the participants were recruited from—were played through external speakers) using a laptop with the TOCS+ platform program (Hodge *et al.*, 2009) for randomization. Participants were instructed to repeat the speech stimuli (28 words total) at a normal loudness level, with 2 practice words at the beginning. This study used a combination of methods for stimulus presentation that were originally designed to increase the likelihood of participation by young children with potential limitations in attention span, as well as potential limitations in cognitive, sensory, and motor functions. These procedures were used successfully in a study of speech intelligibility in children and adults with Down syndrome (Wild *et al.*, 2018). Applying the same procedures with all participants, children and adults, permits the comparability of data across speakers with and without developmental delay or disorder.

B. Acoustic analysis and measurements

Procedures of acoustic analysis were based on results of previous studies that (1) evaluated the accuracy of vowel formant measurements in four acoustic analysis systems (Burris *et al.*, 2014), (2) determined the effect of analysis parameter manipulations on formant measurements in children and adults (Derdemezis *et al.*, 2015), and (3) reviewed methods and data sources for vowel formant frequencies across the lifespan (Kent and Vorperian, 2018). The acoustic analysis procedures used here were the same as those used by Eichhorn *et al.* (2018) in a study of vowels produced by adults of different ages and are as follows: Speech recordings were uploaded to a computer, and the waveforms of each word were segmented using Praat (version 5.1.31, Boersma and Weenink, 2010), and saved as a separate sound file. Next, the vowel portion of each word was analyzed using an upgraded version of TF32 (time-frequency analysis software for 32-bit Windows; Milenkovic, 2010) to measure the frequency for f_o and F1– F4 values. TF32 was chosen for analysis because it does not degrade the signal through downsampling, has a linear predictive coding (LPC) formant-track overlaid on a gray-scale spectrogram for visual inspection of formant patterns (along with a pitch track), and a time-slice spectrum linked to the spectrogram that displays fast Fourier transform (FFT) and LPC spectral slice information. In addition, TF32 allows the user to select a range of LPC coefficients and the optimal dynamic range.

The measurement objective was to determine the extreme formant frequencies in each vowel production to allow comparisons with the classic studies on vowel formants (e.g., Lee *et al.*, 1999; Peterson and Barney, 1952). These extreme values serve to define the acoustic boundaries of vowel production over the lifespan and help establish the formant-frequency extrema as used in various indices of VSA or vowel dispersion (Kent and Vorperian 2018). The measurements were not intended to address vowel inherent spectral changes (Morrison and Assmann, 2013) although such features are certainly of interest in fully characterizing vowel production. Analysis entailed selecting a vowel-specific measurement point/inflection point to estimate formant frequencies because such an approach is suitable for vowels such as /u/ and /æ/ that are often produced with substantial formant shifts (Kent and Vorperian, 2018). Therefore, we first displayed the spectrogram and waveform of the segmented word and used the following criteria for selecting the vowel-specific temporal measurement point: vowel /i/, point of highest frequency of F2; vowel /u/, point of lowest frequency of F2; vowel /ɑ/, point of least separation between F1 and F2 frequencies; and vowel /æ/, point of most evenly spaced formants, while avoiding measurement at a point of decreasing F2-F1 difference (which reflects backing of the vowel). Next, all four formant frequencies F1–F4 measurements were estimated by inspecting (a) the spectrogram (with overlaid LPC formant tracks) and spectral slice (with zoom-in function for greater measurements accuracy, and cepstrum), and using (b) combined displays of the FFT spectrum, LPC spectrum, and cepstrum. Parameter manipulations to optimize the spectrogram for acoustic analysis included the following: (1) The analysis bandwidth of FFT spectrograms was adjusted for each speaker group. The bandwidth for adult male speakers was 300 Hz, and 350–500 Hz for women and children (i.e., speakers with a high f_o). In addition, a narrow band spectrogram with an analysis bandwidth of 50 Hz was used as needed to view the harmonic structure as an additional form of analysis to determine the formant pattern and guard against the possibility of a strong harmonic dominating the LPC analysis. (2) The dynamic range was adjusted to provide the preferred view of the formant pattern by increasing or decreasing the amount of energy present on the spectrogram. (3) The number of coefficients on the time-slice LPC spectrum was adjusted as needed (e.g., increased to differentiate merging formants or identify formants with low energy or decreased to avoid confusing strong harmonics or inter-formant energy with formants). Formant measurements that could not be reliably estimated or appeared to have extreme values (outliers), especially for children and/or higher formants, were scrutinized and re-measured using a consensus analysis approach where two or three examiners assessed the spectrogram and time-slice spectrum displays and used knowledge of acoustics to decide upon analysis values. If the examiners could not come to an agreement, no measurement was recorded for one or more formants (i.e., uncertain measurements were treated as missing data).

The measurements of f_o were made to obtain baseline data to inform formant-frequency estimation (because the accuracy of formant measurement depends, in part, on f_o values), and portray general developmental patterns. The f_o measurements for each vowel production were recorded at the same temporal point as that used for the formants using TF32's pitch determination algorithm. When a value of f_o was questionable (e.g., affected by irregular phonation, such as vocal fry, or when the pitch tracker failed), a narrowband spectrogram with an analysis bandwidth of 50 Hz and the time-slice spectrum FFT display (with 40 ms duration) were compared, and the first harmonic was recorded for the f_o measurement. The value of the first harmonic was interpolated from higher harmonics as appropriate. For example, if the tenth harmonic was of suitably high amplitude, the frequency of this harmonic was divided by ten to obtain the value of the first harmonic. To resolve f_o discrepancies, the f_o measurement was made at a different location than the F1–F4 temporal point. However, when the f_o measurement could not be resolved using any of these methods, no measurement was made.

To assess reliability of acoustic measurements for ages 4–20 years old, a random subset of recordings from eight typically developing (TD) speakers was measured by three raters and intra-class correlation (ICC) calculated for each vowel f_o, and F1–F4 using analysis of variance (ANOVA) variance components estimation in the statistical package SPSS version 25 (SPSS Inc., Chicago, IL). Findings revealed reliability to be excellent for all measurements with ICC estimates >0.927 with the 95% confidence interval lower boundaries >0.884, except for the f_o of vowel /u/ and the F4 of vowels /æ/ and /u/, where reliability was good with ICC estimates >0.877 with 95% confidence interval lower boundaries >0.806.

C. Statistical analysis

Prior to evaluating the research questions, the data were screened for outliers. For each speaker, the mean frequency measures (f_o, F1–F4) for a given vowel were evaluated against the distribution of measurements/observations across speakers of the same age and sex for the same vowel type and frequency measure. An outlier is defined as a measurement that is greater than two standard deviations from the mean frequency measure of the five words from the same vowel for each speaker. Because no outliers were detected, inter-speaker variability was based on mean frequency measures for all vowels from all speakers except for three who had missing F4 mean frequency measurements.[1] The frequency measures used to assess intra-speaker variability similarly included all speakers, except for cases with missing F3 and F4 measurements. Cases were missing if at least one of the measurements from the repetitions was missing. Table I lists the vowel- and frequency-specific numbers of female and male speakers in each analysis for each of the four age-cohorts (defined below). As seen in Table I, the inter-variability analysis had 1.58% of F4 measurements missing, and the intra-variability analysis had 1.44% and 4.21% of F3 and F4 measurements missing, respectively. Despite the number of missing cases for the higher formants F3 and F4, the remaining number of cases per age-cohort made the estimation of both inter- and intra-speaker variabilities possible.

As the individual forms of analysis conducted in this paper cut across the three research questions, we use a similar organization in the presentation of the statistical analysis and Sec. III (Results) that follow. The first set of analyses addressing research questions (1), (2), and the inter-variability portion of research question (3), used only one production for each of the repeated words from each speaker. These analyses focused on simultaneously assessing changes in both the inter-speaker means and inter-speaker variability of formant measurements in relation to speaker age and sex. In these analyses a total of 20 words were used for each speaker, where the mean of f_o and each formant (F1–F4) were computed across the five words of the same vowel type, and submitted for analysis.[2] To better understand inter-speaker variability, we considered a second set of analyses that examine inter-speaker variability in relation to pubertal-

TABLE I. Vowel-specific sample size of female (F) and male (M) speakers in pubertal-stage age-cohorts (years;months) as described in Sec. II (Methods) [pre-pubertal (4;0–7;11), peri-pubertal (8;0–10;2), pubertal (10;3–14;5), and post-pubertal (14;6-20;0)] used in variability analyses for inter-speaker variance and intra-speaker mean frequency difference (f_o, F1–F4).

Inter-variability	Frequency	Pre-pubertal (F/M)	Peri-pubertal (F/M)	Pubertal (F/M)	Post-pubertal (F/M)
/i/ /u/ /æ/ /ɑ/	f_o	22/21	12/13	30/29	33/30
/i/ /u/ /æ/ /ɑ/	F1-F2-F3	22/21	12/13	30/29	33/30
/i/ /u/ /ɑ/	F4	22/21	12/13	30/29	33/30
/ae/	F4	21/21	12/13	30/29	32/29
Intra-variability	Frequency	Pre-pubertal (F/M)	Peri-pubertal (F/M)	pubertal (F/M)	Post-pubertal (F/M)
/i/ /u/ /ɑ/ /æ/	f_o	22/21	12/13	30/29	33/30
/i/ /u/ /ɑ/ /æ/	F1-F2	22/21	12/13	30/29	33/30
/i/	F3	21/20	12/13	30/29	33/30
	F4	20/18	11/13	30/29	33/30
/u/	F3	20/16	12/13	30/29	33/30
	F4	20/16	11/13	30/29	32/30
/æ/	F3	21/21	12/13	30/29	33/30
	F4	19/17	12/13	29/29	31/29
/ɑ/	F3	22/20	12/13	30/29	33/30
	F4	21/20	11/13	29/28	32/30

stage age-cohorts corresponding to the age intervals described in Fitch and Giedd (1999). The distribution of the 190 participants/speakers is summarized in Table I with the sample size of the speaker- and sex-specific measurements in each of the following four age-cohorts: pre-pubertal [4;0–7;11 (years;months)], peri-pubertal (8;0–10;2), pubertal (10;3–14;5), and post-pubertal (14;6–20;0). Finally, a third set of analyses addressing the third research question on intra-speaker variability used only the two repeated words per vowel to measure production consistency within a speaker and its changes across the four age-cohorts. Like the prior set of analyses, these considered differences in the intra-speaker variability seen across the four age-cohorts.

The first set of analyses, noted above, examined developmental changes in the sex-specific frequency (f_o, F1–F4) means and inter-speaker variability for each of the four vowel types. This analysis entailed performing a separate analysis for each frequency measurement and vowel type using variance function regression (VFR; Western and Bloome, 2009). As demonstrated below, VFR enables the simultaneous analysis of mean and variance in relation to studied predictors. Because development underlying the frequency measurements will occur not only at varying rates but also at varying ages across speakers, we have every expectation that the measures will not only show overall mean change, but also changes in inter-speaker variance in relation to age. Such effects also contribute to heteroscedasticity of residuals, a feature that violates the homoscedasticity assumptions of traditional regression models. Importantly, VFR provides a way of not only accounting for heteroscedasticity but simultaneously studying both the mean and variance of development in relation to age. Both aspects of development were considered relevant in understanding change in frequency measurements. VFR has found prior use in a variety of applications, including age-related changes in self-reported health (Zheng *et al.*, 2011), income inequality (Cheng, 2014), cross-national differences in educational achievement (Montt, 2011), among others. In the current analyses, the variables age and sex were studied as predictors of both mean and variance. For each frequency measurement, the combined VFR model can be written as the

$$\text{mean structure}: \quad y_i = \beta_0 + \beta_{\text{male}}\text{male} + \sum_{l=1}^{k_m} \beta_l \text{age}^l + \sum_{j=1}^{p_m} \beta_j \text{male} * \text{age}^j + \epsilon_i = \boldsymbol{x}'_i \boldsymbol{\beta} + \epsilon_i, \tag{1}$$

$$\text{variance structure}: \quad \log(\sigma_i^2) = \lambda_0 + \lambda_{\text{male}}\text{male} + \sum_{l=1}^{k_v} \lambda_l \text{age}^l + \sum_{j=1}^{p_v} \lambda_j \text{male} * \text{age}^j + e_i = \boldsymbol{z}'_i \boldsymbol{\lambda} + e_i, \tag{2}$$

where y_i is the mean frequency measurement across words of the same vowel type for speaker i, $\log(\sigma_i^2)$ is the natural log of the residual (between-participant) variance of y_i, and age^l and $\text{male} * \text{age}^j$ represent the polynomial terms of chronological age and its interaction with speaker-sex, respectively. As described below, the highest order of such terms to be included is determined empirically.

The estimates of model parameters were obtained using the traditionally applied iterative procedure in which the unknown parameters of each equation were updated conditionally upon provisional parameters of the other equation.[3] Initially, a fifth-degree polynomial regression with speaker-sex and age interaction was used to model both mean (m) and variance (v), followed by likelihood ratio (LR) tests to determine the highest order terms (k = age, and p= age $\times$ male), i.e., k_m, p_m, k_v, and p_v, for the best fitting model. The outcomes of the LR tests can, as a result, lead to different models for the mean and variance equations, as well as differences across the formant/vowel types under consideration. The same VFR procedure was applied for each frequency measurement, producing a total of 20 VFRs.

The VFR seeks to model the mean and variance trajectories in relation to age (as displayed in Figs. 1 and 2). Given an initial model that includes all polynomial terms up to the highest order (in this case five) an iterative process was followed that successively removed higher order terms found not to be statistically significant through application of a LR test. When the iterative process reached a point at which removing the remaining term of highest order was statistically significant, that term, along with all lower order terms (whether significant or not), were retained for the final model. The resulting coefficients for the best fitting models are illustrated in Tables II and III. In each table, the presence of estimates for some effects but not others makes apparent the best fitting model determined by VFR. For example, for the f_o frequency /ɑ/ vowel analysis, $k_m = 3$, $p_m = 3$ yielded the best fitting model for the mean structure, implying polynomial terms up to age^3 needed to be included to account for the effect of age, and male $\times$ age^3 to account for the interaction between sex and age (see Table II). For the same f_o and /ɑ/ vowel, the variance structure equation identified $k_v = 3$, $p_v = 1$, implying polynomial terms up to age^3 needed to be included to account for the effect of age, but terms only up to male $\times$ age^1 to account for the interaction between sex and age (see corresponding entry in Table III). While all presented coefficient estimates are relevant to describing the exact nature of the age and sex-by-age interaction effects, for our purposes we simply note that an age effect of some form exists if any term of age is retained in the best-fitted model. Similarly, a speaker-sex effect of some form exists if speaker-sex (male) or any term of speaker-sex (male) and age interaction is retained.

To better evaluate sex differences at specific ages based on the VFR, we also applied tests of the main effect of sex where the age variable was centered at each of the different age values.[4] These analyses are statistically equivalent (each implies the same mean and variance trajectories by age and sex) but provide an analytical mechanism by which to test sex and variance differences at each possible age level (as displayed in Figs. 3 and 4).

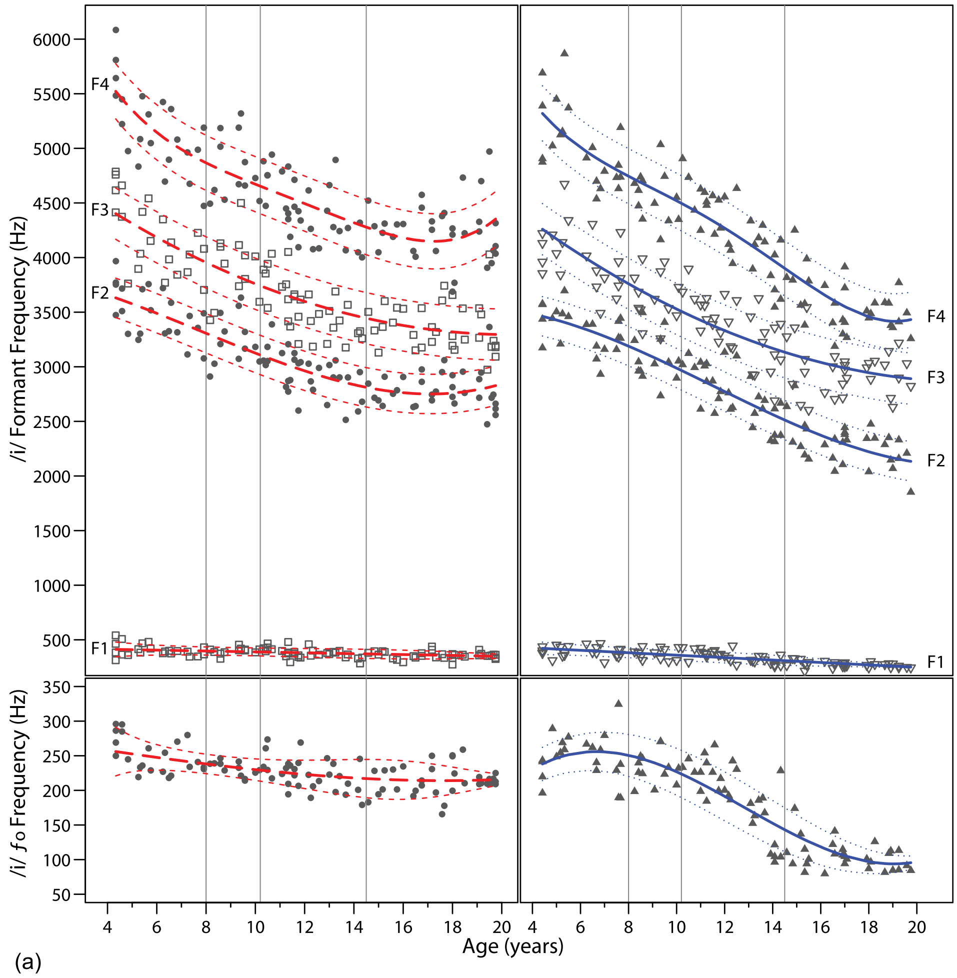
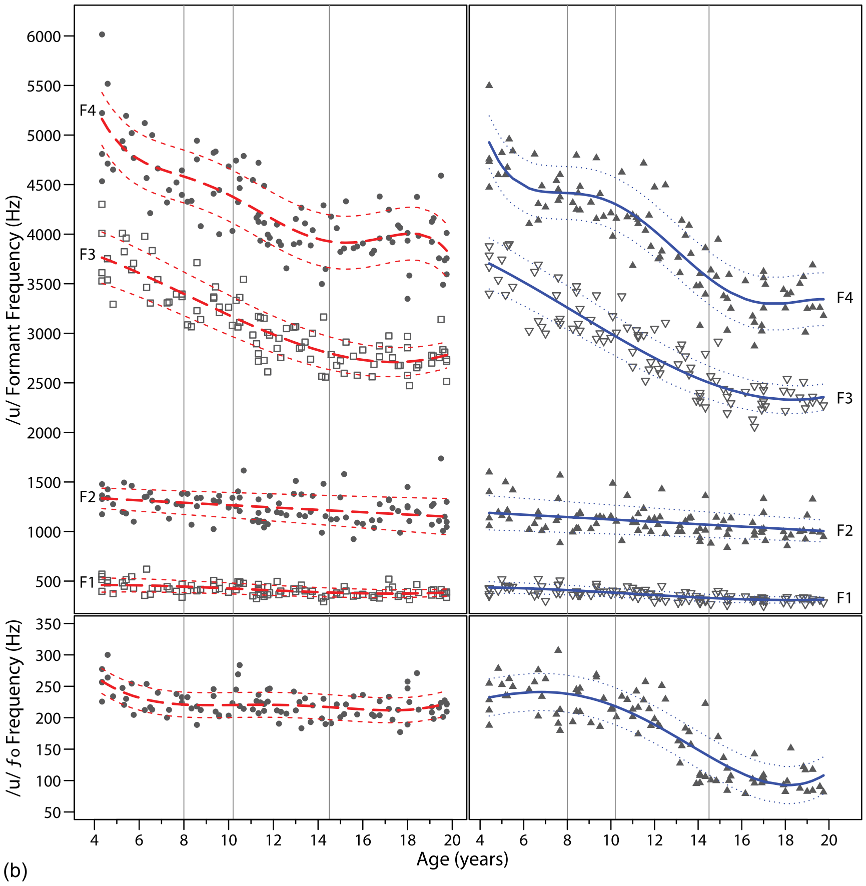


FIG. 1. (Color online) (a)–(d) Display of the vowel mean data of each speaker for f_o and F1–F4: /i/ (a), /u/ (b), /æ/ (c), and /ɑ/ (d). Female speakers are shown in the left panel, and male speakers are shown in the right panel. For each frequency, the variance function regression (VFR) with fifth degree polynomial fits is displayed using thick dashed and thick solid line lines for females and males, respectively, with ±1 standard deviation for each denoted by thin dashed and dotted lines for female and male speakers, respectively. Vertical lines reflect the four pubertal-stage cohorts (years;months) as described in Sec. II (Methods): pre-pubertal (4;0–7;11); peri-pubertal (8;0–10;2); pubertal (10;3–14;5); and post-pubertal (14;6–20;0).

As the statistically significant polynomial terms in many of the VFR analyses imply nonlinear and/or interaction effects, we also rely heavily on graphical inspection of the results based on the models applied when interpreting the findings (of the kind in Figs. 1 and 2). As the coefficients in each of the models work together in defining trajectories, such graphical inspection becomes a more meaningful way of understanding the combined influence of the statistically significant predictors, as opposed to trying to interpret each coefficient separately.

To further assess inter-speaker variability of the frequency measures (f_o, F1–F4) for each of the four vowel types, we followed up the VFR modeling of variance structure with a second set of analyses, applying F-tests to examine variance differences in relation to the different pubertal-stage age-cohorts that are anticipated to correspond to known developmental changes. Our ability to make statistical claims regarding inter-speaker variability differences across the age-cohorts depends on the number of speakers within the relevant cohorts. Thus, unlike the VFR analysis, the F-tests for variance differences are sensitive only to the measurements collected within the relevant intervals. For each frequency and vowel combination, comparisons of all pairs of the four age-cohorts were conducted for male and female speakers. See the supplementary material for the estimated inter-speaker variance (Hz^2) for each frequency type by speaker-sex by age-cohort (with the corresponding results in terms of statistical significance patterns displayed in Fig. 5).[5] The goal of these analyses was to determine whether age-related effects are present in the variance of each frequency (f_o, F1–F4) in males and females across development. Given the multiple comparisons, a Bonferroni correction was applied with an α-level of $0.008\bar{3}$ to control for the inflated type I error rate in assessing significant age-cohort comparisons.

A third set of analyses addressed the second part of research question (3) by using the mean absolute discrepancy values across the repeated words to study developmental change in intra-speaker variability across the four age-cohorts for the males and females. These analyses were again performed by vowel and frequency, implying each participant provided one data point (a mean of two absolute



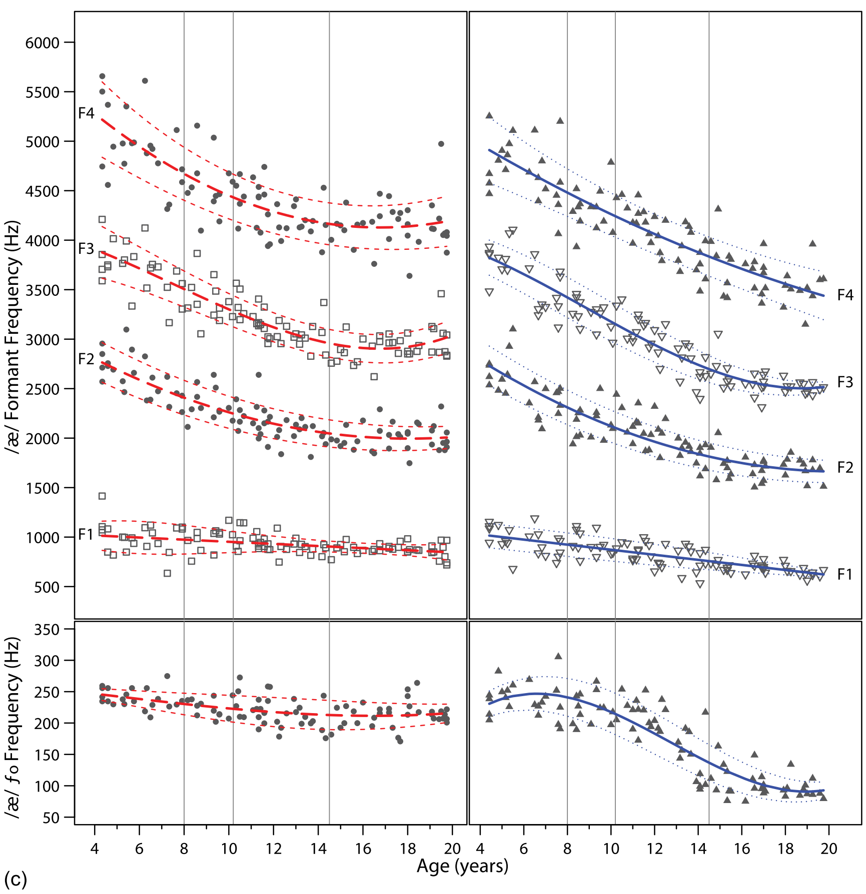


FIG. 1. (*Continued*).

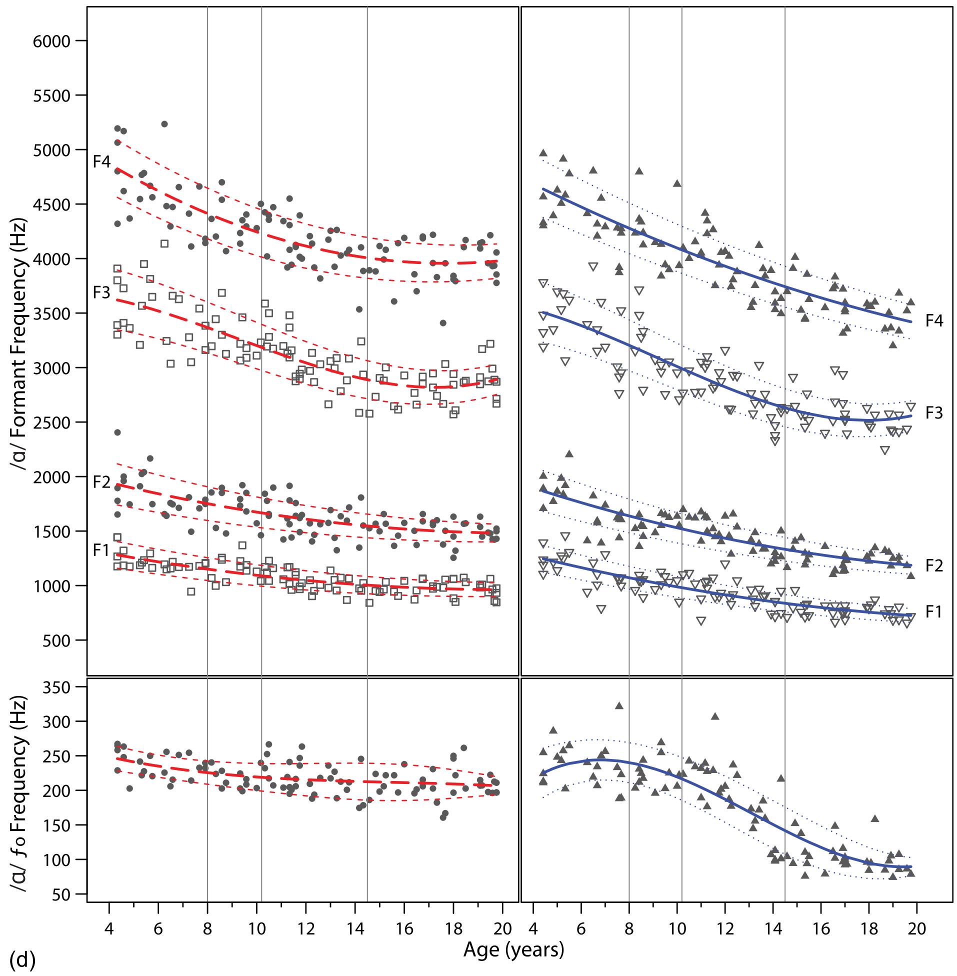


FIG. 1. (*Continued*).

discrepancy values—one for each repeated word) per analysis. Preliminary inspection of the data showed severely right-skewed distributions of the measurements; therefore, a nonparametric Wilcoxon test was applied. The Wilcoxon test compared differences across age-cohorts, as well as sex differences across age-cohorts, using the same Bonferroni correction as above. See the supplementary material for the intra-speaker median absolute discrepancy scores for each frequency type by speaker-sex by age-cohort (with corresponding results in terms of statistical significance patterns displayed in Fig. 6).[5]

III. RESULTS

Figures 1(a)–1(d) show the trajectories of the VFR fit for both mean structure and variance structure of f_o and F1–F4 for each vowel and each sex. In addition, the means for the female and male trajectories (with band of ± 1 standard deviation) are shown in the left panels of Figs. 2(a)–2(d), while the corresponding model-predicted log variances are in the right panels. These graphs give an overview of sex- and age-related changes in f_o and the four formants estimates for each corner vowel. The overall pattern suggests a systematic decrease in all frequencies (f_o, F1–F4) for all vowels, particularly in male speakers. The aberrant trends at the extreme ages, where minor increases or decreases in frequencies are noted (particularly after age 17), can be ignored as they reflect a boundary limitation of the polynomial fit that is typically due to the limited number of measurements at the extremes (De Boor, 1978). The details of the VFR results are shown in Tables II and III. The tables make apparent not only the presence of statistical significance in relation to age, sex, and age $\times$ sex, but also the complexity of the relationships due to the need for higher-order polynomial terms. Therefore, Figs. 1 and 2 help guide the interpretations of the findings. Figures 3 and 4 display the age-specific graphic and numeric values for the means and variances, respectively, using the age-centered VFR models. For inter-speaker variance, the overall pattern suggests a decrease in variance as age increases with the exception of f_o where variance increases until puberty and then decreases [Figs. 2(a)–2(d), right panel, and Fig. 4].

A. Developmental trajectories: Fundamental frequency

1. Fundamental Frequency—Means

Findings confirm the expected general decrease in f_o as age increases for all corner vowels, with the male f_o decreasing at a faster pace than the female f_o, and with sex

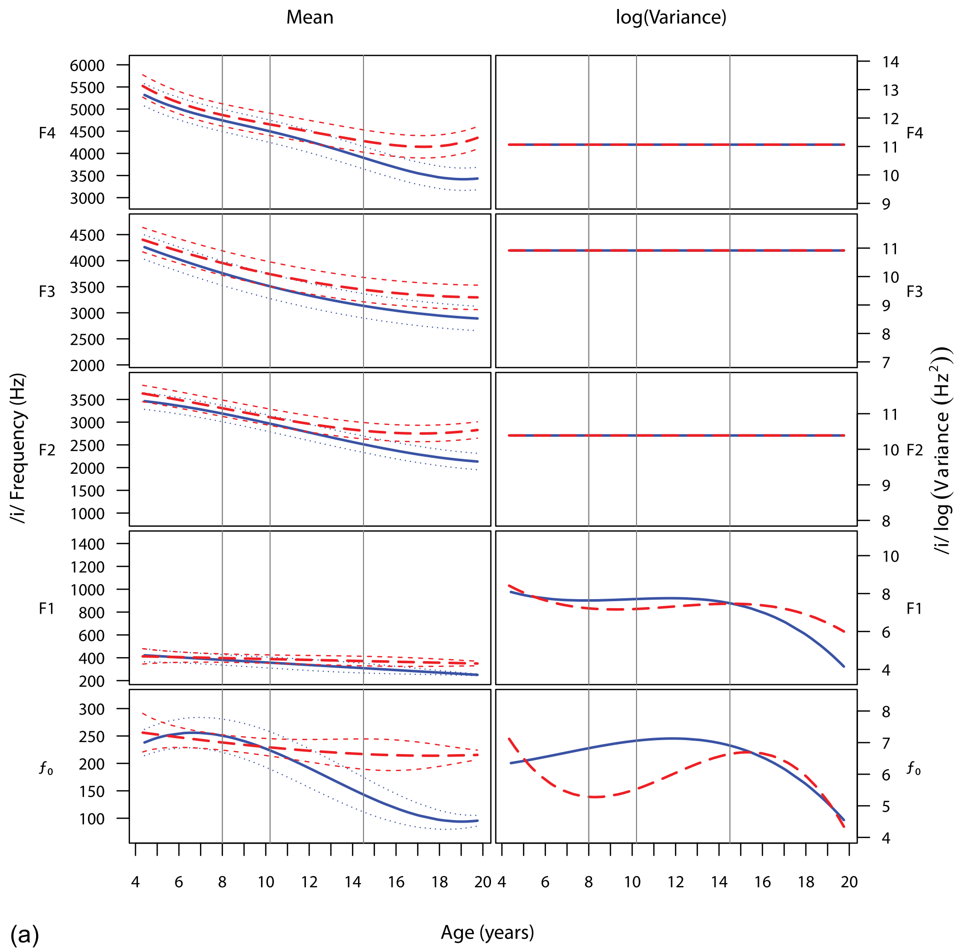


FIG. 2. (Color online) (a)–(d) Estimated VFR mean frequency and variance changes for vowels: /i/ (a), /u/ (b), /æ/ (c), and /ɑ/ (d). The left panel shows estimated mean (±1 standard deviation) for male (thick solid and thin dotted lines) and female (thick dashed line and thin dashed lines) speakers by age. The right panel shows the estimated VFR log residual variance for male and female speakers by age. Note that when there are no significant sex differences in variance, the model predicted variance plots are identical. See the caption for Fig. 1 for a description of the vertical lines.

differences increasing with age (Figs. 1–3). Significant sexual dimorphism first emerges at around age 7, where females have lower f_o. However, those differences decrease then re-emerge after the age of 10 with males having significantly lower f_o (see Fig. 3 with numeric values).

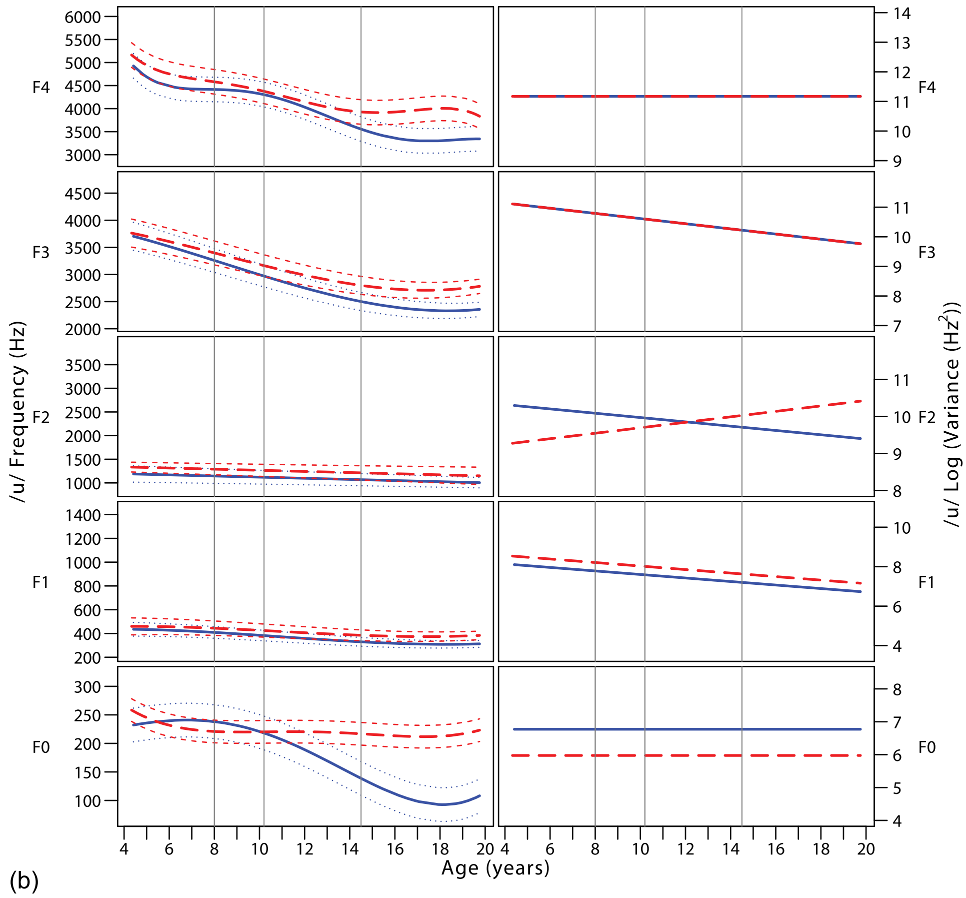
2. Fundamental frequency—Inter-speaker variability

The vowel /u/ stands out for its static f_o variance value in both males and females across development, with males having significantly greater inter-speaker variance than females throughout [Figs. 2(b) and 4]. However, the remaining vowels demonstrate a gradual increase in inter-speaker variance with vowel- and sex-specific peaks, typically during puberty, followed by a general trend of post-pubertal decrease in variance. Overall, male speakers had greater f_o inter-speaker variance than female speakers before age 14 except for /u/ where, as noted above, variance was significant at all ages examined [Figs. 2(a)–2(d), right panel, and Fig. 4]. Figure 5 displays the significant age-cohort comparisons with findings revealing a clear pattern of significant decrease in f_o inter-speaker variance from the pubertal to post-pubertal age-cohorts for all vowels in male speakers only.

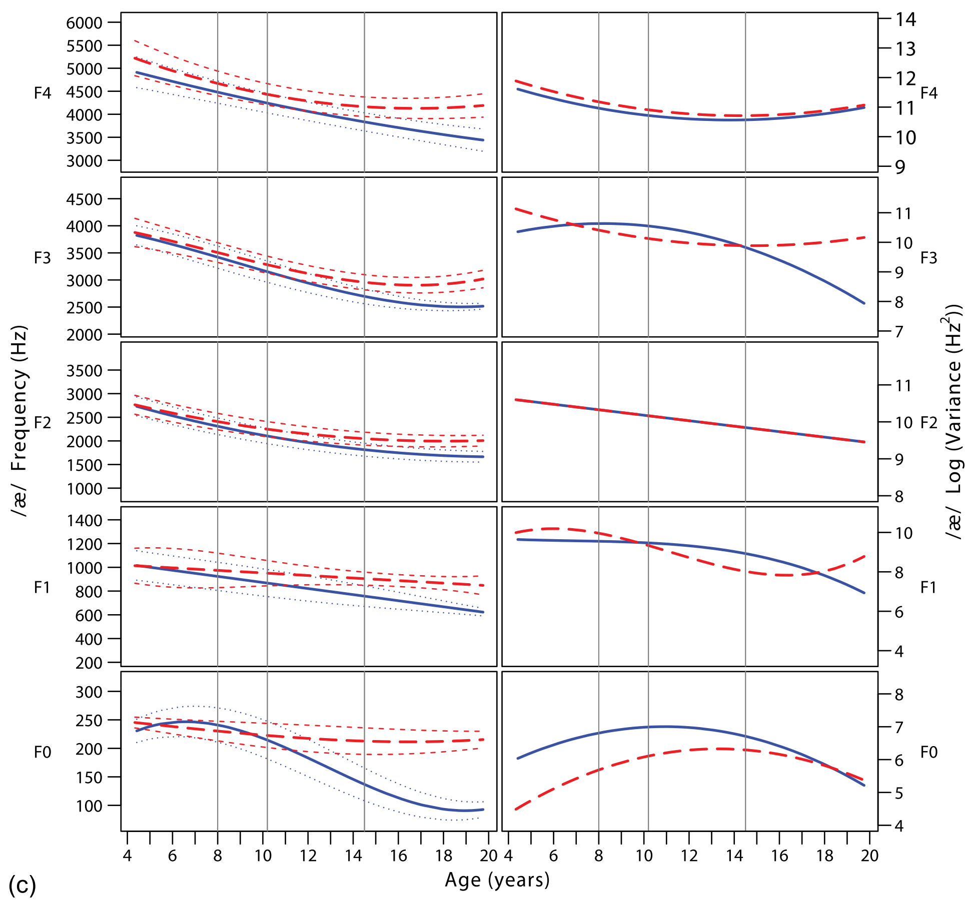
3. Fundamental frequency—Intra-speaker variability

Similar to inter-speaker f_o variability, intra-speaker f_o variability decreased as age increased. However, while inter-speaker f_o variance decreased significantly from pubertal to post-pubertal age-cohorts for all vowels in males, intra-speaker f_o differences decreased significantly from the pre-pubertal to post-pubertal age-cohorts for all vowels in males and the vowels /i/ and /u/ in females (see Fig. 5 versus Fig. 6).

As for sex differences, contrary to the above reported f_o inter-speaker variability where males generally have significantly greater variance than females prior to age 14 (during and before puberty, Fig. 4), significant sex differences in f_o intra-speaker variability were present primarily during post-puberty for all vowels, with males having smaller f_o differences/discrepancies than females (for /i/, $W = 694.50$, $p = 0.0061$, median difference: F = 8.5, M = 5.25; for /u/, $W = 703.50$, $p = 0.0042$, median difference: F = 10.50, M = 6.50; for /æ/, $W = 699.50$, $p = 0.005$, median difference:



(b)



(c)

FIG. 2. (*Continued*).

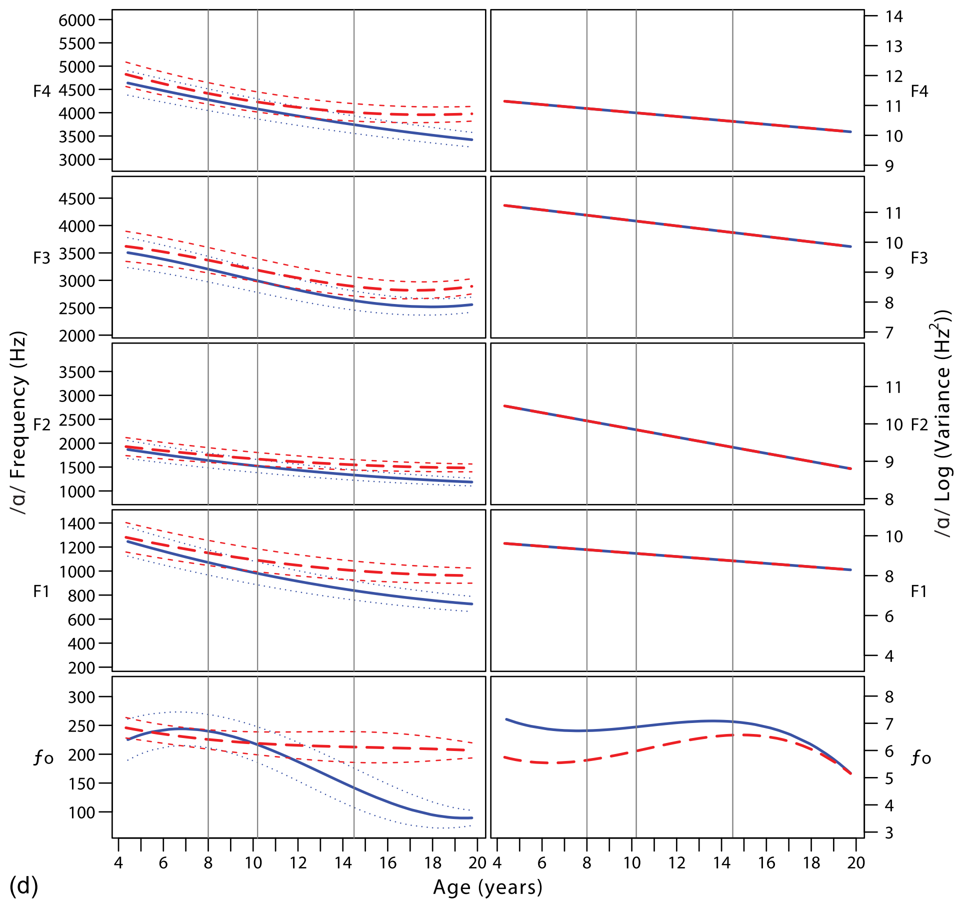


FIG. 2. (*Continued*).

F = 10.5, M = 6.5; for /ɑ/, $W = 707.00$, $p = 0.0036$, median difference: F = 10.5, M = 4.75).

B. Developmental trajectories: Formants

1. Formant frequencies—Means

There is a general trend for F1–F4 frequencies of all vowels to decrease with age particularly in males with an overall trend for males to have lower F1–F4 frequencies than females. The higher formants F3 and F4 have a greater developmental change in linear frequency values than the main vowel formants F1 and F2. In particular, F4 shows distinct mean frequency differences between male versus female speakers, beginning at age 4, as depicted in Figs. 2(a)–2(d). Noteworthy of mention is that despite the drastic developmental changes in the higher formants, such changes are not necessarily associated with greater values of formant scaling factors between children and adults. For example, the formant scaling factor (ratio between frequencies) for 4-year-old boys and male adults is 1.42 for the F1 frequency of vowel /i/ and 1.57 for the F4 frequency of the same vowel. In addition, males tend to have lower F1–F4 frequencies. The trend for male speakers to have a larger decrease in formant frequencies as age increases naturally results in increased sex differences with age. Findings reveal vowel- and formant-specific significant sex differences emerging at a young age, especially for F2 and F4 (e.g., age 4 in F2 for the vowels /i/ and /u/, and F4 for the vowel /æ/; see p-values displayed in Fig. 3). Interestingly, the vowel- and formant-specific sexual dimorphism appears to emerge at different ages. Sexual dimorphism in F1 emerges at age 6 for the vowels /u/ and /ɑ/, age 8 for /æ/, and age 9 for /i/. Sex differences in F2 are present at age 4 for the high vowels /i/ and /u/, but for the low vowel only emerge at age 6 for /ɑ/ and age 7 for /æ/. Similarly, sex differences for F3 emerge at age 5 for the vowel /i/, 6 for /ɑ/, 7 for /u/, and 8 for /æ/. As for F4, sex differences for F4 are present at age 4 for the low front vowel /æ/ and persist throughout development, but emerge at ages 6 and 7 for the back vowels /ɑ/ and /i/, respectively. As for the high back vowel /u/ sex differences emerge at age 5 but dissolve after age 8, and then re-emerge at age 13.

2. Formant frequencies—Inter-speaker variability

The model predicted variance of the formant frequencies is displayed in Figs. 2(a)–2(d) (right panel) and VFR-based age-centered variance in Fig. 4. With the exceptions of the static variance of F2–F4 for the vowel /i/, and F4 for the vowel /u/, as well as the substantial increase of F2 variance for the vowel /u/ among female speakers, the general trend was for variance to decrease with age in both female and male speakers. F-tests were carried out to assess the change in variance across the four age-cohorts for all vowels and formants, except when the model predicted variance remained stable in males and female speakers as a function

TABLE II. The estimated mean structure coefficients [Eq. (1)] of the best-fitting VFR models by frequency and vowel type. Asterisks denote significance levels: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. (Those terms determined by the VFR analysis not to be statistically beneficial are shown as "—" and can be interpreted as fixed 0's.) Note: The relationship between size of estimated coefficient and significance (p-value) is not direct across coefficients due to varying standard errors.

Frequency		/i/	/u/	/æ/	/ɑ/
f_o	Intercept	227.24***	222.08***	221.73***	218.98***
	Male	−46.01***	−46.31***	−47.60***	−43.73***
	Age	−171.44***	−105.29***	−131.70***	−136.36***
	Age × male	−619.42***	−630.01***	−633.02***	−605.63***
	Age^2	58.05*	76.93**	58.78**	50.08*
	Age^2 × male	−174.60***	−182.18***	−160.97***	−189.32***
	Age^3	6.46	−33.02	9.49	−18.39***
	Age^3 × male	174.66***	225.12***	165.17***	181.45***
	Age^4	—	53.80*	—	—
	Age^4 × male	—	—	—	—
F1	Intercept	380.87***	411.14***	1059.83***	1320.62***
	Male	−45.09***	−47.89***	68.52	27.17
	Age	−253.51***	−409.78***	−0.89***	−1.71***
	Age × male	−444.28***	−190.04*	−1.24***	−1.11***
	Age^2	—	69.39	—	338.08***
	Age^2 × male	—	—	—	—
	Age^3	—	85.99	—	—
	Age^3 × male	—	—	—	—
F2	Intercept	3049.07***	1241.47***	2816.34***	1988.38***
	Male	−275.96***	−144.97***	66.27	11.90
	Age	−3694.10***	−748.18***	−4.07***	−2.40***
	Age × male	−2163.87***	—	−1.72***	−1.30***
	Age^2	1213.42***	—	1098.61***	441.23***
	Age^2 × male	−1081.57**	—	—	—
	Age^3	379.53*	—	—	—
	Age^3 × male	—	—	—	—
F3	Intercept	3678.30***	3079.38***	3226.94***	3788.42***
	Male	−268.56***	−239.28***	−213.56***	−46.32
	Age	−4472.05***	−4524.51***	−3994.67***	−4.66***
	Age × male	−1116.79*	−1522.85***	−1859.56***	−1.21*
	Age^2	1152.61***	1294.85***	1487.94***	963.00***
	Age^2 × male	—	—	−434.35	—
	Age^3	—	469.80*	447.84**	465.61*
	Age^3 × male	—	—	—	—
F4	Intercept	4585.90***	4272.66***	4425.19***	4208.37***
	Male	−342.67***	−311.47***	−329.60***	−250.02***
	Age	−5021.77***	−4694.27***	−4060.39***	−3418.26***
	Age × male	−3021.72***	−2208.02***	−1833.86***	−1554.77***
	Age^2	1857.97***	1658.36***	1868.44***	1307.32***
	Age^2 × male	−1474.97**	−1385.87*	−1324.55*	−824.84
	Age^3	229.09	134.37	—	—
	Age^3 × male	—	766.22	—	—
	Age^4	544.95*	−86.61	—	—
	Age^4 × male	—	1141.45*	—	—
	Age^5	—	−744.96**	—	—

of age—see Fig. 4 (i.e., no F-test was carried for F2–F4 for vowel /i/, F4 for vowel /u/). As displayed in Fig. 5, findings revealed significant formant- and vowel-specific changes across age-cohorts, with a general trend of pre-pubertal to post-pubertal significant decrease in variance for both male and female speakers for F1–F3 but not F4. More significant decrease in inter-speaker variability occurred between the male age-cohorts than the female age-cohorts, particularly, for F1 (all vowels except /u/), and F1–F3 in the vowels /ɑ/ and /æ/, implying that male variance decreases more rapidly than female variance.

Sex differences in F2 inter-speaker variance was present only for the vowel /u/ where variability increased in females from pre- to post-puberty, with males having significantly greater variance than female speakers at ages 4 and 5 and significantly smaller variance at age 19 (Fig. 4). Similarly, sex differences in F3 and F4 variance were only present for the vowel /æ/ with F3 variance increasing during the

TABLE III. The estimated variance structure coefficients [Eq. (2)] of the best-fitting VFR models by frequency and vowel type. Asterisks denote significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (Those terms determined by the VFR analysis not to be statistically beneficial are shown as "—" and can be interpreted as fixed 0's.) Note: Female is the reference group. The combination of the parameters without interaction (without × Male) forms the trajectory of female. On the other hand, any model including an interaction term indicates that the male and female trajectories are statistically different according to the LR test. The relationship between size of estimated coefficient and significance (p-value) is not direct across coefficients due to varying standard errors.

Frequency		/i/	/u/	/æ/	/ɑ/
f_o	Intercept	5.94***	5.98***	5.82***	6.01***
	Male	0.60**	0.80***	1.99**	1.82**
	Age	−1.08	—	3.48	1.76
	Age × male	−4.05	—	−0.01*	−0.01
	Age^2	−1.21	—	−6.18***	−3.78*
	Age^2 × male	−6.38*	—	—	—
	Age^3	−8.59***	—	—	−3.58*
	Age^3 × male	6.68*	—	—	—
F1	Intercept	7.28***	7.84***	8.93***	9.99***
	Male	−0.10	−0.43*	0.02	—
	Age	−4.68*	−5.58***	−10.28***	−0.01***
	Age × male	−6.36*	—	0.61	—
	Age^2	−0.41	—	2.93*	—
	Age^2 × male	−6.83*	—	−7.57	—
	Age^3	−4.32**	—	4.56*	—
	Age^3 × male	—	—	−5.73	—
F2	Intercept	10.39***	9.85***	10.92***	10.95***
	Male	—	0.00	—	—
	Age	—	4.62*	−0.01**	−0.01***
	Age × male	—	−8.27*	—	—
	Age^2	—	—	—	—
	Age^2 × male	—	—	—	—
	Age^3	—	—	—	—
	Age^3 × male	—	—	—	—
F3	Intercept	10.92***	10.43***	10.20***	11.62***
	Male	—	—	−0.30	—
	Age	—	−5.46***	−3.88	−0.01***
	Age × male	—	—	−5.98	—
	Age^2	—	—	3.06	—
	Age^2 × male	—	—	−8.48**	—
	Age^3	—	—	—	—
	Age^3 × male	—	—	—	—
F4	Intercept	11.06***	11.17***	11.64***	11.43***
	Male	—	—	−0.31	—
	Age	—	—	0.00	−0.01**
	Age × male	—	—	0.00	—
	Age^2	—	—	3.15*	—
	Age^2 × male	—	—	—	—
	Age^3	—	—	—	—
	Age^3 × male	—	—	—	—

peri-pubertal ages in males, and males having significantly larger F3 variance than females after age 17; also, F4 variance was significantly smaller in males than female at all ages studied.

3. *Formant frequencies—Intra-speaker variability*

The overall trend for formant frequency (F1–F4) intra-speaker variability (Fig. 5) to decrease as age increases is similar to that of inter-speaker variability (Fig. 6) where both intra- and inter-speaker variability reach their respective smallest values during post-puberty. Comparison of Figs. 5 and 6 shows striking differences in F4, where there is a general absence of inter-speaker variability in F4 but not for intra-speaker variability.

IV. DISCUSSION

This section addresses the three research questions that this study examined. (1) What is the trajectory of developmental change for f_o and all four formants across the corner vowels? (2) At what age will speaker-sex differences be evident in f_o and formant frequencies? (3) What is the pattern of inter- (between) and intra- (within) speaker variability across development for the corner vowels in f_o and formant frequencies?

A. The trajectory of developmental change for f_o and all four formants

As stated in the Introduction, we expected a progressive decrease in f_o and the formant frequencies of all vowels but also some non-uniform changes across age, vowel-type, formants, and speaker-sex. The following discussion is keyed to Figs. 1(a)–1(d) and 3.

The trajectories for f_o exhibit the expected overall decrease with age, but with much greater effects for males than females, as discussed in detail in Sec. IV B on sex differences. The mean values of f_o appear to decrease for both sexes beginning at about 7 years of age, reaching adult values at about 14 for girls and 16 for boys. Studies of both gross and microscopic anatomy show that laryngeal development in children is a protracted process that extends to late adolescence. Features of gross anatomy have been revealed by cadaver dissections (Kahane, 1978; Litman *et al.*, 2003; Wysocki *et al.*, 2008) and imaging methods (Rogers *et al.*, 2014; Wani *et al.*, 2016). The general conclusions are that (a) the laryngeal structures grow in size throughout childhood but maintain their relative proportions, (b) the larynx descends in the neck (resulting in lengthening of the vocal tract) with a primary descent by about 4 years of age and secondary descent during adolescence particularly in males, and (c) the vocal folds lengthen continuously in both sexes but relatively more in boys. Microscopic and histological studies reveal that the lamina propria develops over an extended period, reaching adult-like characteristics at 12 (Boseley and Hartnick, 2006; Hartnick *et al.*, 2005) or even later (Ishii *et al.*, 2000; Sato, 2018). These macro- and micro-anatomic changes likely account for age-related decreases in mean f_o and variability of f_o.

For the most part, the formant frequency trajectories for all four formants follow a continuously decreasing pattern that is roughly monotonic for some formants and vowels (e.g., all formants of vowel /a/) but not monotonic for others (e.g., F4 of vowel /u/). In their review article, Vorperian and Kent (2007) identified evidence of abrupt changes in formant frequencies at certain ages, specifically, an overall jump in vowel acoustic space in adolescent boys and a limited jump in the low vowel region of

		f_o (Hz)				F1 (Hzz)				F2 (Hz)				F3 (Hz)				F4 (Hz)			
Vowel	Age	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value
/i/	4	258	230	27	0.0727	414	426	-12	0.324	3657	3483	174	0.0348	4448	4325	124	0.0819	5622	5427	195	0.0932
	5	253	246	6	0.5265	410	415	-5	0.645	3576	3428	148	0.0216	4312	4171	142	0.0286	5350	5189	160	0.0766
	6	248	255	-7	0.3127	406	404	2	0.8735	3490	3360	130	0.0101	4185	4025	160	0.0065	5145	5009	137	0.0537
	7	243	256	-13	0.0415	402	393	9	0.3709	3399	3279	120	0.0037	4066	3888	177	8.00E-04	4990	4866	124	0.032
	8	238	251	-12	0.0547	398	382	16	0.0765	3307	3189	118	0.0015	3955	3760	195	0	4867	4745	122	0.0179
	9	234	241	-7	0.2975	393	370	23	0.0049	3215	3091	124	7.00E-04	3853	3640	213	0	4764	4633	132	0.0097
	10	230	227	3	0.6166	389	359	30	1.00E-04	3125	2988	138	3.00E-04	3760	3529	231	0	4672	4520	152	0.0039
	11	226	210	16	0.0062	385	348	37	0	3041	2881	160	0	3675	3426	249	0	4583	4400	183	8.00E-04
	12	223	191	32	0	381	337	44	0	2963	2773	190	0	3598	3332	267	0	4494	4269	225	1.00E-04
	13	220	172	48	0	377	326	51	0	2894	2666	227	0	3530	3246	284	0	4405	4128	278	0
	14	218	152	66	0	373	315	58	0	2836	2563	273	0	3471	3169	302	0	4319	3978	341	0
	15	216	134	82	0	369	304	66	0	2791	2464	327	0	3420	3100	320	0	4241	3825	416	0
	16	215	118	97	0	365	292	73	0	2762	2373	389	0	3378	3040	338	0	4181	3679	502	0
	17	214	106	109	0	361	281	80	0	2751	2291	459	0	3344	2989	356	0	4150	3552	598	0
	18	214	97	117	0	357	270	87	0	2759	2222	538	0	3319	2946	374	0	4164	3458	706	0
	19	215	94	121	0	353	259	94	0	2789	2166	624	0	3303	2911	391	0	4241	3416	824	0
	20	216	97	119	0	349	248	101	0	2843	2126	718	0	3295	2885	409	0	4402	3448	954	0
Vowel	Age	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value
/u/	4	267	230	37	0.0169	460	436	23	0.1528	1338	1193	145	0	3791	3750	42	0.5205	5317	5172	145	0.4755
	5	245	236	9	0.3385	460	433	26	0.078	1326	1182	145	0	3704	3638	66	0.2677	4940	4689	251	0.0154
	6	232	240	-8	0.2594	457	427	29	0.0311	1315	1170	145	0	3608	3517	91	0.0963	4751	4486	265	0.0033
	7	224	241	-17	0.0143	452	419	32	0.0087	1303	1158	145	0	3505	3390	115	0.02	4653	4426	226	0.0142
	8	221	238	-17	0.012	444	409	35	0.0015	1291	1146	145	0	3398	3259	139	0.0018	4581	4415	166	0.0482
	9	220	232	-12	0.0846	436	397	38	1.00E-04	1279	1134	145	0	3291	3127	164	0	4501	4392	109	0.1324
	10	220	221	-1	0.9113	426	385	41	0	1267	1122	145	0	3185	2997	188	0	4398	4324	75	0.2678
	11	221	206	14	0.0141	416	372	45	0	1255	1110	145	0	3083	2871	212	0	4276	4201	75	0.2792
	12	221	189	31	0	406	359	48	0	1243	1098	145	0	2988	2752	237	0	4148	4032	116	0.1071
	13	220	169	50	0	397	346	51	0	1231	1086	145	0	2903	2642	261	0	4034	3838	196	0.006
	14	218	149	69	0	389	335	54	0	1219	1074	145	0	2830	2545	285	0	3952	3643	308	0
	15	216	129	87	0	382	325	57	0	1207	1062	145	0	2772	2462	310	0	3915	3477	439	0
	16	213	111	102	0	377	317	60	0	1195	1050	145	0	2732	2398	334	0	3926	3359	567	0
	17	212	99	113	0	374	311	63	0	1183	1038	145	0	2711	2353	358	0	3968	3303	666	0
	18	213	93	120	0	375	309	66	0	1171	1026	145	0	2714	2331	383	0	4004	3302	702	0
	19	217	97	120	0	378	310	69	0	1159	1014	145	0	2742	2335	407	0	3968	3332	636	0
	20	226	114	112	0	386	314	72	0	1147	1002	145	0	2798	2367	431	0	3758	3338	420	0.0431
Vowel	Age	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value
/æ/	4	247	224	23	0.0307	1017	1026	-9	0.79070	2802	2786	16	0.751	3906	3862	44	0.6214	5277	4964	314	0.0271
	5	242	238	4	0.5446	1006	1000	6	0.85018	2692	2655	37	0.4336	3814	3764	50	0.4897	5104	4837	268	0.017
	6	238	246	-7	0.1913	996	975	21	0.46635	2589	2532	58	0.1804	3715	3656	58	0.3087	4946	4714	232	0.008
	7	234	246	-12	0.0488	985	949	36	0.16908	2495	2417	78	0.0446	3611	3541	70	0.1283	4801	4595	206	0.0026
	8	230	241	-11	0.1012	974	924	51	0.03104	2409	2310	99	0.0049	3506	3421	85	0.028	4670	4480	190	6.00E-04
	9	227	231	-4	0.4926	963	898	65	0.00190	2331	2211	120	2.00E-04	3402	3299	103	0.0032	4554	4370	184	2.00E-04
	10	223	218	6	0.3598	953	872	80	0.00002	2261	2120	141	0	3301	3177	125	2.00E-04	4451	4263	188	1.00E-04
	11	220	201	19	0.0013	942	847	95	0	2199	2038	161	0	3206	3057	149	0	4363	4161	202	0
	12	217	183	34	0	931	821	110	0	2146	1964	182	0	3120	2943	177	0	4289	4062	226	0
	13	215	164	51	0	921	796	125	0	2100	1898	203	0	3045	2837	208	0	4228	3968	260	0
	14	213	146	68	0	910	770	140	0	2063	1840	223	0	2983	2741	242	0	4182	3878	305	0
	15	212	128	84	0	899	744	155	0	2034	1790	244	0	2937	2657	280	0	4150	3792	359	0
	16	212	113	98	0	888	719	170	0	2013	1748	265	0	2910	2590	321	0	4132	3710	423	0
	17	212	101	110	0	878	693	184	0	2000	1715	285	0	2904	2540	365	0	4129	3632	497	0
	18	212	93	119	0	867	668	199	0	1995	1689	306	0	2922	2510	412	0	4139	3558	581	0
	19	214	91	123	0	856	642	214	0	1998	1672	327	0	2966	2504	462	0	4163	3489	675	0
	20	216	94	122	0	846	616	229	0	2010	1663	347	0	3039	2523	516	0	4202	3423	779	0
Vowel	Age	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value	Female	Male	Difference	p-value
/ɑ/	4	248	218	31	0.0608	1294	1268	26	0.3954	1946	1896	50	0.2627	3636	3532	105	0.1303	4867	4683	185	0.0954
	5	241	234	7	0.4701	1255	1215	40	0.1619	1893	1827	66	0.1105	3583	3464	119	0.0601	4739	4576	163	0.0633
	6	235	242	-7	0.3297	1218	1165	53	0.0403	1842	1761	82	0.0309	3519	3385	134	0.0207	4621	4473	148	0.0329
	7	230	244	-14	0.0305	1183	1117	66	0.0047	1795	1698	97	0.0047	3446	3298	148	0.0047	4513	4374	139	0.0127
	8	226	240	-15	0.0276	1151	1071	80	2.00E-04	1752	1639	113	3.00E-04	3367	3204	163	6.00E-04	4414	4278	136	0.004
	9	222	232	-10	0.1483	1122	1029	93	0	1711	1583	128	0	3285	3107	177	0	4324	4186	139	0.0014
	10	219	219	0	0.9984	1095	988	106	0	1674	1530	144	0	3202	3010	192	0	4245	4097	148	6.00E-04
	11	217	204	13	0.0313	1070	950	120	0	1640	1480	160	0	3120	2914	207	0	4175	4012	163	2.00E-04
	12	215	187	28	0	1048	915	133	0	1609	1434	175	0	3043	2822	221	0	4115	3930	184	0
	13	214	169	45	0	1029	882	146	0	1582	1391	191	0	2974	2738	236	0	4064	3852	212	0
	14	213	151	62	0	1012	852	160	0	1558	1351	206	0	2914	2664	250	0	4023	3778	245	0
	15	212	133	79	0	997	824	173	0	1537	1315	222	0	2866	2601	265	0	3992	3707	285	0
	16	211	117	94	0	985	799	186	0	1519	1282	238	0	2834	2554	279	0	3971	3640	331	0
	17	210	104	106	0	976	776	200	0	1505	1252	253	0	2819	2525	294	0	3959	3576	383	0
	18	209	95	114	0	969	755	213	0	1494	1225	269	0	2824	2516	309	0	3957	3516	440	0
	19	208	90	118	0	964	738	226	0	1486	1202	284	0	2853	2530	323	0	3964	3460	505	0
	20	206	90	116	0	962	722	240	0	1481	1182	300	0	2907	2569	338	0	3982	3407	575	0

FIG. 3. (Color online) Graphic and numeric display of VFR-based mean frequency (Hz) measurements of ages 4–20 for female and male speakers, and corresponding differences and statistical significance *p*-values by frequency and corner vowels.

the vowel acoustic space. Abrupt drops in formant frequency are not readily apparent in Figs. 1(a)–1(d), where the overall pattern is one of smooth decrease. This is likely due to the consistent methodology employed across all ages in this study that addresses the various methodological issues noted in the Introduction (Kent and Vorperian, 2018). The data in the present study differ from previously reported data in some respects. One difference is in the data for vowel /u/, especially for F1 and F2. In the present study, these formants changed very little with age in either sex. The mean frequency of F2 for this vowel decreased less than 300 Hz over the age range of 4–20 years old for both males and females. In addition, the mean F1 and F2 frequencies for vowel /u/ in the present study are substantially lower than those in the studies of Lee *et al.* (1999; who used the word *boot*) and Hillenbrand *et al.* (1995; who used the word *who'd*). In both the present study and the study by Eichhorn *et al.* (2018), the F1 and F2 values for vowel /u/ in adults agree with those of Peterson and Barney (1952) but not so well with those of Lee *et al.* (1999) and Hillenbrand *et al.* (1995). Given the correspondence of the present results with those of Peterson and Barney (1952), it does not seem that generational or diachronic differences are involved. The discrepant results for this vowel are most likely the consequence of either dialectal variations or differences in word selection and measurement procedures. Regarding the last mentioned possibility, the words *boot* and *who'd* may have been produced with /u/ fronting, which induces an increase in F2 frequency. Fronting of vowel /u/ (also called *goose fronting* because *goose* is a frequently used keyword for this vowel) has been noted in nearly all varieties of North American English (Labov *et al.*, 2006). In addition, the procedure followed in the present study was to measure the formant frequencies of vowel /u/ at the point in time when F2 reached its lowest frequency.

The present data, when combined with those of Eichhorn *et al.* (2018), show the pattern of sex-specific changes in f_o and the first four formant frequencies over the age range of 4–92 years old. The data for female speakers reach adult values by about 16 years of age, whereas the data for male speakers reach adult values at about 20. Eichhorn *et al.* reported that formant frequencies are essentially stable throughout adulthood but f_o decreases significantly with age in women. These results can be interpreted to mean that vocal tract length reaches its adult size in adolescence for

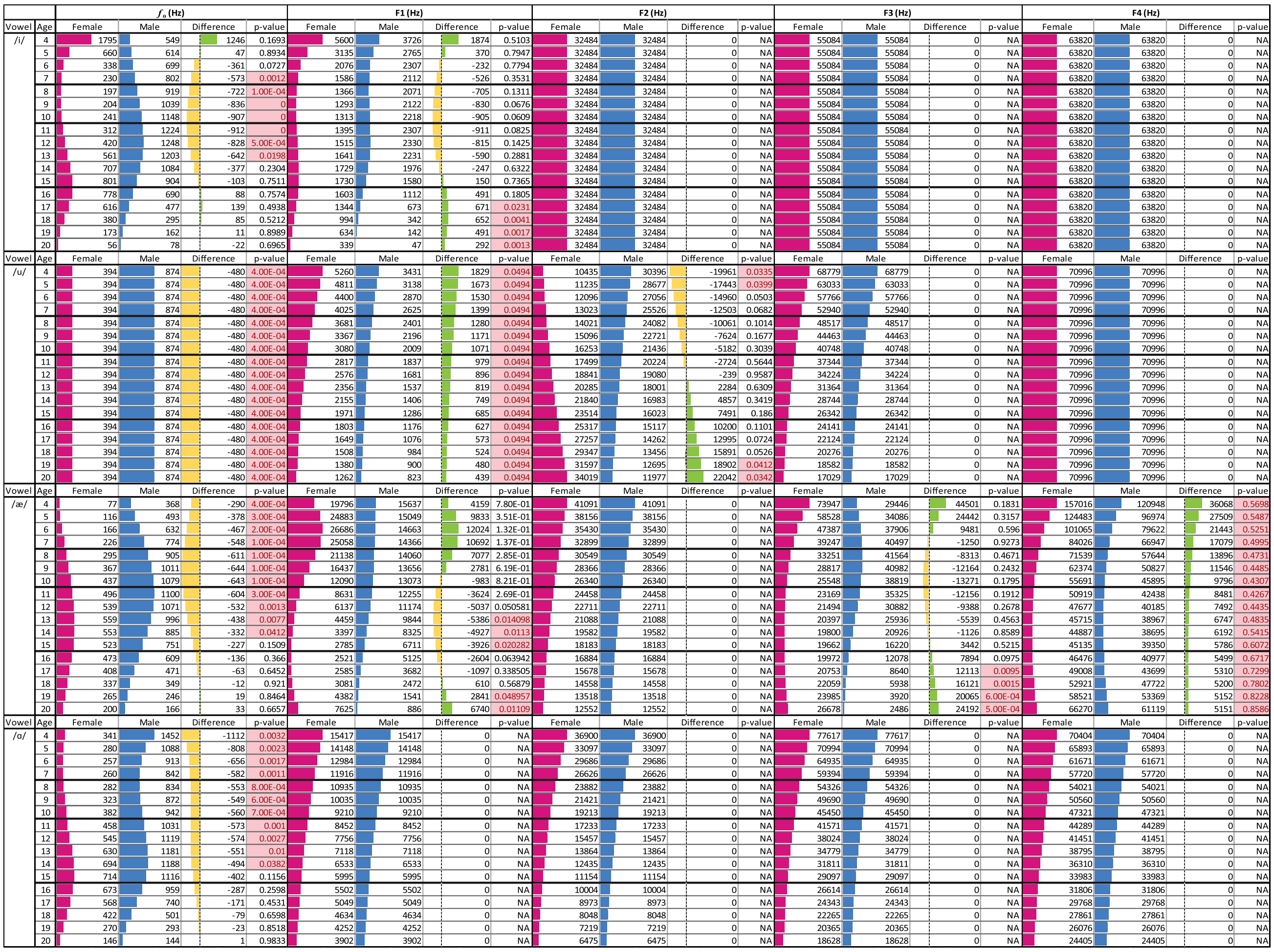


FIG. 4. (Color online) Graphic and numeric display of VFR-based age-centered variance (Hz^2) of ages 4–20 for female and male speakers with corresponding differences and statistical significance p-values by frequency and corner vowel.

females or early adulthood for males and changes little, if at all, throughout adulthood in healthy individuals.

B. Sex differences in f_o and formant frequencies

Given previously reported data (Lee *et al.*, 1999; Maturo *et al.*, 2012; Sorenson, 1989), we expected sex differences in f_o to emerge at about 12 years of age. However, the mean data in the present study show significant f_o sex differences emerging at age 7 with a steady gradual decrease of f_o in males beginning at about age 7, while females show little if any change around this age period. Similarly, Nicollas *et al.* (2008) reported that boys have a lower f_o than girls, even before mutation. The maturational change in f_o in females is accomplished primarily between the ages of 7 and 14, whereas in males it is accomplished primarily between the ages of 7 and 16, over which period there is a drop of approximately one octave. These results differ from those in a large pediatric database reported by Maturo *et al.* (2012) who concluded that boys reach the adult mean f_o at about 16 with a transition period beginning at about 12 years of age. They also concluded that girls reach the adult mean f_o at around age 14, with a transition period beginning around age 11.

Sex differences in formant frequencies appeared at the earliest ages under study but varied with vowel and formant. Typically, once differences emerged, they persisted through age 20. For F1 frequency, sex differences emerged between the ages of 6 and 9 with differences for the back vowels /u/ and /ɑ/ emerging earlier than the front vowels /æ/ and /i/. For F2 frequency, a sex difference was present at age 4 for the high vowels /i/ and /u/, with differences evident for the low vowels /ɑ/ and /æ/ at ages 6 and 7, respectively. For F3 frequency, a significant sex difference emerged at age 5 for the high-front vowel /i/ and present for all vowels by age 8. For F4 frequency, sex differences appeared for all vowels at the age interval of 4–8 years with differences present for the low-front vowel /æ/ at age 4, followed by the vowels /ɑ/ and /i/ at ages 6 and 7, respectively. Previous studies have reported sex differences in vowel formants for children as young as 3 or 4 years of age (Perry *et al.*, 2001; Whiteside, 2001; Yang and Mu, 1989), and the present data confirm and extend these results. It can be concluded that sex differences in one or more of the formant frequencies of the corner vowels are evident by at least the age of 4.

The origin and significance of these sex differences are uncertain. Speech is sexually dimorphic, strikingly so in male:female ratios compared with other physical differences (Rendall *et al.*, 2005). Sexual dimorphism in the acoustic

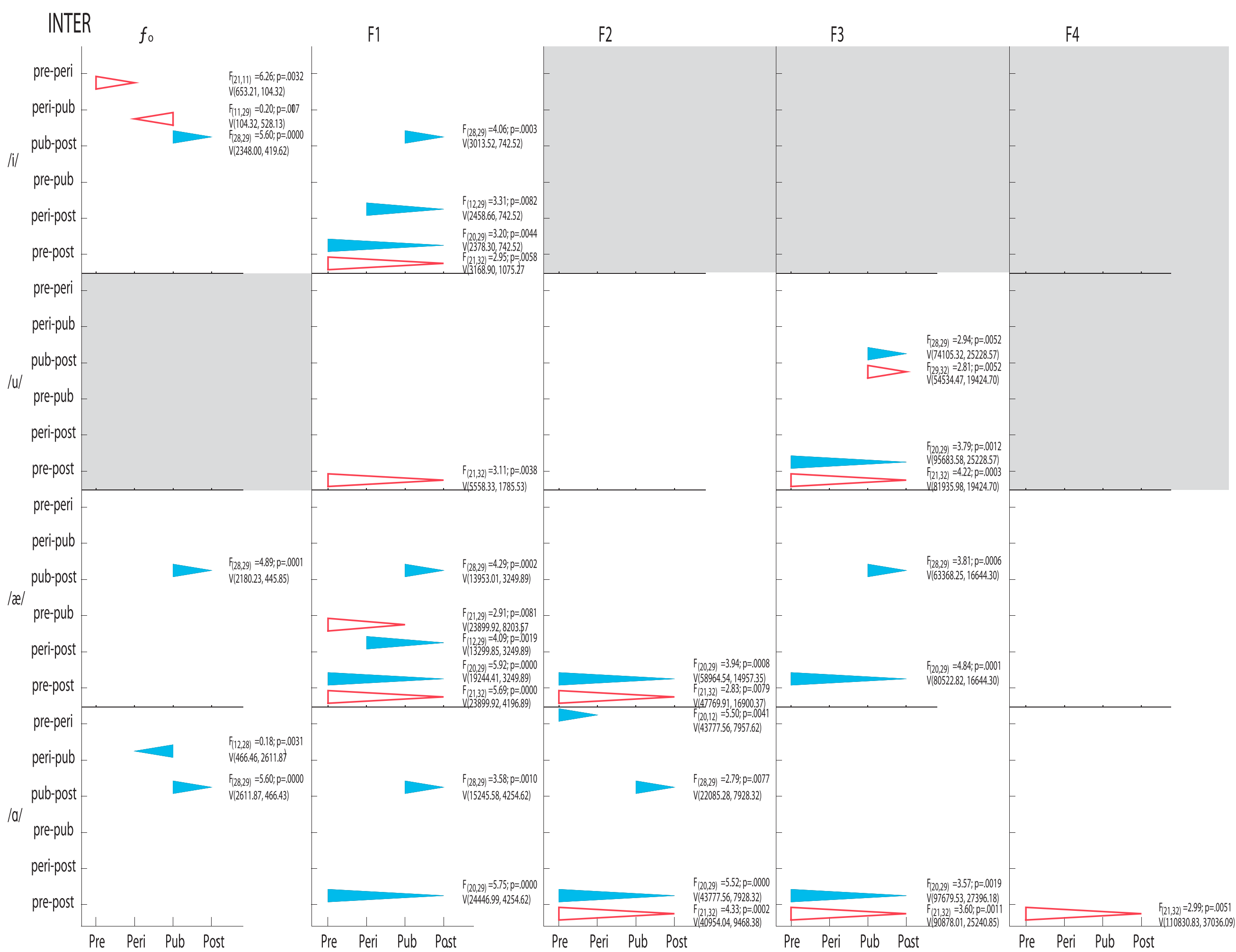


FIG. 5. (Color online) Inter-speaker variability. Significant changes in variance of f_o, F1–F4 across the four age-cohorts for each of the corner vowels are displayed for male and female speakers with filled and open horizontal triangular bars, respectively. The base of the triangular bar represents the maximal variance at the pubertal-stage age-cohort. The pointed apex of the triangular bar reflects the direction of change in variance. Grayed-out plots indicate that the F-test was not carried out since the model predicted variance remained stable as a function of age (see Fig. 4).

signal of speech arises for two primary reasons: (1) anatomic and physiologic differences between males and females, and (2) articulatory and phonatory adjustments that speakers make to sound more like one sex than the other. Regarding the first point, Fitch and Giedd (1999) concluded that sexual dimorphism in the vocal tract is not evident until puberty. However, Vorperian *et al.* (2011) concluded that pre-pubertal sex differences exist first in the oral region, then the pharyngeal region of the vocal tract once growth rate differences between males and females are accounted for. A possible limitation to the available anatomic data is that measures of only the length of the vocal tract (Fant, 1960), or portions of the vocal tract, do not account for the acoustic differences between males and females. That is, it is necessary to obtain data on the regional volumes of the vocal tract during development in both sexes. Only recently have data been reported on sex differences in the hypopharynx in adults (Zhang *et al.*, 2019); however, it is not known when sexual dimorphism emerges and whether it is present in children. Regarding the second point, it has been reported that acoustic differences between the sexes result from learning gender-specific speech patterns (Cartei *et al.*, 2014; Cartei *et al.*, 2012; Cartei and Reby, 2013; Johnson, 2006; Pisanski *et al.*, 2016). It is not possible to reach a definitive conclusion from the present data as to the relative roles of the anatomic and speech-learning interpretations. More refined anatomic data are required particularly of the oral and pharyngeal regions. For example, Kelly *et al.* (2017) document the presence of pubertal and pre-pubertal sexual dimorphism of the inferior portion of the mandible with males having greater dimensions in the antero-posterior and medial-lateral planes (e.g., gonion width, gonion angle, and gnathion angle), but if and how such differences in mandibular measurement alter the oral-pharyngeal region is needed. It is likely that the different developmental patterns in the first four formants may hold clues as to anatomic differences in the growing vocal tracts of boys and girls. We tentatively conclude that the vowel-formant interactions in sex differences point to anatomic factors that may interact with speech-learning phenomena.

Another question relating to sexual dimorphism is: when do males and females achieve the adult values of formant frequencies? Comparing the data for males and females [Figs. 1(a)–1(d)], it appears that formant frequencies for

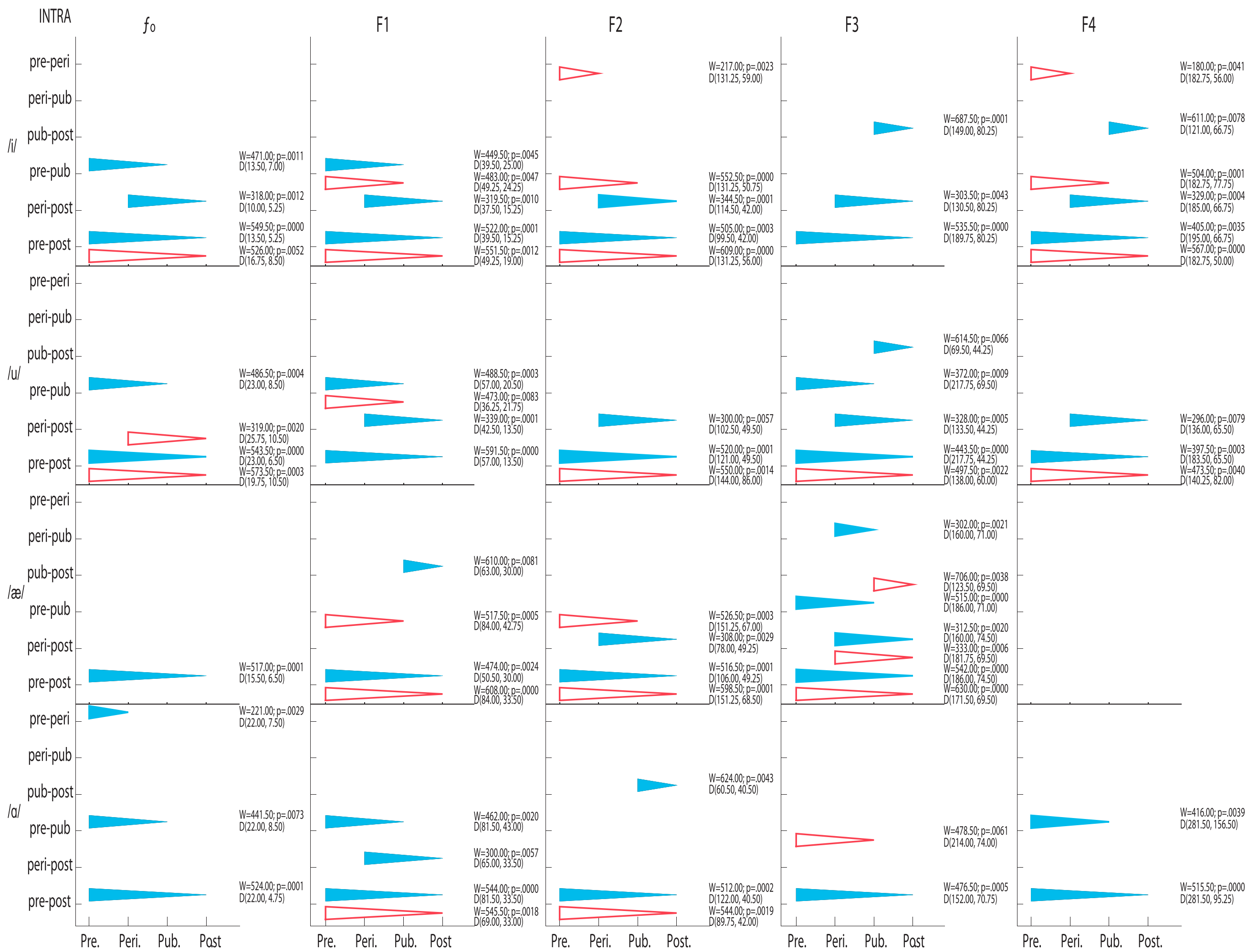


FIG. 6. (Color online) Intra-speaker variability. Significant differences in f_o, F1–F4 across the four age-cohorts for each of the corner vowels in males (shaded triangles) and female (open triangles) speakers. The base of the triangular bar represents the maximal difference at the pubertal-stage age-cohort, and the pointed apex of the triangular bar reflects the direction of change in difference.

females asymptote on adult values at around age 14, whereas formant frequencies for males do not asymptote on adult values until about age 20, the upper limit of age in this study. Lee *et al.* (1999) observed that formant frequencies for males and females diverge beginning at age 11 and progress until about age 15. They concluded that the growth spurt of the vocal tract in males occurs between ages 10 and 15. This conclusion is consistent with findings from imaging studies of the vocal tract confirming a rapid growth rate in vocal tract length in males up to age 15 (Vorperian *et al.*, 2009); however, findings from imaging studies also indicate that vocal tract lengthening continues beyond the age of 15 in males but not as much in females (Fitch and Giedd, 1999; Vorperian *et al.*, 2009). If the present results are compared with those in Eichhorn *et al.* (2018; a companion study using the same methods), it appears that formant frequencies in males asymptote at adult values at ages 19–20, which invites the inference that vocal tract length reaches its maximum at this period. Formant frequencies in later years of adulthood exhibit variable patterns across studies, but based on findings by Eichhorn *et al.* (2018) who used a larger number of participants than most other studies and applied consistent methodology across the young, middle, and older adult age-cohorts, it does not appear that additional lengthening of the vocal tract necessarily occurs.

C. Inter- and intra-speaker variability across development for the corner vowels in f_o and formant frequencies

Adults are capable of reliable vowel production for both sustained vowels (Vogel *et al.*, 2011), citation form vowels (Heald and Nusbaum, 2015), and even after perturbation of articulation by a bite block (Lindblom *et al.*, 1977; Lindblom and Sundberg, 1971). For citation speech, Heald and Nusbaum reported some within-day variation in f_o and F1 but no significant changes in f_o and F1–F3 between days. The authors concluded that adults have a high level of internal precision and consistency. The developmental question is: At what age is precise and reliable vowel production achieved? An answer to this question is important for purposes such as automatic speech recognition (where error rates are higher for children than adults) and clinical assessment of speech disorders (where atypical variability may be a sign of a disorder such as childhood apraxia of speech). At least three major factors should be considered in the interpretation of acoustic data pertaining to the precision of vowel production.

(1) Age-related measurement error is particularly relevant to formant frequencies, given that the error of measurement is related to vocal fundamental frequency (Lindblom, 1962; Chen *et al.*, 2019). It is also likely that the error of formant estimation varies across formants, with the higher formants F3 and F4 being more susceptible to error given their larger bandwidths and lower energy. The risk is that measurement error is not easily distinguished from variability arising from developmental or other processes.
(2) Variability in acoustic and physiologic measures is commonly taken as an index of maturity of speech motor control, and used as evidence that maturity is not reached until late adolescence (Cheng *et al.*, 2007; Walsh and Smith, 2002) and perhaps even as late as 30 years of age (Schötz *et al.*, 2013). This conclusion applies to vowels as well as overall speech production. Yang and Fox (2013) concluded that children's vowel production is marked by substantial developmental change besides the effect of vocal tract lengthening. They noted that the "acoustical development of vowels from children to adult norms is a long-term process" (Yang and Fox, 2013, p. 1266).
(3) Within the protracted development noted in factor (2) above, maturity apparently is reached at different times for different aspects (e.g., spatial versus temporal) of speech production (Lee *et al.*, 1999; Nittrouer, 1995; Smith and Goffman, 1998; Stathopoulos, 1995). Therefore, maturation of speech motor control is a multilayered process of overlapping biological (anatomic and physiologic) and linguistic developments, rather than a monolithic process. Furthermore, speech motor control is not necessarily a continuous, monotonic process. Smith and Zelaznik (2004) concluded that late childhood (7– 12 years of age) is a plateau in the development of coordinative synergies for speech production. It is likely that motor control adapts to ongoing changes in anatomy, physiology, and phonology, each of which has a developmental pattern.

In the data from the present study, the overall trend is that both inter- and intra-speaker variability decreased with age of speaker. Although age-dependent measurement error [as discussed above in factor (1)] cannot be entirely excluded as at least a partly explanatory factor; it is likely that the patterns of change in variability reflect anatomic growth and motor control (and their interaction). Specifically, increased inter-speaker variability is probably associated with periods of rapid growth during which individual differences are large (e.g., somewhat different ages of onset of puberty, and different rates of growth between speakers), whereas increased intra-speaker variability is likely more reflective of maturing motor control.

Developmental differences in inter- versus intra- variability of the first four formants may have implications for determining anatomic versus motoric contributions to variability. For example, inter-speaker variability for the low vowels /ɑ/ and /æ/ were significantly greater than the high vowels /i/ and /u/, as also reported by Yang and Fox (2013). Factors that might explain the better precision of high over low vowels are that the high vowels are associated with (a) somatosensory feedback of tongue contact with the palate, teeth, or both (Gick *et al.*, 2017; Mitsuya *et al.*, 2015), (b) lateral bracing against the upper structures of the oral cavity, which stabilizes articulation (Gick *et al.*, 2017), and (c) saturation effects in the relationship between articulation and acoustic result (Perkell *et al.*, 1997). A longitudinal study design that includes acoustic and physiologic measures of inter- and intra- variability will likely help gain a clearer understanding of the key contributors to inter- versus intra-variability.

D. Comparison with data from previous research

As pointed out in Kent and Vorperian (2018), published data on the corner vowels are not in complete agreement, and the differences affect both the position and the shape of the vowel quadrilateral in the traditional F1-F2 plane. The differences could arise from several factors, including dialectal differences among studies, methodological differences in formant-frequency estimation, and selection of words for the vowels of interest. A notable difference, as reported in Sec. IV A, include differences between the present data and those of Lee *et al.* (1999) in the F2 frequency of vowel /u/. The present data for F1 frequency of the high vowels /i/ and /u/ are lower than those in Hillenbrand *et al.* (1995) for both male and female adults. These differences complicate efforts to establish normative data on metrics such as VSA or measures of formant-frequency dispersion. For example, VSA is arguably one of the most frequently reported acoustic measures of disordered speech, but there does not appear to be a common source of normative data for the clinical interpretation of VSA values. Most studies using this metric to evaluate clinical populations report their own normative data for comparison. The same comment applies to alternative measures, such as measures of vowel formant dispersion. Accounting for factors, such as the use of consistent data collection and analysis methodology across development—as done in this study using methodology consistent with that of Eichhorn *et al.* (2018) in the adult population—may contribute to a reliable normative database across the lifespan. Similar studies across different regions and ages less than 4 can further help establish normative data.

The present study shows a relatively early emergence of sex differences in f_o and formant frequencies (particularly the higher formants). These differences likely reflect anatomic growth and remodeling, but sociocultural factors cannot be discounted. The acoustic properties of vowels become sex distinctive well before puberty.

ACKNOWLEDGMENTS

This research was supported by National Institutes of Health (NIH) research Grant No. R01 DC6282 entitled "*MRI and CT Studies of the Developing Vocal Tract*" from the National Institute of Deafness and other Communicative

Disorders (NIDCD), and also, core grants, Grant Nos. P30 HD03352 and U54 HD090256, from the National Institute of Child Health and Human Development (NICHHD). The authors thank the following individuals for assistance with data collection and acoustic analysis: Carlyn Burris, Ekaterini Derdemezis, Erin Douglas, Julie Eichhorn, Katie Lester, Jennifer Lewandowski, Lela Murchison, Allison Petska, and Elaine Romenesko, as well as Ying Ji Chuang and Courtney A. Miller for assistance with figures and tables, and Sophie Blankenheim for assistance with references.

[1]Three speakers (two females and one male, ages 4, 18, and 18, respectively) had missing F4 mean frequency measurements for the vowel /æ/ because either all five data points or four out of the five data points for F4 were missing.

[2]Among repeated words, the first recording was selected unless it had one or more missing frequency (f_o or F1–F4) measurements, in which case the second recording was used, provided it had fewer missing measurements.

[3]Initially, the mean structure parameters in Eq. (1) were estimated using ordinary least squares to yield provisional estimates of $\hat{\beta}$, and for each observation its corresponding residual, $\tilde{\epsilon}$. The $\hat{\epsilon}^2$ were, in turn, used to define a provisional estimate of σ_i^2. This estimate of σ_i^2 was then modeled as an outcome in Eq. (2) using a gamma regression model with log link to determine provisional coefficient estimates $\hat{\lambda}$. Then, the inverse of the predicted $\hat{\sigma^2}$ from the gamma regression were used as weights to re-estimate the parameters in Eq. (1), and the process was repeated until convergence to yield the final estimates.

[4]Specifically, by centering age at each of the possible age values, we altered the age at which the speaker-sex (male) term evaluates a sex difference for both the mean and variance. By testing the significance of the male term, we therefore tested the existence of speaker-sex difference at the specific age of the centering, and thus evaluated the statistical significance of sex difference at every possible age location for the mean frequency measure and the inter-speaker variance [values, respectively, displayed in Figs. 3 and 4 in Sec. III (Results)].

[5]See supplementary material at https://doi.org/10.1121/1.5131271 for a summary of the results on inter-speaker variance (Hz^2) estimates by speaker-sex for each age-cohort; and a summary of the results of the intra-speaker median absolute discrepancy scores by speaker-sex for each age-cohort.

Assmann, P. F., Nearey, T. M., and Bharadwaj, S. (**2008**). "Analysis of a vowel database," Can. Acoust. **36**(3), 148–149.

Bele, I. V. (**2006**). "The speaker's formant," J. Voice **20**(4), 555–578.

Boersma, P., and Weenink, D. (**2010**). "Praat: Doing phonetics by computer (version 5.1.32) [computer program]," http://www.fon.hum.uva.nl/praat (Last viewed September 1, 2013).

Boseley, M. E., and Hartnick, C. J. (**2006**). "Development of the human true vocal fold: Depth of cell layers and quantifying cell types within the lamina propria," Ann. Oto. Rhinol. Laryn. **115**(10), 784–788.

Burris, C., Vorperian, H. K., Fourakis, M., Kent, R. D., and Bolt, D. M. (**2014**). "Quantitative and descriptive comparison of four acoustic analysis systems: Vowel measurements," J. Speech Lang. Hear. Res. **57**(1), 26–45.

Buzaneli, E. C. P., Zenari, M. S., Kulcsar, M. A. V., Dedivitis, R. A., Cernea, C. R., and Nemr, K. (**2018**). "Supracricoid laryngectomy: The function of the remaining arytenoid in voice and swallowing," Int. Arch. Otorhinolaryngol. **22**(03), 303–312.

Cartei, V., Cowels, W., Banerjee, R., and Reby, D. (**2014**). "Control of voice gender in pre-pubertal children," Brit. J. Dev. Psychol. **32**(1), 100–106.

Cartei, V., Cowles, H. W., and Reby, D. (**2012**). "Spontaneous voice gender imitation abilities in adult speakers," PloS One **7**(2), e31353.

Cartei, V., and Reby, D. (**2013**). "Effect of formant frequency spacing on perceived gender in pre-pubertal children's voices," PLoS One **8**(12), e81022.

Chen, W.-R., Whalen, D. H., and Shadle, C. H. (**2019**). "F0-induced formant measurement errors result in biased variabilities," J. Acoust. Soc. Am. **145**(5), EL360–EL366.

Cheng, S. (**2014**). "A life course trajectory framework for understanding the intracohort pattern of wage inequality," Am. J. Sociol. **120**(3), 633–700.

Cheng, H. Y., Murdoch, B. E., Goozée, J. V., and Scott, D. (**2007**). "Physiologic development of tongue–jaw coordination from childhood to adulthood," J. Speech Lang. Hear. Res. **50**(2), 352–360.

De Boor, C. (**1978**). *A Practical Guide to Splines* (Springer, New York).

Derdemezis, E., Vorperian, H. K., Kent, R. D., Fourakis, M., Reinicke, E. L., and Bolt, D. M. (**2015**). "Optimizing vowel formant measurements in four acoustic analysis systems for diverse speaker groups," Am. J. Speech-Lang. Pathol. **25**(3), 335–354.

Disner, S. F. (**1980**). "Evaluation of vowel normalization procedures," J. Acoust. Soc. Am. **67**(1), 253–261.

Donegan, P. (**2013**). "Normal vowel development," in *Handbook of Vowels and Vowel Disorders*, edited by M. J. Ball and F. E. Gibbon (Psychology, New York), pp. 42–78.

Eguchi, S., and Hirsh, I. J. (**1969**). "Development of speech sounds in children," Acta Oto-laryngol. Suppl. **257**, 1–51.

Eichhorn, J. T., Kent, R. D., Austin, D., and Vorperian, H. K. (**2018**). "Effects of aging on vocal fundamental frequency and vowel formants in men and women," J. Voice **32**(5), 644.e1–664.e9.

Fant, G. (ed.) (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fant, G., and Pauli, S. (**1974**). "Spatial characteristics of vocal tract resonance modes," in *Proc. of SCS-74, Speech Commun.*, edited by G. Fant, Stockholm, Vol. 2, pp. 121–132.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**(3), 1511–1522.

Gick, B., Allen, B., Roewer-Després, F., and Stavness, I. (**2017**). "Speaking tongues are actively braced," J. Speech Lang. Hear. Res. **60**(3), 494–506.

Hagiwara, R. E. (**1995**). "Acoustic realizations of American /r/ as produced by women and men (formant frequencies)," in *UCLA Working Papers in Phonetics*, Vol. 90.

Hamdan, A. L., Khandakji, M., and Macari, A. T. (**2018**). "Maxillary arch dimensions associated with acoustic parameters in prepubertal children," Angle Orthod. **88**(4), 410–415.

Hartnick, C. J., Rehbar, R., and Prasad, V. (**2005**). "Development and maturation of the pediatric human vocal fold lamina propria," Laryngoscope **115**(1), 4–15.

Heald, S. L. M., and Nusbaum, H. C. (**2015**). "Variability in vowel production within and between days," PLoS One **10**(9), e0136791.

Hillenbrand, J., and Gayvert, R. T. (**1993**). "Vowel classification based on fundamental frequency and formant frequencies," J. Speech. Lang. Hear. Res. **36**(4), 694–700.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5), 3099–3111.

Hodge, M., Daniels, J., and Gotzke, C. L. (**2009**). "TOCS+ intelligibility measures (version 5.3) [computer software]" (University of Alberta, Edmonton, Canada).

Ishii, K., Akita, M., Yamashita, K., and Hirose, H. (**2000**). "Age-related development of the arrangement of connective tissue fibers in the lamina propria of the human vocal fold," Ann. Otol. Rhinol. Laryngol. **109**(11), 1055–1064.

Johnson, K. (**2006**). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," J. Phonetics **34**(4), 485–499.

Kahane, J. C. (**1978**). "A morphological study of the human prepubertal and pubertal larynx," Dev. Dynam. **151**(1), 11–19.

Kelly, M. P., Vorperian, H. K., Wang, Y., Tillman, K. K., Werner, H. M., Chung, M. K., and Gentry, L. R. (**2017**). "Characterizing mandibular growth using three-dimensional imaging techniques and anatomic landmarks," Arch. Oral Biol. **77**, 27–38.

Kent, R. D. (**1976**). "Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies," J. Speech Hear. Res. **19**(3), 421–447.

Kent, R. D., and Vorperian, H. K. (**2018**). "Static measurements of vowel formant frequencies and bandwidths: A review," J. Commun. Disord. **74**, 74–97.

Kulak Kayikci, M. E., Akan, S., Ciger, S., and Ozkan, S. (**2012**). "Effects of Hawley retainers on consonants and formant frequencies of vowels," Angle Orthod. **82**(1), 14–21.

Labov, W., Ash, S., and Boberg, C. (**2006**). *The Atlas of North American English: Phonetics, Phonology and Sound Change: A Multimedia Reference Tool* (De Gruyter, Mouton).

Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford, UK), p. 193.

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**(3), 1455–1468.
Leino, T., Laukkanen, A. M., and Radolf, V. (**2011**). "Formation of the actor's/speaker's formant: A study applying spectrum analysis and computer modeling," J. Voice **25**(2), 150–158.
Lin, P. T., Gould, W. J., Fukazawa, T., and El-Assuooty, A. (**1989**). "Acoustic analysis of voice in tonsillectomy," J. Voice **3**(1), 81–86.
Lindblom, B. (**1962**). "Accuracy and limitations of sona-graph measurements," in *Proceedings of the Fourth International Congress of Phonetic Sciences*, pp. 188–202.
Lindblom, B., Lubker, J., and Gay, T. (**1977**). "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation," J. Acoust. Soc. Am. **62**(S1), S15.
Lindblom, B. E., and Sundberg, J. E. (**1971**). "Acoustical consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Am. **50**(4B), 1166–1179.
Litman, R. S., Weissend, E. E., Shibata, D., and Westesson, P. L. (**2003**). "Developmental changes of laryngeal dimensions in unparalyzed, sedated children," Anesthesiology **98**(1), 41–45.
Maturo, S., Hill, C., Bunting, G., Ballif, C., Maurer, R., and Hartnick, C. (**2012**). "Establishment of a normative pediatric acoustic database," Arch. Otolaryngol. **138**(10), 956–961.
Milenkovic P. (**2010**). "Time-frequency analysis software program for 32-bit Windows (alpha version 1.2) [computer software]," available at http://userpages.chorus.net/cspeech/ (Last viewed July 19, 2016).
Mitsuya, T., MacDonald, E. N., Munhall, K. G., and Purcell, D. W. (**2015**). "Formant compensation for auditory feedback with English vowels," J. Acoust. Soc. Am. **138**(1), 413–424.
Montt, G. (**2011**). "Cross-national differences in educational achievement inequality," Sociol. Ed. **84**(1), 49–68.
Morrison, G. S., and Assmann, P. F. (**2013**). *Vowel Inherent Spectral Change* (Springer, Heidelberg, Berlin).
Munson, B., and Solomon, N. P. (**2004**). "The effect of phonological neighborhood density on vowel articulation," J. Speech Lang. Hear. Res. **47**(5), 1048–1058.
Nicollas, R., Garrel, R., Ouaknine, M., Giovanni, A., Nazarian, B., and Triglia, J. M. (**2008**). "Normal voice in children between 6 and 12 years of age: Database and nonlinear analysis," J. Voice **22**(6), 671–675.
Nittrouer, S. (**1995**). "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," J. Acoust. Soc. Am. **97**(1), 520–530.
Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., and Guiod, P. (**1997**). "Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models," Speech Commun. **22**(2-3), 227–250.
Perry, T. L., Ohde, R. N., and Ashmead, D. H. (**2001**). "The acoustic bases for gender identification from children's voices," J. Acoust. Soc. Am. **109**(6), 2988–2998.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**(2), 175–184.
Pisanski, K., Mora, E. C., Pisanski, A., Reby, D., Sorokowski, P., Frackowiak, T., and Feinberg, D. R. (**2016**). "Volitional exaggeration of body size through fundamental and formant frequency modulation in humans," Sci. Rep. **6**, 34389.
Rendall, D., Kollias, S., and Net, C. (**2005**). "Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," J. Acoust. Soc. Am. **117**(5), 944–955.
Rogers, D. J., Setlur, J., Raol, N., Maurer, R., and Hartnick, C. J. (**2014**). "Evaluation of true vocal fold growth as a function of age," Otolaryng. Head Neck **151**(4), 681–686.
Sato, K. (**2018**). "Growth and development of the human vocal fold mucosa," in *Functional Histoanatomy of the Human Larynx*, edited by K. Sato (Springer, Singapore), pp. 199–211.
Schötz, S., Frid, J., and Löfqvist, A. (**2013**). "Development of speech motor control: Lip movement variability," J. Acoust. Soc. Am. **133**, 4210–4217.
Smith, A., and Goffman, L. (**1998**). "Stability and patterning of speech movement sequence in children and adults," J. Speech Lang. Hear. Res. **41**(1), 18–30.
Smith, A., and Zelaznik, H. N. (**2004**). "Development of functional synergies for speech motor coordination in childhood and adolescence," Dev. Psychobiol. **45**(1), 22–33.
Sorenson, D. N. (**1989**). "A fundamental frequency investigation of children ages 6–10 years old," J. Commun. Disord. **22**(2), 115–123.
Stathopoulos, E. T. (**1995**). "Variability revisited: An acoustic, aerodynamic, and respiratory kinematic comparison of children and adults during speech," J. Phonetics **23**(1-2), 67–80.
Sundberg, J. (**1974**). "Articulatory interpretation of the 'singing formant,'" J. Acoust. Soc. Am. **55**(4), 838–844.
Sundberg, J., and Nordström, P.-E. (**1976**). "Raised and lowered larynx-the effect on vowel formant frequencies," STL-QPSR **17**(2–3), 035–039.
Švancara, P., Horácek, J., Vokrál, J., and Cerný, L. (**2006**). "Computational modelling of effect of tonsillectomy on voice production," Logoped. Phoniatr. Vocol. **31**(3), 117–125.
Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., and Honda, K. (**2006**). "Acoustic roles of the laryngeal cavity in vocal tract resonance," J. Acoust. Soc. Am. **120**(4), 2228–2238.
Vogel, A. P., Fletcher, J., Snyder, P. J., Fredrickson, A., and Maruff, P. (**2011**). "Reliability, stability, and sensitivity to change and impairment in acoustic measures of timing and frequency," J. Voice **25**(2), 137–149.
Vorperian, H. K. (**2000**). "Anatomic development of the vocal tract structures as visualized by MRI," Doctoral dissertation, University of Wisconsin, available at https://search.library.wisc.edu/catalog/999923575302121 (Last viewed June 14, 2019).
Vorperian, H. K., and Kent, R. D. (**2007**). "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," J. Speech Lang. Hear. Res. **50**(6), 1510–1545.
Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J., and Gentry, L. R. (**2009**). "Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study," J. Acoust. Soc. Am. **125**(3), 1666–1678.
Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., and Chung, M. K. (**2011**). "Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study," J. Speech Lang. Hear. Res. **54**(4), 995–1010.
Walsh, B., and Smith, A. (**2002**). "Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes," J. Speech Lang. Hear. Res. **45**(6), 1119–1133.
Wani, T. M., Bissonnette, B., Rafiq Malik, M., Hayes, D., Jr., Ramesh, A. S., Al Sohaibani, M., and Tobias, J. D. (**2016**). "Age-based analysis of pediatric upper airway dimensions using computed tomography imaging," Pediatr. Pulm. **51**(3), 267–271.
Western, B., and Bloome, D. (**2009**). "Variance function regression for studying inequality," Sociol. Methodol. **39**(1) 293–326.
Whiteside, S. P. (**2001**). "Sex-specific fundamental and formant frequency patterns in a cross-sectional study," J. Acoust. Soc. Am. **110**, 464–478.
Wild, A., Vorperian, H. K., Kent, R. D., Bolt, D. M., and Austin, D. (**2018**). "Single-word speech intelligibility in children and adults with Down syndrome," Am. J. Speech-Lang. Pathol. **27**(1), 222–236.
Wysocki, J., Kielska, E., Orszulak, P., and Reymond, J. (**2008**). "Measurements of pre- and postpubertal human larynx: A cadaver study," Surg. Radiol. Anat. **30**(3), 191–199.
Yang, J., and Fox, R. A. (**2013**). "Acoustic development of vowel production in American English children," in *Interspeech*, pp. 1263–1267.
Yang, S., and Mu, L. (**1989**). "An investigation of the third formant of /a/ in prepubertal children," J. Voice **3**(4), 321–323.
Zhang, J., Honda, K., Wei, J., and Kitamura, T. (**2019**). "Morphological characteristics of male and female hypopharynx: A magnetic resonance imaging-based study," J. Acoust. Soc. Am. **145**(2), 734–748.
Zheng, H., Yang, Y., and Land, K. C. (**2011**). "Variance function regression in hierarchical age-period-cohort models: Applications to the study of self-reported health," Am. Sociol. Rev. **76**(6), 955–983.